

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

The Path of Power and the Path of Law: How Europe got from feudal anarchy to peace, prosperity and democracy

**Permalink**

<https://escholarship.org/uc/item/23s230sz>

**Author**

Koppel, Bill

**Publication Date**

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Path of Power and the Path of Law:

How Europe got from feudal anarchy to peace, prosperity and democracy

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Political Science

by

William Edward Koppel

2020

© Copyright by  
William Edward Koppel  
2020

## ABSTRACT OF THE DISSERTATION

The Path of Power and the Path of Law:

How Europe got from feudal anarchy to peace, prosperity and democracy

by

William Edward Koppel

Doctor of Philosophy in Political Science

University of California, Los Angeles, 2020

Professor Arthur Stein, Chair

This study provides a new explanation of Europe's transition from feudal anarchy a thousand years ago to peace, prosperity and democracy in the 20th century. I argue that Europe's long transition out of feudal anarchy was the consequence of a series of conflicts and settlements caused by a single problem. Actors repeatedly reached the point where they recognized that competitions in raising relative power were making everyone worse off. To stop the competitions, they agreed to divide the assets over which they were competing according to terms that were *independent* of relative power. They formed contracts whose terms were tied to anchors in the world that did not move. This device reduced each actor's incentive to invest resources in increasing his relative power, for such investments would not shift the contract terms in his favor. The anchor contracts stopped the competitions in raising relative power. Actors also needed a way to enforce the anchor contracts that did not rely primarily on *power* based enforcement tools. For such tools could reignite the competitions in raising relative power. Hence actors developed *anchor* based enforcement tools.

Reliance on anchor contracts created a vulnerability to disputes, even when there were no power shifts or uncertainties about relative power. The single most important cause of disputes was legal incompatibility problems. At the times and places where actors had a high capacity for *legal incompatibility management* (LIM), the disputes were resolved smoothly and development proceeded. At the times and places where actors had low LIM capacity, by contrast, the disputes became intractable and development was hindered. In the long run, Europe oscillated between these two worlds. Where actors had high LIM capacity, they lived

in a world of law that enjoyed political stability, good economic governance, social capital formation and long-run development. Where actors had low LIM capacity, they lived in a world of power that suffered political instability, misgovernance, social capital depletion and underdevelopment. These oscillations occurred at both the domestic and international levels. A single causal model explains the evidence on both levels. The model thus provides a unified explanation of Europe's development.

The theory developed in this study, which I call Contractual Realism, differs from the three main paradigms in the literature on conflict, cooperation and long-run transformation. Contractual Realism identifies a single factor that drives the three variables that are thought to be independent drivers in the main paradigms: power politics, institutional design, and social identity roles. In this sense Contractual Realism resolves the debates between the three conventional paradigms.

## 1. The roots of power politics

The main paradigm in the literature for explaining conflict is Conventional Realism. According to this paradigm, actors divide valuable assets according to relative power at all times – not only in war but in peacetime as well. The main cause of conflict is assumed to be shifts in power in peacetime (and uncertainty about power shifts in peacetime).

Contractual Realism takes a different view. It assumes that actors divide assets according to relative power during conflicts and competitions in raising relative power. But at other times they divide assets according to anchor contracts – agreements whose terms are invariant to shifts in relative power. At those times they live in a world of law rather than a world of power. The main cause of conflict is assumed to be disputes over legal incompatibilities. Once such a dispute becomes intractable, actors can no longer divide assets by anchor contracts. So they revert to dividing assets by relative power and that triggers a competition in raising relative power. In short, they revert to the world of power politics. There are two sub-equilibriums in this world. There is the sub-equilibrium of *conflict* – defined as unilateral asset seizures that elicit an immediate response of a reciprocal or retributive kind. And there is the sub-equilibrium of *instability* – defined as unilateral asset seizures that do not elicit an

immediate response of a reciprocal or retributive kind. As long as actors remain in the world of power politics, they switch back and forth between these two sub-equilibriums. From this perspective, Conventional Realism only explains why actors switch from the instability sub-equilibrium to the conflict sub-equilibrium within the world of power. It does not explain why the actors reverted from the world of law to the world of power in the first place. Contractual Realism explains this. According to Contractual Realism, reversion to the world of power is caused by legal incompatibility disputes in an institutional environment with a low capacity for Legal Incompatibility Management (LIM). Through European history, the largest and longest oscillations between the world of law and the world of power, at both the domestic and international levels, were caused by this factor.

## **2. The roots of good governance and misgovernance**

The main paradigm in the literature for explaining good governance and misgovernance is Conventional Institutionalism. According to this paradigm, there are three main stories about the institutional design that produces good governance and the institutional design flaw that produces misgovernance. In the first story, good governance requires that one actor be strong enough to enforce law and order among the other actors in the system. (In the domestic system the strong enforcer is the state. In the international system the strong enforcer is a hegemonic state.) According to this story, if the other actors in the system are too strong relative to the enforcer, the result is over-*decentralization* of power, institutional instability and misgovernance. In the second story, good governance requires that the other actors be strong enough to keep the enforcer in check. If they are not strong enough, then the result is over-centralization of power, institutional rigidity and misgovernance. In the third story, good governance requires just the right balance of power between the enforcer and the other actors in the system. If the balance tilts too much in the enforcer's direction, the result is abuse of power, rigidity and misgovernance. If the balance tilts too much in the others' direction, the result is a vacuum of power, instability and misgovernance.

Contractual Realism takes a different view. Good governance does not require a particular distribution of power between the enforcer and the other actors in the system. Good governance requires anchor contracts, anchor based enforcement methods, and the capac-

ity for Legal Incompatibility Management (LIM). When these institutions of anchor based governance are present, there will be a relatively low level of investment in power-based enforcement tools by *both* the enforcer and the other actors in the system. Conversely, when anchor based governance institutions are absent, there will be a relatively high level of investment in power-based enforcement assets by *both* the enforcer and the others. In this case the resulting institutional design will depend on the type of disputes that are most salient due to the absence of anchor based governance. When the salient disputes are about *economic* exclusion, the system veers in the direction of over-centralization of power and institutional rigidity, resulting in misgovernance. When the salient disputes are about *political* exclusion, the system veers in the direction of over-decentralization of power and institutional instability, resulting in misgovernance. In either case, the problem of misgovernance is not due to the wrong balance of power between the enforcer and the others. It is due to the absence of anchor based governance institutions. Through European history, it was the lack or loss of anchor based governance that generated these two structures of misgovernance – at both the domestic and international levels.

### **3. The roots of divisive identity politics**

The main paradigm in the literature for explaining the long-run transformation of actors and systems is Conventional Constructivism. According to this paradigm, the main obstacle to long-run development is social, political and economic stratification that divides the system into dominant social groups and subjected social groups. The divide between the dominant and subjected groups is exacerbated by cleavages between different tribal, ethnic, religious and national identity groups. These cleavages and inequalities then generate divisive identity politics.

Contractual Realism takes a different view. Divisive identity politics is driven by the absence or failure of anchor based governance institutions. Through European history, the decay of anchor based governance led to the emergence of divisive identity politics among previously cohesive groups. Conversely, the development of anchor based governance enabled actors in divided societies to cooperate and achieve social capital formation across identity lines that were previously divisive.

The dissertation of William Edward Koppel is approved.

Nico Voigtlander

Robert Trager

Arthur Stein, Committee Chair

University of California, Los Angeles

2020



This study is dedicated to the memory of the people who died in the conflicts studied in these pages, and the hope that a better understanding of these conflicts may help to avoid similar tragedies in the future.

## TABLE OF CONTENTS

1	A new explanation of Europe's long run development . . . . .	1
2	Oscillations between the world of power and the world of law . . . . .	7
3	Plan of the study . . . . .	17
<b>I</b>	<b>Critical tests</b>	<b>34</b>
4	A framework for critical tests based on Lakatosian principles . . . . .	35
5	Critical tests (I): Theories of Europe's development at the international level . . . . .	65
6	Critical tests (II): Theories of Europe's development at the domestic level	87
<b>II</b>	<b>Explaining history</b>	<b>112</b>
7	Rationality, equilibrium and optimality in long-run history: Two views	113
8	Historical explanation <i>versus</i> statistical explanation . . . . .	162
9	Inference strategies for path dependent histories . . . . .	181
<b>III</b>	<b>Contractual Realism: Causal links between power and law</b>	<b>200</b>
10	The world the power, conflict and high-cost deterrence . . . . .	201
11	The world of law, justice and low-cost deterrence . . . . .	224
12	Hard-core assumptions, causal claims and observable implications . . . .	244

<b>IV</b>	<b>The path of power and the path of law in domestic politics</b>	<b>258</b>
13	From the world of power to the world of law in Medieval England . . . .	259
14	From the world of law to the world of power in Medieval Germany . . .	292
15	Worlds of power and law in the democratization battles of the 19th century	320
<b>V</b>	<b>From the path of law to the path of power in international politics: The outbreak of hegemonic wars</b>	<b>326</b>
16	Theories of hegemonic war: A review and critique . . . . .	327
17	The Contractual Realist explanation of hegemonic wars . . . . .	418
18	International security competition and the motive of income loss avoidance . . . . .	437
19	Observable implications of the hypothesis of income loss avoidance . . .	478
20	Case study: The outbreak of the Thirty Years War . . . . .	504
21	Case study: The outbreak of the French Revolutionary Wars . . . . .	556
22	Case study: The outbreak of World War II . . . . .	630
<b>VI</b>	<b>The path of power unleashed: Grand strategy in hegemonic wars</b>	<b>650</b>
23	Theories of grand strategy in hegemonic wars: A review and critique .	651
24	The Contractual Realist theory of grand strategy in hegemonic wars . .	669
25	Case study: Grand strategy in the Thirty Years War . . . . .	685

26 Case study: Grand strategy in the French Revolutionary Wars . . . . .	694
 VII The path of power at home: Domestic politics in hegemonic wars	 704
27 A Contractual Realist theory of domestic politics in a unitary actor at war	705
28 Case study: Domestic politics in Austria in the Thirty Years War . . . .	713
29 Case study: Domestic politics in Japan in World War II . . . . .	717
 VIII From the path of power back to the path of law: Settlements of hegemonic wars	 727
30 Stability and instability in hegemonic war settlements: Theory and cases	728
 IX Contractual Realism resolves the inter-paradigm debates	 749
31 Contractual Realism explains the causal drivers of the conventional paradigms	750
 X Appendices	 788
32 A formal model of the world of power, conflict and high cost deterrence	789
33 A formal model of oscillations between the world of law and the world of power . . . . .	812
34 A formal model of the world of law, justice and low cost deterrence . .	820
35 A formal model of international security competition and hegemonic war	853

36 A formal model of how enforcement induces a competition in raising relative power to punish and evade punishment . . . . .	870
37 Original causal assumptions of this study: Quotations from my earliest studies of the topic . . . . .	890

## List of Figures

15.1	Dispute between state 1 and state 2 over income	474
15.2	Deterrent war between state 1 and state 2 over a dispute	475
15.3	Disputes with small and large stakes	476
15.4	Impact of spillover problem on army size and incomes	477
33.1	The contractual anchor shifting game with two technologies	819
34.1	The contractual anchor shifting game with one technology	852
35.1	State 1's total income and contributions of factors $T$ and $L$	866
35.2	State 1's marginal products of military and civilian labor	866
35.3	State 1's income from launching a sudden attack	867
35.4	State $i$ 's productivity deflator	867
35.5	Number of men effectively entering civilian workforce	868
35.6	State 1's incomes from peace and from launching an attack	868
35.7	Maximum stable transfer fraction	869
36.1	Equilibrium detection probability	881
36.2	State 1's income from the detection equilibrium	882
36.3	State 2's income from the detection equilibrium	883
36.4	State 1's incomes from detection and private max equilibria	884
36.5	State 2's incomes from detection and private max equilibria	885
36.6	State 1's incomes from detection and social max equilibria	886
36.7	State 2's incomes from detection and social max equilibria	887
36.8	Sum of two states' incomes in detection and social max equilibria	888
36.9	Joint income loss from the detection equilibrium	889

## List of Tables

2.1	Cases with high and low LIM capacity	16
9.1	Cases with high and low LIM capacity	257
12.1	Conditions for democratization in 19th century Europe	321
13.1	Key facts about Germany in World War I	370
13.2	Key facts about Germany in World Wars I and II	373
13.3	Hegemonic wars as anomalies for power shift theories	385

## ACKNOWLEDGMENTS

This study began as a doctoral dissertation more than two decades ago. After some long leaves of absence from graduate school, I am now completing this study as a doctoral dissertation. It has gone through many changes over the years. The core assumptions of the theory have remained the same though. The theory makes three central assumptions. First, actors divide valuable assets according to anchor contracts – agreements whose terms are invariant to shifts in relative power. Second, the main cause of conflicts is legal incompatibility disputes. Third, legal incompatibility disputes provoke competitions in raising relative power and these competitions cause conflict some of the time and instability all of the time. These assumptions have remained the same since my earliest studies of the topic.

Four aspects of this study have changed over the years. First, the number of observable implications derived from the theory has expanded. Second, the number of historical cases explained by the theory has expanded. Third, my understanding of how to conceptualize and measure the cause and effect variables has evolved. Fourth, my understanding of the proper scope conditions of the theory has evolved.

Among political scientists and political economists who are concerned about methodology, there has been some debate about whether and how it can be legitimate to modify a theory in response to anomalies. The main source on this question remains the study by Imre Lakatos entitled "The methodology of scientific research programs". According to that study, it is legitimate to amend a theory in response to anomalies as long as the theory grows in explanatory power as a result. If the amended theory explains new cases it never explained before – as well as the anomalous cases it was intended to explain – then it is scientifically legitimate. But if the amended theory only explains the anomalous cases it was intended to explain (and the previous cases it already explained), then it is scientifically illegitimate. Using these standards, I claim that the theory presented in this study outperforms previous theories in the literature on these subjects. The purpose of this book is to lay out the methodology, theory and evidence that support this claim.

I want to thank the people who supported me in the production of this study. Within academia, I want to thank the chair and members of my doctoral dissertation committee



at UCLA. I thank Arthur Stein, the chair of the committee, for providing encouragement and advice at several stages in the development of my thinking about this research agenda. At the outset he endorsed my goal of using the theory to try to resolve the inter-paradigm debates. A few years later he advised me to extend the theory to explain the similarities and differences between the three types of domestic crisis that caused hegemonic wars in European history: secession, revolution and depression crises. He also advised me to look and see if the theory might explain the cases of Britain's imperial wars in the 19th century. Some years later he endorsed my goal of using the theory to explain cases of conflict, settlement and development at the domestic level. I thank Ron Rogowski for advising me to look and see if there might be other conflicts and settlements – beyond hegemonic wars – that are explained by the theory. I thank Ron Rogowski and Nico Voigtlander for advising me to use the Cobb-Douglas functional form in my model of conflict. That suggestion generated observable implications that explained a number of additional historical cases beyond what I was expecting. I thank Marc Trachtenberg for advising me to figure out the connections between my theory and the balance of power. I thank Robert Trager for advising me to distinguish empirically between changes in force size due to the shift from peace to war and changes in force size due to shifts in the costs and benefits of using force. All remaining errors are my responsibility alone.

I am also grateful for the lessons I learned in the courses of several professors in the departments of political science and economics at UCLA. Exposure to their way of thinking about political economy had a deep impact on my thinking about this research. In this regard I thank: Harold Demsetz, Jeff Frieden, Tom Hubbard, Ben Klein, Ron Rogowski, Richard Rosecrance, Arthur Stein, George Tsebelis and the late Michael Wallerstein. I am grateful for the lessons I learned about empirical analysis in Charles Ragin's course at Northwestern University. I am grateful for the lessons I learned by reading the methodology articles of Richard Langlois of the University of Connecticut. And I am grateful for the lessons I learned by reading Ronald Findlay's article on conflict and trade-offs between the factors of production. These lessons played a major role in shaping this study.

I am particularly grateful for the help I received from the late Jack Hirshleifer. I audited his course on conflict modeling. He wrote a letter of recommendation for my successful

fellowship application in 2002. When he saw that my research was based in part on one of his conflict models, he kindly offered to serve on my dissertation committee. He also made time to meet with me on a few occasions to discuss my modeling plans. In particular, he encouraged me to formalize my model of how legal incompatibility disputes cause conflicts and competitions in raising relative power. For anyone familiar with Jack Hirshleifer's work on conflict models, the impact of his work on this study will be evident.

Beyond academia, I want to thank my friends and family for their support and encouragement. I thank my friends Cliff Brown, Steve Brown, Adam Goldman, Sharon Goldman, and Josiah Spurr. I also thank my sisters and their families, our mother, the memory of our father, and the memory of our grandparents. Without their support and encouragement I could not have completed the study.

## VITA

### Education

- 1986 BA, political science, Duke University
- 1990 MA, political science, Northwestern University
- 2010 MA, political science, UCLA
- 2010 CPhil, political science, UCLA

### Fellowships

- 1991-92 Dean's Fellowship, UCLA
- 2002-03 Dissertation Research Fellowship, UC Institute for Global Conflict and Cooperation; award declined due to health issue requiring leave of absence

### Employment

- 1990-91 research assistant, Brookings Institution, Washington DC
- 1991-95 research assistant, Center for American Politics and Public Policy, UCLA
- 1992-93 teaching assistant, Department of Science, UCLA
- 1993-97 teaching associate, Department of Political Science, UCLA
- 2010-19 work-study position, UCLA Student Call Center, UCLA Fund

# CHAPTER 1

## A new explanation of Europe's long run development

A thousand years ago Europe was mired in feudal anarchy. Conflict was pervasive. Society was antagonistic. The economy was stagnant. Government was unresponsive. By the late 20th century Europe was a land of peace, prosperity and democracy. How did Europe make this transition? The question is significant for both academic and practical reasons. Scholars still debate the evolution of the modern state in Europe over the last thousand years and the development of international institutions in Europe over the last 500 years. Practitioners currently debate the conditions for improving state capacity in developing countries and deepening international cooperation among all countries.

In this study, I argue that Europe's long transition out of feudal anarchy was the consequence of a series of conflicts and settlements caused by a single problem. This problem arose repeatedly at both the domestic and international levels. Each time it arose and caused conflict, the result was political breakdown, social capital depletion, economic misgovernance and lower economic growth. Each time the problem was solved and the period of conflict ended, the result was a return to political stability, social capital formation, good economic governance and higher economic growth. To explain these oscillations in political, social and economic development, I develop a unified theory of conflict and settlement. It is unified in the sense that it explains conflicts and settlements of many different kinds at both the domestic and international levels. The theory's implications are corroborated by the historical evidence presented in the case study chapters. The theory and the evidence, taken together, provide a new explanation of Europe's transition from feudal anarchy to peace, prosperity and democracy.

# 1. The world of power and the world of law

Throughout the last thousand years of European history, actors repeatedly reached the point where they recognized that competitions in raising relative power were making everyone worse off. To stop the competitions, they agreed to divide the assets over which they were competing according to terms that were *independent* of relative power. They formed contracts whose terms were tied to anchors in the world that did not move. This device eliminated each actor's incentive to invest resources in increasing his relative power, for such investments would not shift the contract terms in his favor. Relative power continued to shift due to random factors beyond anyone's control. But no actor had an incentive to make *intentional* investments in his power at another actor's expense. These anchor contracts stopped the competitions in raising relative power.<sup>1</sup> Actors also needed a way to enforce the anchor contracts that did not rely primarily on *power* based enforcement tools. For such tools could reignite the competitions in raising relative power. Hence actors developed *anchor* based enforcement tools.<sup>2</sup>

At the times and places where actors relied on anchor contracts and anchor based enforcement, they achieved political stability, good economic governance and social capital formation. These factors promoted development. The reliance on anchor contracts created a vulnerability, however. Since the contract terms were independent of relative power, the terms could provoke disputes even when there were no power shifts, no uncertainties about relative power, and no other concerns about relative power. Through European history the single most important cause of disputes was legal incompatibility problems. A *legal incompatibility* was a situation where one term of a contract or law gave one actor the right to an asset, but another term of the *same* contract or law gave another actor the right to the same asset. When such situations arose, neither actor could concede the asset in question to the other actor because that would give him an incentive to create legal incompatibilities

---

<sup>1</sup>The first social scientist to develop the idea of an anchor contract was Nicholas Rowe, though he did not call it by that name (Rowe 1989, 1990). However, he did not derive the observable implications of his idea or test the idea on empirical evidence. I undertake these tasks in this study. I first used Rowe's idea to explain key events and developments in European history in my early studies of this topic (Koppel 1997-2002; 2006).

<sup>2</sup>While Rowe's study is seminal, he neglected to explain how anchor contracts are enforced without reliance on power-based enforcement tools that would reignite competitions in raising relative power. I first identified and characterized the mechanism of anchor based enforcement in Koppel (2011).

intentionally in the future for his own gain. When each actor demanded his right to the asset and stood firm over the demand, the result was a dispute.

Two outcomes were possible in such disputes. At the times and places where actors had a high capacity for *legal incompatibility management* (LIM), the disputes were resolved smoothly and development proceeded. At the times and places where actors had low LIM capacity, by contrast, the disputes caused instability, misgovernance and social capital depletion. These outcomes hindered development.<sup>3</sup>

In the long run, Europe oscillated between these two worlds. Where actors had high LIM capacity, they lived in a world of law that enjoyed political stability, good economic governance, social capital formation and long-run development. Where actors had low LIM capacity, by contrast, they lived in a world of power that suffered political instability, misgovernance, social capital depletion and underdevelopment. These oscillations occurred at both the domestic and international levels. A single causal model explains the evidence on both levels. The model thus provides a unified explanation of Europe's development.

## **2. New answers to old questions**

The theory developed in this study, which I call Contractual Realism, differs from the three main paradigms in the literature on conflict, cooperation and long-run transformation. Contractual Realism identifies a single factor that drives the three variables that are thought to be independent drivers in the main paradigms: power politics, institutional design, and social identity roles. In this sense Contractual Realism resolves the debates between the three conventional paradigms.

### **2.1 The roots of power politics**

The main paradigm in the literature for explaining conflict is Conventional Realism. According to this paradigm, actors divide valuable assets according to relative power at all times –

---

<sup>3</sup>Rowe's study neglected to explain why some disputes cause conflicts while other disputes are resolved without conflicts. In my early studies of this topic (Koppel 1997-2002; 2006), I identified legal incompatibility problems as a cause of conflict and instability and characterized some of the reasons why stability depends on the capacity to manage legal incompatibility problems.

not only in war but in peacetime as well. The main cause of conflict is assumed to be shifts in power in peacetime (and uncertainty about power shifts in peacetime).

Contractual Realism takes a different view. It assumes that actors divide assets according to relative power during conflicts and competitions in raising relative power. But at other times they divide assets according to anchor contracts – agreements whose terms are invariant to shifts in relative power. At those times they live in a world of law rather than a world of power. The main cause of conflict is assumed to be disputes over legal incompatibilities. Once such a dispute becomes intractable, actors can no longer divide assets by anchor contracts. So they revert to dividing assets by relative power and that triggers a competition in raising relative power. In short, they revert to the world of power politics. There are two sub-equilibriums in this world. There is the sub-equilibrium of *conflict* – defined as unilateral asset seizures that elicit an immediate response of a reciprocal or retributive kind. And there is the sub-equilibrium of *instability* – defined as unilateral asset seizures that do not elicit an immediate response of a reciprocal or retributive kind. As long as actors remain in the world of power politics, they switch back and forth between these two sub-equilibriums. From this perspective, Conventional Realism only explains why actors switch from the instability sub-equilibrium to the conflict sub-equilibrium within the world of power. It does not explain why the actors reverted from the world of law to the world of power in the first place. Contractual Realism explains this. According to Contractual Realism, reversion to the world of power is caused by legal incompatibility disputes in an institutional environment with a low capacity for Legal Incompatibility Management (LIM). Through European history, the largest and longest oscillations between the world of law and the world of power, at both the domestic and international levels, were caused by this factor.

## **2.2 The roots of good governance and misgovernance**

The main paradigm in the literature for explaining good governance and misgovernance is Conventional Institutionalism. According to this paradigm, there are three main stories about the institutional design that produces good governance and the institutional design flaw that produces misgovernance. In the first story, good governance requires that one actor be strong enough to enforce law and order among the other actors in the system. (In

the domestic system the strong enforcer is the state. In the international system the strong enforcer is a hegemonic state.) According to this story, if the other actors in the system are too strong relative to the enforcer, the result is over-*decentralization* of power, institutional instability and misgovernance. In the second story, good governance requires that the other actors be strong enough to keep the enforcer in check. If they are not strong enough, then the result is over-centralization of power, institutional rigidity and misgovernance. In the third story, good governance requires just the right balance of power between the enforcer and the other actors in the system. If the balance tilts too much in the enforcer's direction, the result is abuse of power, rigidity and misgovernance. If the balance tilts too much in the others' direction, the result is a vacuum of power, instability and misgovernance.

Contractual Realism takes a different view. Good governance does not require a particular distribution of power between the enforcer and the other actors in the system. Good governance requires anchor contracts, anchor based enforcement methods, and the capacity for Legal Incompatibility Management (LIM). When these institutions of anchor based governance are present, there will be a relatively low level of investment in power-based enforcement tools by *both* the enforcer and the other actors in the system. Conversely, when anchor based governance institutions are absent, there will be a relatively high level of investment in power-based enforcement assets by *both* the enforcer and the others. In this case the resulting institutional design will depend on the type of disputes that are most salient due to the absence of anchor based governance. When the salient disputes are about *economic* exclusion, the system veers in the direction of over-centralization of power and institutional rigidity, resulting in misgovernance. When the salient disputes are about *political* exclusion, the system veers in the direction of over-decentralization of power and institutional instability, resulting in misgovernance. In either case, the problem of misgovernance is not due to the wrong balance of power between the enforcer and the others. It is due to the absence of anchor based governance institutions. Through European history, it was the lack or loss of anchor based governance that generated these two structures of misgovernance – at both the domestic and international levels.



### **2.3 The roots of divisive identity politics**

The main paradigm in the literature for explaining the long-run transformation of actors and systems is Conventional Constructivism. According to this paradigm, the main obstacle to long-run development is social, political and economic stratification that divides the system into dominant social groups and subjected social groups. The divide between the dominant and subjected groups is exacerbated by cleavages between different tribal, ethnic, religious and national identity groups. These cleavages and inequalities then generate divisive identity politics.

Contractual Realism takes a different view. Divisive identity politics is driven by the absence or failure of anchor based governance institutions. Through European history, the decay of anchor based governance led to the emergence of divisive identity politics among previously cohesive groups. Conversely, the development of anchor based governance enabled actors in divided societies to cooperate and achieve social capital formation across identity lines that were previously divisive.

## **3. An opportunity for critical tests**

Contractual Realism unites the strands of European history into a single narrative based on one causal factor: the presence or absence of anchor based governance. The theory thus explains the connections between arenas that are usually thought to be separate – the political, economic and social arenas; the international and domestic arenas; the arena of violent conflict and the arena of nonviolent conflict. This approach transcends disciplinary and theoretical boundaries in the social sciences in a way that may seem counterintuitive to some. The test of this approach is in its explanatory power compared to other theories of conflict, settlement and development in European history. The aim of this book is to present the methodology, theory and evidence that are needed to conduct comparative tests of its explanatory power.

## CHAPTER 2

# Oscillations between the world of power and the world of law

European history can be viewed as a series of oscillations between conflict and settlement. These cycles drove Europe's political, social and economic development at both the domestic and international levels. The periods of conflict were associated with political breakdown, social capital depletion, economic misgovernance and lower economic growth. The periods of settlement were associated with political stability, social capital formation, good economic governance and higher economic growth. To explain these patterns of variation, I propose a theory that identifies the cause of the conflicts and the condition for their stable settlement.

### 1. A contractual approach to conflict and settlement

In Europe as elsewhere, people faced the problem of how to divide valuable assets among themselves without conflict. All kinds of actors faced this problem: individuals, groups and states. They faced this problem with all sorts of assets: economic, social and political. They faced this problem at all levels: local, regional, national and international. It was ultimately a *contracting* problem, because there were two basic types of contract that actors could use to divide assets amongst themselves. One type would guarantee stability and security in the allocation of assets across actors. The other type would not. From this contractual perspective, European history is a long story of oscillations between periods when the first type of contract was predominant within a particular state or among states, and other periods when the second type of contract was predominant.

## 1.1 Two types of contract

The first type of contract is based on terms that shift with changes in the relative power of the contracting parties. This type of contract causes instability and insecurity in the allocation of assets across actors. Since the contract terms shift with changes in the actors' relative power, each actor has an incentive to invest resources in raising his relative power to shift the terms in his favor. The result is competitions in raising relative power that are destabilizing and mutually self-defeating. The competitions are destabilizing because every time an actor sees others increasing their power, he is driven to increase his power in response, even if it means seizing assets from other actors unilaterally or imposing costs on them against their will to force them to relinquish shares of their assets. The competitions are also mutually self-defeating because each actor incurs the cost of investing in power assets, yet no actor succeeds in raising his power relative to the others. Such competitions are thus a waste of everyone's resources – a deadweight cost in the language of economics. To avoid such wasteful competitions, actors form a second type of contract – a contract whose terms do *not* shift with changes in the relative power of the contracting parties.

In this second type of contract, the terms of the contract are linked to objects in the real world that do not move even when there is a shift in relative power among the contracting parties. I refer to this type of contract as an *anchor contract*, because the real world objects that do not move are, in effect, anchors. Such a contract eliminates the incentive to invest resources in raising relative power. For even if a person makes such investments, they will not shift the contract terms in his favor. The terms are tied to anchors that remain stationary despite changes in relative power.

Since the world is changing all the time due to random factors beyond anyone's control, the anchors also move due to random factors beyond anyone's control. Whenever this happens one of the contractors receives a "random variable benefit" from the other one. For example, suppose the borderline between two agricultural estates is a river, and the property titles state that the borderline is "the river". The course of the river may change due to random factors beyond either landowner's control (such as a storm, fire or landslide). If this happens, then landowner A will receive a strip of land from landowner B – the strip between the old river course and the new river course. This is landowner A's random variable benefit.

Landowner B can allow landowner A to receive this benefit as long as B knows that A did not move the river in his own favor intentionally (for example, by shoveling loads of earth into the river’s path upstream). But if landowner B knows that landowner A moved the river intentionally to gain a strip of land, then B can retaliate by withholding A’s random variable benefit *at other times* when the river moves in A’s favor due to random forces beyond his control. As long as A expects to incur this kind of punishment for moving the river intentionally, he will be deterred from moving it intentionally – even if B lacks a private army to impose punishment by power-based methods (or a third-party backer to impose punishment by power-based methods). I refer to this method of enforcement as *anchor-based enforcement*. It is composed of three elements:

1. **An adjudicative institution.** Its function is to distinguish between intentional and unintentional movements of the anchors.
2. **An enforcement bargain.** Each actor allows the other actor to receive his random variable benefit from *unintentional* movements of the anchors in his favor as long as he acquiesces willingly, without resistance, when found guilty of moving an anchor in his favor *intentionally* for his own gain.<sup>1</sup>
3. **An enforcement threat.** If one actor is certain that the other moved an anchor in his own favor intentionally for his own gain, then the first actor withholds the second actor’s random variable benefit at the other times when an anchor moves in his favor due to random forces beyond his control.

I refer to these elements together as a *justice bargain*. As long as these elements are functioning, actors can rely on anchor-based enforcement. If any of these elements breaks down, however, then actors must revert to power-based enforcement.<sup>2</sup>

## 1.2 Anchor-based enforcement versus power-based enforcement

These two methods of enforcement differ in important respects. First, they differ in the target’s response to being punished. With power-based enforcement, the target of the

---

<sup>1</sup>That is, as long as the guilty party accepts the adverse verdict and gives up his illicit gain from his opportunism without resistance.

<sup>2</sup>I first characterized the mathematics of anchor-based enforcement in an earlier analysis of this topic (Koppel11).

enforcement action may resist if his relative power is high enough. So the enforcer must maintain and deploy power-based tools to be able to impose punishment against the target's resistance if necessary. By contrast, with anchor-based enforcement, the target acquiesces willingly to the punishment – without resistance – so the enforcer does not need to employ power-based enforcement tools to impose punishment against the target's will. Because it does not rely on power based enforcement tools and does not require a conflict, anchor based enforcement does not provoke wasteful competitions in raising relative power.

The second difference is in the costliness of enforcement actions to the enforcer – the costliness of imposing punishments on violators. With anchor-based enforcement, the punishment takes the form of withholding the violator's random variable benefit. The enforcer actually benefits from imposing this punishment, because he gets to keep the random variable benefit for himself. By contrast, with power-based enforcement, the enforcer finds it costly to impose punishment on the violator. The costly element is usually armed force, which is costly to make and use. But non-kinetic tools of power based enforcement are also costly to use. An economic boycott is costly to impose because the boycotters must forego lucrative trading opportunities with the target. A labor strike is costly to impose on management because the workers lose wages during the strike. A lockout is costly to impose on workers because management loses output and profits during the lockout.

These differences between the two methods of enforcement create strong incentives for actors. In light of the cost differential, actors prefer to use anchor-based enforcement when possible. Throughout European history this was the method of choice because it was the least costly method. This method thus saved resources for productive economic investment and promoted higher economic growth rates. However, if any element of the justice bargain breaks down, then actors must revert to power based enforcement. Each actor can continue to use power-based enforcement as long as his cost of maintaining and applying power-based enforcement tools (per week or month) is less than the growth rate of the economic base he secures with those tools. If the cost rate rises above the growth rate, however, then he has limited options. He can lower the cost rate, increase the growth rate, secure financial loans, find allies to support his cause, capitulate temporarily to buy time to rebuild his power base, or fix the justice bargain so he can return to anchor-based enforcement. If he uses none of

these strategies, and simply continues to use power-based enforcement tools even though the cost rate is above the growth rate, then his economic base will continue to shrink. Eventually it will be gone, or to put it another way, the economy will be driven into the ground. This risk creates a strong incentive to use one of the other strategies.

## **2. Irreconcilable disputes and reversion to power-based enforcement**

The main cause of breakdowns in justice bargains is the problem of *legal incompatibility*. This is a situation where one term of a contract gives one actor the right to an asset, but another term of the contract gives another actor the right to the same asset. This problem causes a dispute over the asset. Each actor claims that the other actor is in violation of the contract. Yet each actor expects that if he concedes the asset to the other one, then the other one will create more legal incompatibilities intentionally in the future to gain more concessions. So each actor refuses to agree to any compromise that would allow the other actor to reap a net gain from the dispute (such as a compromise settlement of "splitting the difference" in the dispute). When both actors stand firm in this way, the dispute becomes irreconcilable.

Once the dispute reaches this point, anchor contracts and anchor-based enforcement are no longer an effective method of securing ownership rights to the asset. For the actors are no longer in consensus on the meaning of the anchor contract's terms. So they revert to power-based enforcement to secure their ownership rights. Each actor imposes costs on the other actor against his will using power-based enforcement tools – to punish him for his violation, to deter further violations, and to compel him to relinquish shares of the asset. Each actor resists punishment, however, in the belief that he has done nothing wrong and he must punish the other actor's violation or else he will continue to commit violations *ad infinitum*. The result is a period of instability and conflict that continues as long as the legal incompatibility problem remains unsolved.

Each actor's goal is to impose more punishment costs on the other actor than he will gain in the end from a compromise settlement of "splitting the difference" in the dispute. Once each actor has imposed this much costs on the other one, each one can agree to settle the dispute by splitting the difference. For such a settlement will leave the other one having

incurred more costs in the conflict than he gains from the settlement. So the other one will have no incentive to create such legal incompatibilities intentionally in the future.

This kind of conflict succeeds in deterring actors from creating legal incompatibilities intentionally for their own gain. Yet legal incompatibilities still arise unintentionally due to random factors beyond anyone's control. When this happens, each actor stands firm and refuses to compromise in the belief that he is legally in the right. Each actor interprets the other actor's intentional act of standing firm, however, as evidence that he created the legal incompatibility intentionally for his own gain. So each one concludes that he must wage conflict to deter the other one from creating legal incompatibilities intentionally again in the future. Moreover, the larger is the asset at stake in the current dispute, the longer the conflict needs to last to reach the point where each actor has incurred more costs in the conflict than he gains from the compromise settlement at the end of the conflict.

Unintentional legal incompatibilities cause conflict in this way even if there are no large shifts in relative power between the two actors and no uncertainties about relative power between them. Hence the conflict cannot be prevented simply by stabilizing the power balance between the actors and providing them with full information about relative power. Moreover, the conflict continues as long as the legal incompatibility problem remains unsolved. So the conflict cannot be terminated simply by stabilizing the power balance and providing the actors with full information about relative power.

Conflict can be avoided if there is an institution that has the capacity to identify *unintentional* legal incompatibilities and conceptualize new ways of partitioning rights to disputed assets. Conflict is inevitable, however, if there is no such institution or existing institutions are insufficient because the legal incompatibility is too novel in nature. Thus there are two causes of conflict in this theory:

1. The emergence of a legal incompatibility.
2. The absence or insufficiency of institutions for identifying unintentional legal incompatibilities and conceptualizing new ways of partitioning rights to disputed assets.

For example, in medieval England the courts developed a formal pleading procedure that had the capacity to identify unintentional legal incompatibilities and formulate options for

partitioning asset rights. In medieval Germany, by contrast, the judicial procedures lacked this capacity. In early modern Germany, the Peace of Westphalia was supported by institutions that had this capacity, but the Peace of Augsburg was not. Between 1550 and 1950, seven hegemonic wars erupted between the great powers of Europe. Each conflict was caused by the emergence of a legal incompatibility problem that was so novel in nature that existing institutions did not have the capacity to identify it and conceptualize options for partitioning the disputed asset. Each conflict was terminated by a settlement that attempted to solve the legal incompatibility problem and establish institutions with these capacities. Three of the seven settlements were based on anchor contracts and institutions having these capacities. The other four settlements were not.

### 3. Introducing "Contractual Realism"

I call this theory contractual realism because it identifies the contractual foundations of conflict and settlement. And it is realistic about both the inevitability of conflicts and the incentive to avoid conflicts and competitions in raising relative power through better contracting. This theory also identifies the contractual foundations of power, institutions and shared identity – and demonstrates that they are linked by a single, unified causal mechanism. In this way *contractual realism* resolves the "inter-paradigm debate" and a number of related debates. It thereby sets new research agendas that link numerous literatures previously thought to be independent. The literature review chapter at the beginning of each Part of this study discusses the theoretical debates that contractual realism addresses. Part X of this study explains how contractual realism resolves the inter-paradigm debate.<sup>3</sup>

**The main observable implication of Contractual Realism** The main observable implication of this theory is that European history is expected to exhibit a series of oscillations between two completely different modes of behavior. In periods when actors divided assets by the *contractual method*, we should expect to observe low levels of conflict and spending on power-based enforcement tools, and high rates of economic growth, social capital formation and good governance. In the other periods, when the contractual method failed, we

---

<sup>3</sup>I first wrote that this theory resolves the inter-paradigm debate in Koppel (2000b).



should expect to observe actors dividing assets according to *relative power on the margin* – a completely different mode of behavior. In these periods we should expect to observe high levels of conflict and spending on power-based enforcement tools, low or negative economic growth, social capital depletion and political breakdown. Finally, we should expect that each of the conflict periods ended with the formation of new anchor-term contracts, justice bargains and adjudicative institutions.

The clearest instances of this pattern of oscillation were the great power wars in Europe between 1550 and 1950. Since the Renaissance a major war has broken out between the largest states in Europe every 50 to 100 years. In each war one great power tried to conquer neighboring states and establish hegemony over the region. A coalition of the other great powers banded together to oppose the conqueror and force it to disgorge its conquests. Peace was then maintained for another 50 or 100 years – until the next conqueror arose. These conflicts are called “hegemonic wars” because in each case one great power sought to gain hegemony over the continent and the other great powers resisted.

These wars shaped Europe’s long-run development on many levels. On the political level, these wars forced governments to increase their domestic capacities in many ways. The final settlement of each conflict required states to develop treaty law and international institutions beyond levels seen previously in Europe or elsewhere. On the economic level, these conflicts destroyed more lives and wasted more economic resources than any other wars in human history. Yet at the same time these conflicts forced states to develop new technologies to wage war effectively, promote recovery and improve competitiveness in a challenging world. On the social level, each war created deep social enmities within and across countries, thereby destroying social capital that had accumulated over centuries. In the aftermath of each war states had to develop new means of community building and social capital formation to heal the social rifts. In all of these ways and more, hegemonic wars shaped Europe’s long-run development as profoundly as domestic politics did.

Previous studies attributed these wars to shifts in relative power, uncertainty about relative power or commitment problems related to relative power. Yet the conventional wisdom is contradicted by the historical evidence and suffers from deep logical problems. In this study I provide a different explanation of why these wars erupted and why they

shaped Europe's long-run development so profoundly. I show that these wars were caused by legal incompatibility disputes that confounded anchor based governance and forced actors to descend into the world of power. These wars were settled, in turn, by the formation of new anchor based governance systems that enabled actors to return to the world of law.

Oscillation between the world of power and the world of law occurred at the domestic level too. In the table on the following page, I list historical cases of oscillation at both the domestic and international levels. The cases are divided into categories based on the main causal variables in this study. The primary causal variable is the capacity for a Legal Incompatibility Management (LIM capacity). Among the cases with low LIM capacity, the key causal variable is the type of legal incompatibility disputes that caused the reversion to the world of power. On one hand there were disputes over economic exclusion. On the other hand there were disputes over political exclusion. The similarities and differences between these two kinds of disputes were evident throughout European history at both the domestic and international levels. The domestic pattern of evidence is similar enough to the international pattern of evidence to deserve explanation. The theory and case studies in this book offer an explanation.

<b>LIM capacity</b>	<b>LOW</b>	<b>LOW</b>	<b>HIGH</b>
Enforcement method	power-based	power-based	anchor-based
Political outcome	instability	instability	stability
Economic outcome	misgovernance	misgovernance	good governance
Social outcome	social capital depletion	social capital depletion	social capital formation
<b>Main cause of disputes</b>	<b>political exclusion</b>	<b>economic exclusion</b>	<b>neither</b>
Swings in relative coalition size	frequent, narrow swings	infrequent, wide swings	neither
Resulting distribution of relative power	unstable balance	unipolar predominance	neither
Empirical cases in domestic politics: 1000-1500	Germany	Italy	England France
Empirical cases in domestic politics: 1500-1800	France	Netherlands	England Germany (Prussia)
Empirical cases in International politics: 1500-1950	Peace of Augsburg (1600-18) Thirty Years War (1618-48) Nine Years War (1689-97) War of Spanish Succession (1701-13) Peace of Utrecht (1713-45) French Rev. Wars (1792-99) World War I (1914-18)	Peace of Augsburg (1555-99) Napoleonic Wars (1800-14) Peace of Versailles (1919-39) World War II (1939-45)	Bretton Woods system (1945-) Peace of Westphalia (1648-89) Peace of Vienna (1815-53)

## CHAPTER 3

### Plan of the study

This chapter is divided into four sections. The first section provides definitions of key terms. The second section provides a thematic outline of the study. I outline the main claims of the study and explain how each claim advances the argument of the study. The third section provides a summary of the empirical evidence for these claims. The fourth section provides a sequential outline of the study. I summarize each part of the study from one to ten and explain how it advances the claims and overall argument of the study.

#### 1. Definitions of key terms

**Asset.** A material resource or contractual right that generates an income stream over time. This includes both economic and political assets. Economic assets include land and market entry rights. Political assets include the right to exercise political authority and the right to participate in political decisions.

**Anchor contract (or fixed term contract).** An agreement whose terms are invariant to shifts in relative power between the contracting parties. The terms are tied to anchors in the world that do not move even when relative power shifts among the parties. (For example, property titles divide land according to terms that are tied to features of the landscape that do not move. A constitution distributes political authority and participation rights according to terms that are tied to features of the institutional landscape that do not move.) An anchor contract may be an explicit contract or an implicit contract. Either way, the defining feature is that there is a distribution of rights and obligations between the contracting parties and mutual expectations of adherence to that distribution.

**Variable term contract.** An agreement whose terms vary with shifts in relative power between the contracting parties. For example, the Peace of Augsburg distributed political authority according to a term that said the ruler of each territory determines the religion of his subjects. The identity of each territory's ruler changed with shifts in relative power within and between territories.

**Competition in raising relative power.** An unrestricted scramble to amass and deploy the capacity to impose costs on other actors. The competition is measured by the strategies that actors employ to impose costs on each other. (These strategies are outlined in Chapter 10.)

**Instability.** An equilibrium in which actors commit unilateral asset seizures that do not provoke an immediate response of a reciprocal or retributive kind.

**Conflict.** An equilibrium in which actors commit unilateral asset seizures that provoke an immediate response of a reciprocal or retributive kind.

**Stability.** An equilibrium with no unilateral asset seizures.

**Misgovernance** An equilibrium in which stronger actors violate the rights of weaker actors (in part by committing unilateral asset seizures against them). At the international level, strong states violate the territorial boundaries and political independence of weak states. At the domestic level, elites impose high corruption and tax rates on non-elites.

**Good governance.** An equilibrium in which stronger actors respect the rights of weaker actors.

**Social capital depletion.** An equilibrium in which widespread insecurity of asset rights generates pervasive social distrust. Social capital depletion is measured by three indicators:

- indirect subversion rather than direct dialogue;
- jurisdictional separation; and
- discourses of social disapprobation.

**Social capital formation.** An equilibrium in which widespread security of asset rights generates pervasive social trust. Social capital formation is measured by three indicators:

- direct dialogue rather than indirect subversion;
- jurisdictional integration; and
- discourses of social approbation.

## 2. Thematic organization of the study

The overall argument of the study is composed of eight causal claims. The first four claims concern the oscillation from the world of law to the world of power. The second four claims, which are mirror images of the first four, concern the return oscillation from the world of power to the world of law.

### 2.1 Claims about the oscillation from the world of law to the world of power

**Claim 1a.** The absence or failure of anchor contracts causes costly competitions in raising relative power. Anchor contracts are *absent* when settlements are based on variable term contracts rather than fixed term (anchor) contracts. Anchor contracts *fail* when there are legal incompatibility disputes in an environment with a low capacity for Legal Incompatibility Management (LIM).

**Claim 1b.** The competition for power causes conflict. Conflict is simply a part of the competition for power – one of many ways that actors compete for power. While conflicts are tragic, analytically, the best way to understand the causation of conflict is to understand the causation of competitions in raising relative power. They are caused by the absence or failure of anchor contracts (claim 1a).

**Claim 1c.** The competition for power causes instability, misgovernance and social capital depletion. To compete for power, actors commit unilateral asset seizures – the definition of *instability*. Strong actors commit seizures against weak actors – the definition of *misgovernance*. Both patterns cause widespread insecurity and pervasive

distrust – the definition of *social capital depletion*.

**Claim 1d.** The competition for power produces a characteristic structure of misgovernance that depends on the type of disputes that are most salient. When the salient disputes are over *political exclusion*, the equilibrium is a series of unstable balances of power between the opposing sides. When the salient disputes are over *economic exclusion*, the equilibrium is a stable preponderance of power on one side or the other.

## 2.2 Claims about the oscillation from the world of power to the world of law

**Claim 2a.** Anchor contracts stop the competitions in raising relative power. Because the contract terms are invariant to shifts in relative power, actors have less incentive to invest in power to shift the terms in their favor.

**Claim 2b.** Settlements based on anchor contracts can terminate conflicts in a stable way. By contrast, settlements based on variable term contracts can cause a temporary cessation of conflict. But the competition for power will continue and the settlement will be unstable.

**Claim 2c.** Anchor based governance is the main condition for stability, good governance and social capital formation. Anchor contracts generate *stability* in the distribution of assets among actors, so the distribution is invariant to shifts or differences in relative power. Strong actors respect the rights of weak actors – the definition of *good governance*. Both patterns generate security of asset ownership and trust that stronger actors will not take advantage of power disparities – the definition of *social capital formation*.

**Claim 2d.** Anchor based governance enables actors to avoid both structures of misgovernance – unstable balances of power and stable preponderances of power.

### **3. Outline of the empirical evidence that supports these claims**

#### **3.1 Evidence at the international level for oscillations from the world of law to the world of power**

**Claim 1a.** There were seven hegemonic wars in Europe between 1550 and 1950. Each war was caused by a legal incompatibility dispute in an environment with a low capacity for Legal Incompatibility Management (LIM). Once each war began, states descended into an unrestricted competition in raising relative power.

**Claim 1b.** To compete for power, states employed five main strategies (see Chapter 10), including the commission of unilateral asset seizures against each other.

**Claim 1c.** Each war was characterized by pervasive instability, misgovernance and social capital depletion.

**Claim 1d.** In the wars caused by *political* exclusion disputes – secessions and revolutions – there were unstable balances of power. In the wars caused by *economic* exclusion disputes – trade tariffs and trade embargoes – there were stable preponderances of power.

#### **3.2 Evidence at the international level for oscillations from the world of power to the world of law**

**Claim 2a.** Of the seven peace settlements, three settlements were based on anchor contracts. The other four settlements were based on variable term contracts.

**Claim 2b.** The settlements based on anchor contracts stopped the conflicts *and* the competitions in raising relative power. The settlements based on variable term contracts stopped the conflicts, but failed to stop the competitions in raising relative power.

**Claim 2c.** During the settlement periods based on anchor contracts, there was stability, good governance and social capital formation. During the settlement periods



based on variable term contracts, there was instability, misgovernance and social capital depletion.

**Claim 2d.** During the settlement periods based on anchor contracts, actors avoided both structures of misgovernance – unstable balances of power and stable preponderances of power. During the settlement periods based on variable term contracts, both structures of misgovernance continued.

### **3.3 Evidence at the domestic level for oscillations from the world of law to the world of power**

#### **3.3.1 Medieval Period**

**Claim 1a.** In the medieval period, Germany and northern Italy had low capacities for legal incompatibility management. Both regions experienced pervasive competitions in raising relative power.

**Claim 1b.** Both regions experienced pervasive conflict.

**Claim 1c.** Both regions experienced instability, misgovernance and social capital depletion.

**Claim 1d.** In Germany the main disputes were over political exclusion. So the region experienced unstable balances of power. In Italy the main disputes were over economic exclusion. So the region experienced stable preponderances of power in each city-state.

#### **3.3.2 Early Modern Period**

**Claim 1a.** In the early modern period, France and the Netherlands had low capacities for legal incompatibility management. Both states experienced pervasive domestic competitions in raising relative power.

**Claim 1b.** Both states experienced pervasive domestic conflict.

**Claim 1c.** Both states experienced instability, misgovernance and social capital depletion.

**Claim 1d.** In France the main disputes *among elites* were over political exclusion. So there were unstable balances of power between elite factions, institutions and regions. The main disputes *between elites and non-elites* were over economic exclusion. So they experienced stable preponderances of power on one side or the other. In the Netherlands the main disputes were over economic exclusion. So the state experienced stable preponderances of power.

### 3.3.3 Nineteenth Century

**Claim 1a.** In the nineteenth century, Germany and Spain experienced domestic legal incompatibility disputes that went unresolved through the century. Both states experienced domestic competitions in raising relative power through the century.

**Claim 1b.** Both states experienced domestic political conflicts that prevented democratization.

**Claim 1c.** Both states experienced instability, misgovernance and social capital depletion.

**Claim 1d.** In Germany the main disputes were over political exclusion. So the state experienced unstable balances of power domestically. In Spain the main disputes were over economic exclusion. So the state experienced stable preponderances of power domestically.

## 3.4 Evidence at the domestic level for oscillations from the world of power to the world of law

### 3.4.1 Medieval Period

**Claim 2a.** In the medieval period, England's legal system was based on anchor contracts and a high capacity for legal incompatibility management. The state experienced relatively few domestic competitions in raising relative power (compared to medieval Germany and Italy).

**Claim 2b.** England experienced relatively few domestic conflicts (compared to Ger-

many and Italy).

**Claim 2c.** England experienced relatively high levels of stability, good governance and social capital formation (compared to Germany and Italy).

**Claim 2d.** Actors in England largely avoided the structures of misgovernance that plagued Germany and Italy – unstable balances of power and stable preponderances of power.

### 3.4.2 Early Modern Period

**Claim 2a.** In the early modern period, Prussia's legal, political and administrative systems were based on anchor contracts and a high capacity for legal incompatibility management. The state experienced relatively few competitions in raising relative power (compared to early modern France and the Netherlands).

**Claim 2b.** Prussia experienced relatively few domestic conflicts (compared to France and the Netherlands).

**Claim 2c.** Prussia experienced relatively high levels of stability, good governance and social capital formation (compared to France and the Netherlands).

**Claim 2d.** Actors in Prussia largely avoided the structures of misgovernance that plagued France and the Netherlands – unstable balances of power and stable preponderances of power.

### 3.4.3 Nineteenth Century

**Claim 2a.** In the nineteenth century, England and France experienced domestic legal incompatibility disputes that were resolved. Both states experienced domestic competitions in raising relative power during the disputes. But the competitions slowed down significantly after the disputes were resolved.

**Claim 2b.** Both states experienced movements toward democratization after their domestic disputes were resolved.

**Claim 2c.** Both states experienced relatively high levels of stability, good governance and social capital formation (compared to Germany and Spain).

**Claim 2d.** Both states largely avoided the structures of misgovernance that plagued Germany and Spain – unstable balances of power and stable preponderances of power.

## 4. Sequential organization of the study

### 4.1 Part I

This part outlines the critical tests that are conducted in the study. A critical test is a comparison between two theories to determine which theory has more explanatory power, which theory has more anomalies, and which theory handles its anomalies more scientifically. Using these standards, I claim that the theory presented in this study, Contractual Realism, outperforms previous theories in the literature on conflict, settlement and long-run development in Europe. To enable evaluation of this claim, I employ the methodology of critical tests outlined by Imre Lakatos in his study "The methodology of scientific research programs". I outline this methodology and apply it to the debate between the two dominant paradigms in the field of international relations theory: Conventional Realism and Conventional Institutionalism (chapter 4). I then use this methodology to review and critique previous theories of Europe's development at the international level (chapter 5) and the domestic level (chapter 6).

In chapter 4, I describe how scientific progress is defined and measured by the two dominant paradigms in the field of international relations theory: Conventional Realism and Conventional Institutionalism. I show that they define progress by the methodologies of verificationism and dogmatic falsificationism first outlined by Lakatos. I then summarize his argument for why a theory can never be proven by the verificationist approach or disproven by the dogmatic falsificationist approach. He explains that for this reason scientists turn to the approach of *methodological* falsificationism. They use *methods* to protect a paradigm from false negatives and false positives (e.g discarding a valid theory by mistake and accepting an invalid theory by mistake). To protect a paradigm from false negatives, they use the method of *ceteris paribus* conditions. To protect a paradigm from false positives, they

use the method of "*observational theories*" to turn raw data into "observations" for use in testing theories. I then outline the observational theories that are used by Conventional Realism and Conventional Institutionalism. Conventional Realism employs an observational theory that turns raw data on the outbreak of conflicts into observations about shifts in relative power, uncertainty about relative power and commitment problems related to relative power. Conventional Institutionalism employs an observational theory that turns raw data on international cooperation into observations about costly punishment as an enforcement device. I then summarize Lakatos' argument that scientists tend to retain observational theories even after falsifying evidence accumulates – contrary to the rules of methodological falsificationism which require that the theories should be discarded – because scientific progress is not made by testing a theory against evidence. It is made by testing one theory against another theory in terms of their comparative ability to explain the evidence. I argue that Conventional Realism and Conventional Institutionalism have retained their observational theories despite falsifying evidence, because no one has yet produced a new theory that outperforms them in explanatory power.

I summarize Lakatos' argument that scientific progress occurred when innovators developed a new observational theory and a new substantive theory that together explained (1) what previous paradigms explained, (2) the anomalies they failed to explain, and (3) novel facts they never claimed to explain. The new paradigm developed in this study, Contractual Realism, is composed of a new observational theory and a new substantive theory. The observational theory turns the raw data on the outbreak of conflicts into observations about legal incompatibility disputes. The observational theory also turns the raw data on international cooperation into observations about anchor contracts and anchor based punishment mechanisms.<sup>1</sup> I argue that this combination of a new observational theory and a new substantive theory provides greater explanatory power than existing theories.

In chapters 5 and 6, I review existing theories of conflict, settlement and long-run development in Europe. I show how theorists amended these theories to account for anomalies. And I show that Contractual Realism can explain these anomalies without needing to be amended in such ways.

---

<sup>1</sup>An anchor contract is an agreement whose terms are invariant to shifts in relative power. An anchor based punishment is beneficial for the punisher to impose rather than costly, as with conventional punishments.

## **4.2 Part II**

This part explains the methodological foundations on which this study is based. In chapter 7, I discuss the role of rationality, equilibrium and optimality in long-run history. There are two views of its role in the literature. Methodologists have concluded that the first view is flawed and the second view is more appropriate for a study of long-run history.

In chapter 8, I discuss the difference between historical explanation and statistical explanation. I describe the factors that confound statistical explanation as a method for explaining history. I show that the alternative method of historical explanation can be just as rigorous, in its own way, as statistical explanation. I discuss the elements of the method that I employ in this study to be rigorous.

In chapter 9, I discuss the inference fallacies that occur when statistical methodology is applied to the explanation of path dependent histories. I then outline the inference strategies that I use (within the methodology of historical explanation) to avoid these fallacies.

## **4.3 Part III**

This part presents Contractual Realism – a new theory of conflict, settlement and long-run development.

## **4.4 Part IV**

This part provides a Contractual Realist theory of domestic political, social and economic development in Europe from 1000 to 1900. The theory assumes that domestic development progressed at times and places when actors relied on anchor based governance, including the capacity for Legal Incompatibility Management (LIM). In regions where this condition was met, a positive spiral of development ensued. There were relatively low levels of conflict and spending on power-based tools of enforcement. Measures of social capital formation rose (e.g direct dialogue, jurisdictional integration, social trust and discourses of social approbation). Corruption was reduced, good governance spread, and economic growth increased. By contrast, in regions where anchor based governance was absent or decayed, a negative spiral ensued. There were high levels of conflict and spending on power-based enforcement

tools. Measures of social capital *depletion* rose (e.g. indirect subversion, jurisdictional disintegration, social distrust and discourses of social disapprobation). Corruption increased, good governance lapsed, and economic growth slowed.

## 4.5 Part V

This part provides a Contractual Realist theory that explains the hegemonic wars in Europe between 1550 and 1950. According to this theory, each war was caused by an international spillover problem that descended into deadlock because of a legal incompatibility. One state had developed a new domestic activity that increased its income but, at the same time, generated international spillover effects that decreased another state's income. The first state cited an international law that gave it the right to continue its new domestic activity. The second state cited another international law that gave it the right to be free of the spillover effects and income losses caused by the activity. Each state expected that if it were to capitulate in the dispute over the spillover problem without a fight, then it would incur large income losses. Each state decided to go to war over the dispute to avoid these income losses, regardless of whether relative power was shifting in its favor or against it. Each state continued to wage war as long as the spillover problem and the legal incompatibility remained unresolved. Even if the state had to drop out of the fighting for a while because its relative power had fallen too low, it waited until its power could be recharged and entered the fighting again rather than capitulate in the dispute over the spillover problem and the legal incompatibility.

The theory generates many observable implications that explain many aspects of each state's crisis bargaining strategy and entry into military hostilities:

- the speed of its assessment of the threat of income losses posed by the spillover problem and the legal incompatibility;
- the speed of its decision to issue diplomatic threats in the dispute over the spillover problem and the legal incompatibility;
- the credibility of its diplomatic threats in the eyes of the other state;
- the speed of its decision to wage war over the dispute rather than capitulate;

- the speed of its entry into military hostilities;
- the sustainability of its military action.

These strategies varied from one state to another in each war, depending on whether it was the state producing the international spillovers or the state suffering from them, and from one war to another, depending on the type of international spillover problem that had caused the war. The theory explains these variations well – under the assumption that each state’s decision whether to wage war over the dispute or settle it peacefully was driven by the legal incompatibility *alone*. Hence the successful explanation of this pattern of historical evidence supports the thesis that (1) each war was caused by the international spillover problem and the legal incompatibility, and (2) the eruption of each war marked an oscillation from a peacetime international system – in which states had been dividing assets according to anchor contracts – to a wartime international system in which states could no longer rely on this method because of the legal incompatibility problem – and relied instead on relative power on the margin.

#### 4.6 Part VI

In this part I extend this theory to explain the grand strategies of the same states in these hegemonic wars. The extension assumes that in each war, each state’s *ultimate* goal was to secure its income rights that were threatened by the international spillover problem and the legal incompatibility. The first state sought to secure its right to continue engaging in its new domestic activity that was increasing its income, even though it was producing international spillovers that lowered the second state’s income. In turn, the second state sought to stop its income losses from the spillovers by stopping the first state from engaging in its new domestic activity. Each state’s *immediate* goal was to impose costs on the other state by military means in order to stop it from expanding its share of the income rights in dispute. Consequently, each state’s share of the rights shifted back and forth with increases or decreases in its military power to impose costs on the other state. Since there were no functioning anchor contracts to constrain the allocation of the disputed rights between the two states, the result was an unrestrained competition in raising relative power. In this competition each state employed a combination of domestic resource mobilization, alliance



formation and territorial expansion to increase its power at the other state's expense, not as an end in itself but as a means to the end of securing its income rights in the dispute. Its only higher priority was to keep its domestic economy running fast enough to maintain the wages of its civilian workers (to avoid labor strikes and material shortages) and maintain the tax revenues needed to pay its soldiers (to avoid mutinies). Each state transformed its resource, alliance and territorial strategies completely under the pressures of unrestrained international security competition.

This extension of the theory generates observable implications that explain each state's choices on the main strategic trade-offs faced by states at war:

- the choice whether to increase its power by mobilizing domestic resources or forming alliances;
- the choice whether to form alliances by contracting allies voluntarily or by imposing hegemony on other states coercively;
- the choice whether to form a selective coalition of allies or an all-inclusive coalition of allies;
- the choice whether to seek victory by increasing its own alliance size or decreasing the opponent's alliance size.

These choices varied from one state to another in each war, depending on whether it was the state producing the international spillovers or the state suffering from them, and from one war to another, depending on the type of international spillover problem that had caused the war. The theory explains these variations well – under the assumption that each war was caused by the dispute over income rights provoked by the spillover problem and the legal incompatibility. Hence the successful explanation of this pattern of historical evidence supports the thesis that when states oscillate from a peacetime system to a wartime system, (1) they shift from dividing assets by the contractual method to dividing assets according to relative power on the margin; (2) this induces a competition in raising relative power; (3) the competition forces each state to produce and deploy power by all available means, not as an end in itself but as a means to the end of securing its income rights in the dispute; and (4) the result is an oscillation from a world of stability where asset rights are secured by contractual means to a world of conflict where asset rights are secured by military means.

## 4.7 Part VII

In this part I extend the theory to explain the domestic political changes that occurred in the state suffering the income losses from the international spillovers. After each war began, this state took military action to stop the other state from producing the international spillovers. To take action it produced military power up to the point that the *cost of producing the power* was just equal to the *benefit from using the power* in the war. Although the cost of producing the power was borne by domestic constituents, they were willing to bear the cost because they expected that they would be even worse off if they were to capitulate in the international dispute without a fight and accept the income losses from the international spillovers passively. As the military campaign to stop the spillovers proceeded, however, the domestic constituents incurred *both* the cost of the income losses from the international spillovers *and* the cost of producing military power to stop them. Hence their leaders were highly constrained to organize the production of the military power in a cost-minimizing way, simply to avoid placing even more burdens on the constituents (since that might provoke labor strikes by civilian workers or mutinies by soldiers). By implication, the leaders were constrained to organize the production of military power in a way that would minimize the *sum* of the two costs their constituents were bearing in the war (e.g. the cost of producing military power and the cost of the income losses from the international spillovers). To minimize costs the leaders transformed domestic institutions and policies in many areas.

The domestic changes varied from one war to another depending on the type of international spillover problem that caused the war, and in particular, whether it increased or decreased the domestic resource cost of producing military power. In cases where the spillovers *increased* the domestic resource cost of producing military power, leaders transformed domestic institutions and policies to enable the *government to compensate the domestic constituents* for the material resources they were sacrificing by awarding them with political decentralization at home (to avoid losing their allegiance to the war effort). By contrast, in cases where the spillovers *decreased* the domestic resource cost of producing military power, leaders transformed domestic institutions and policies to enable the *domestic constituents to compensate the government* for reorganizing production by awarding it with political centralization at home (to avoid losing its allegiance to the war policy).

The two kinds of cases differed in a range of areas:

- institutional decentralization *versus* centralization
- political consensualism *versus* authoritarianism
- constituent group entitlements *versus* sacrifices
- capacity privatization *versus* nationalization
- social divisiveness *versus* cohesion
- military expansion through domestic vote buying *versus* military expansion through domestic rallying around the flag
- policy cleavages over the risk of military over-expansion through domestic privatization *versus* policy cleavages over the risk of military over-expansion through domestic delusion.

The theory explains these differences in domestic politics well – under the assumption that each state produced military power up to the point that its marginal cost of producing the power was just equal to its marginal benefit from using the power in the war. Hence the successful explanation of this pattern of evidence supports the thesis that when states switch from a peacetime international system to a wartime system: (1) each state needs to produce a much larger *quantity* of power-based enforcement tools to secure its asset rights from other states; (2) each state bears a much greater cost to secure its asset rights from other states; (3) each state's domestic constituents are willing to bear the additional cost because they expect that they would be even worse off if they were to give up the asset rights without a fight; (4) each state produces as much power as possible given the constraints it faces; (5) each state is constrained by domestic constituents to produce the income-maximizing amount of power, or at least the income loss-minimizing amount; (6) this domestic political constraint forces policy makers to reconstruct domestic institutions to minimize the cost of producing power; (7) the domestic institutions needed to secure a state's assets in international competition *during wartime* are significantly different from the domestic institutions needed to do so *in peacetime*; because (8) states divide assets on the margin of relative power during wartime, whereas they divide assets by the contractual method during peace time.

## **4.8 Part VIII**

In this part I apply the theory to explain variations in the stability of peace settlements following hegemonic wars. The theory implies that each settlement would be stable in the long run only if its terms formed: (1) an anchor contract that resolved the international spillover problem and the legal incompatibility that caused the war, (2) a justice bargain that enabled enforcement of the contract; and (3) rules for adjudicating disputes over the contract in novel situations not anticipated by the framers. I present case studies which show that the settlements of three hegemonic wars met these conditions and were stable as a result, while the settlements of three other hegemonic wars did not meet these conditions and were unstable as a result. This pattern of evidence provides further proof that the main cause of catastrophic wars in European history was the breakdown of anchor contracts and justice bargains. And the main condition for stable settlements was the construction of new anchor contracts and justice bargains.

## **4.9 Part IX**

In this part I explain how the theory developed and tested in this study – Contractual Realism – resolves the inter-paradigm debates. Each of the conventional paradigms has its maintained assumptions. And each one's maintained assumptions are logically inconsistent with the maintained assumptions of the other two paradigms. These inconsistencies persist to this day (despite claims from some quarters that they are overdrawn or nonexistent). By contrast, Contractual Realism provides a single set of logically consistent assumptions that succeeds in making all three of the main points of the three conventional paradigms. It thereby resolves the inter-paradigm debates and provides new answers to a range of explanatory questions based on a single, unified set of causal assumptions.

Part I

# Critical tests

## CHAPTER 4

### A framework for critical tests based on Lakatosian principles

In recent decades theorists of international relations have engaged in deep debate about how to measure scientific progress in the field. The main metric of progress has been the "methodology of scientific research programs" proposed by Lakatos (1970). According to this methodology, progress is made when a scientist discovers a new theory that explains as much as previous theories explained *and more*. The new theory receives additional support if the scientist can identify empirical anomalies that confounded the previous theories – events that contradicted their predictions – and explain those events too. The new theory receives further support if the scientist can identify flaws in the "observational theories" that were used to measure the main variables in the previous (substantive) theories. Finally, the new theory receives conclusive support if the scientist can show that previous theories were only able to grow in explanatory power by increasing the number of *ceteris paribus* conditions attached to their causal claims, whereas the new theory can grow in explanatory power with less *ceteris paribus* conditions. In this chapter and the next two chapters, I will show that when these metrics are used, the new theory presented in this study – Contractual Realism – constitutes a contribution to scientific progress in the fields of IR theory and comparative politics.

#### 1. Lakatos' methodology of scientific research programmes: A digestible summary

When a political scientist mentions the name *Lakatos* or the phrase *philosophy of social science*, listeners may expect to hear a dry discourse on methodology or an opaque narrative on

epistemology. The concern is heightened when the speaker starts to use terms like "observational theory" and "ceteris paribus condition". Yet actually, the main points of Lakatos' view of science are straightforward and easy to understand.

First, theories are essentially *unprovable*. No single theory can explain everything. Hence no matter how many data points one finds to confirm a theory's predictions, one might always find another data point that refutes its predictions. Therefore, no amount of confirming evidence – no matter how large – can ever *prove* a theory to be valid. Scientists thus face a risk of "false positives": Accepting a theory as valid after finding a mountain of confirming evidence for it, when actually the theory is invalid.

Second, theories are essentially *undisprovable*. No matter how many data points one finds that refute a theory's predictions, one might always find another data point that confirms its predictions. Scientists thus face a risk of "false negatives": Rejecting a theory as invalid after finding some evidence that refutes its predictions, when actually the theory is valid.

Third, the main solution to these problems is *methodological* theory building and theory testing. A *method* is a set of conventions that scientists use to delimit the scope of their theories and code raw data into empirical "observations" for use in testing their theories. The purpose of the scope limitations is to avoid a particular kind of false negative: Rejecting a theory for failing to explain events that it was never designed or intended to explain in the first place. The purpose of the conventional coding rules (and the observational theories on which they are based) is to avoid a particular kind of false positive: Retaining a theory after it runs into empirical anomalies, when the only way to retain it is by modifying the coding rules to get around those anomalies.

Fourth, even this methodological view of science fails to explain the main cases of scientific progress in human history. For this view still assumes that the main task of science is to test each theory against empirical evidence. In fact the main cases of scientific progress arose from testing one theory against another theory. Whichever theory explained as much as its rival explained *and more* won the contest – but only as long as the winning theory complied with two methodological rules for handling empirical anomalies.

1. Anomalies could be handled by narrowing the scope of the theory, but only as long as the theory was still able to grow in explanatory power within its narrower scope. (This

is "novel fact" rule 1.)

2. Anomalies could be handled by rewriting the observational theory used to code data for testing the substantive theory, but only as long as the new combination of an observational theory and a substantive theory was still able to grow in explanatory power. (This is "novel fact" rule 2.)

In discussions of Lakatos by IR theorists, the main focus has been on the "novel fact" aspect of these rules: As a theory is amended over time to handle anomalies, it must still grow in explanatory power; if the theory is amended to handle anomalies without increasing its explanatory power, then the resulting sequence of theories is "degenerative".<sup>1</sup> The focus on novel facts is understandable. Yet it runs the risk of drawing attention away from the specific observational theories and *ceteris paribus* conditions that were developed to save particular IR theories from anomalies. Are these observational theories really necessary to explain the outcomes in question? Or can they be explained by means of some new observational theories that offer more explanatory power with less "massaging" of the raw data to explain away anomalies? I contend that Contractual Realism offers such new observational theories. Similarly, are these *ceteris paribus* conditions really necessary to explain the outcomes in question? Or can they be explained in a new way without such narrow *ceteris paribus* conditions? I contend that Contractual Realism offers such a new way.

This shift in focus – from novel facts to observational theories and *ceteris paribus* conditions – leads to the fifth point in Lakatos' explanation of scientific progress. The flaws in existing theories do not really become apparent until someone develops a new theory that outperforms those theories in explanatory power. Only then does it become apparent that the empirical anomalies encountered by the existing theories are, in fact, critical tests. When this critical moment arrives, scientists can see the observational theories and *ceteris paribus* conditions that were used to save the existing theories from anomalies in a new light – and recognize that they constituted degenerative problemshifts. None of this is apparent, however, before someone develops a new theory that outperforms the existing theories in explanatory power. Hence innovators cannot wait for the existing theories to run into anomalies – and the anomalies to accumulate into refutations – before searching for a new

---

<sup>1</sup>Elman.



theory. For such anomalies alone can never refute the existing theories! Only a new theory that outperforms the existing theories in explanatory power can reveal such anomalies to be critical tests that produce refutation of the existing theories.

Therefore, the task of the innovator is to "try to look at things from different points of view, to put forward new theories which anticipate novel facts, and to reject theories which have been superseded by more powerful ones."<sup>2</sup> In the effort to look at things from different points of view, the innovator should look for a new substantive theory, a new observational theory, or a new combination of both types of theory that offers the greatest explanatory power.

"The problem is then *shifted* from the old problem of replacing a [substantive] theory refuted by 'facts' to the new problem of how to resolve inconsistencies between closely associated theories [e.g. a substantive theory and an observational theory]. Which of the mutually inconsistent theories should be eliminated? The sophisticated falsificationist can answer that question easily: one had to try to replace first one, then the other, then possibly both, and opt for that new set-up which provides the biggest increase in corroborated content, which provides the most progressive problemshift."<sup>3</sup>

I contend that Contractual Realism offers this new set-up – a new substantive theory and a new observational theory – that provides the most progressive problemshift.

The rivalry between old and new theories leads to Lakatos' sixth and final point. Not *all* discoveries of a new theory reveal the anomalies in existing theories to be critical tests. Hence several theoretical schools of thought may persist at the same time – based on mutually contradictory assumptions – with each school encountering anomalies but persisting anyway. In other words, each school simply ignores some of the anomalies that it encounters. The practice of ignoring some anomalies helps to avoid false negatives: Rejecting an entire school of thought as "refuted" after it encounters anomalies, when the school may turn out to be valid later – after further research – so it should be retained until a *conclusive* critical test arises.

Given this rivalry between imperfect schools of thought, each school must develop rules to distinguish between the anomalies it will take seriously – as signs that it needs to amend

---

<sup>2</sup>Lakatos 1970, p. 122.

<sup>3</sup>Quotes in original. Lakatos 1970, p. 130.

its theories – and the anomalies it will ignore. The primary rule is to distinguish between the school’s ”hard core” causal assumptions – which are to be retained no matter how many anomalies the school encounters – and the school’s ”protective belt” theories – which may be amended or discarded after encountering anomalies that are too numerous or fundamental to ignore. In the phase of unresolved competition between the existing schools of thought, each school continues to produce research based on its hard core assumptions despite encountering anomalies and being contradicted by the other schools’ hard core assumptions. When a new school of thought is proposed and suggests a critical test, therefore, the task of the new school is to show that it can explain all that the existing schools explained *as well as the anomalies they ignored*, so those anomalies can no longer be ignored.

## 2. Overview of this chapter and the next two chapters

I undertake three tasks in the remainder of this chapter. First, I explain the Lakatosian methodology in detail, because elements of it can provide new insight into debates about the causes of conflict and the conditions for stable settlement.<sup>4</sup> Second, I use the Lakatosian methodology to reinterpret the debate between Conventional Realism and Conventional Institutionalism and show that it provides an opportunity for a critical test (as Lakatos defined that term). Third, I summarize how Contractual Realism, the new theory proposed in this study, resolves that debate in a new way and thereby offers a progressive problemshift (in the Lakatosian sense).

In the next two chapters, I use this Lakatosian framework to assess the existing theories of Europe’s development at the international and domestic levels and determine whether they are scientifically progressive or degenerative. I conclude that they are degenerative and that Contractual Realism offers a progressive problemshift.

---

<sup>4</sup>In explaining Lakatos’ work, I provide full quotations, because scholars still debate what Lakatos meant to say, so it is essential to establish a solid foundation for my interpretations and applications in this study.

### 3. Lakatosian methodology, the intra-paradigm debates and Contractual Realism

Lakatos developed this methodology to answer three basic questions. When scientific progress occurred through history, *how* did it occur? What does that tell us about how scientists went about their work? And what does this tell us about how to pursue scientific progress in the future? To answer these questions, Lakatos identified three different explanations of scientific progress, pinpointed the flaws in them, and then offered a fourth explanation that seems to provide a better account of scientific progress through history and a better guide to scientific inquiry now. Consider each of the four accounts in turn.

#### 3.1 Account 1: "Verificationism"

This is the simplest explanation of scientific progress. It attributes progress to the collection of empirical evidence that *verifies* theories. A scientist proposes a theory, say,  $T$  causes  $E$ . He collects empirical data on variables that seems to reflect the main causal factor  $T$  and the main effect  $E$ , say, variables  $X$  and  $Y$ . If  $X$  and  $Y$  are correlated, then this data set *verifies* the theory that  $T$  causes  $E$ .

#### Theory

Initial conditions	+	Substantive theory's main causal factor	=	Effect
$I_0$	+	$T_{subst.}$	=	$E$

#### Evidence

$I_0$	+	$X_1$	=	$Y_1$
$I_0$	+	$X_2$	=	$Y_2$
$\vdots$		$\vdots$		$\vdots$
$I_0$	+	$X_n$	=	$Y_n$

According to this account, the history of scientific progress is a long series of *verifications*.

Lakatos argues that verificationism is flawed for one main reason. No amount of *empirical* evidence on variables  $X$  and  $Y$  can ever prove the *theoretical* statement that  $T$  causes  $E$ . As Lakatos phrased it, "facts cannot prove propositions... and no [amount of] logic[al analysis of the empirical facts] can infallibly increase [theoretical] content.... [A]ll theories are equally unprovable."<sup>5</sup> For no matter how many data points the scientist collects – no matter how large  $n$  is – there may always be another data point in which  $X$  is correlated with **not**- $Y$ , thereby refuting the theory that  $T$  causes  $E$ .

### Counter-evidence

$I_0$	+	$X_{n+1}$	=	<b>not</b> - $Y_{n+1}$
-------	---	-----------	---	------------------------

As long as such data is possible, it may be the case that  $T$  does *not* cause  $E$ , in which case the previous evidence that seemed to verify that  $T$  causes  $E$  was a "false positive". Thus verificationism is an unreliable method of scientific inquiry because it creates a risk of "false positives" (accepting bad theories). For the same reason, verificationism cannot explain the history of scientific progress. For it cannot explain how scientists distinguished the truly good theories from the false positives.

#### 3.1.1 Implications for this study

The leading theory of conflict in the political science literature is conventional realism. This theory holds that states divide assets according to relative power at all times, not only in war but in peace as well. By implication, state policies are driven by relative power factors more than anything else. And the main causes of conflict are shifts in relative power and uncertainties about relative power. Members of the conventional realist school have gathered much evidence that verifies the theory that states divide assets according to relative power at all times, even in peacetime.

The leading alternative school of thought in the political science literature is conventional institutionalism. Members of this school have gathered much evidence that falsifies the theory that states divide assets according to relative power in peacetime. Their evidence

---

<sup>5</sup>Emphasis in original. Lakatos 1970, pp. 94–5.

shows that states divide assets according to a different criterion in peacetime: opportunities for mutual gains from exchange. In the 1970s, members of this school studied transnational relations and interdependence and showed the many ways that actors reap mutual gains from exchange across interstate boundaries. In the 1980s, this school studied international governance methods and showed how states enforce agreements to reap mutual gains from exchange. In the 1990s, this school characterized the conditions under which states can reap mutual gains from exchange without concern for the distribution of gains across states (e.g. relative gains concerns). In the 2000s, this school showed that when there is a hegemonic state, the hegemon and the other states can reap mutual gains by agreeing to establish an international hierarchy that enforces international order. All of this evidence falsifies the conventional realist theory that states divide assets according to relative power in peacetime.

According to Lakatos, the conventional realists can never prove their theory that states divide assets according to relative power in peacetime simply by gathering evidence that appears to verify the theory. For the conventional institutionalists will always be able to gather more evidence that states divide assets according to opportunities for mutual gains from exchange instead, thereby falsifying the realist theory. Thus the realist theory is unprovable by the method of verificationism. If the theory is going to be proven valid then it will have to be proven by some other method.

### 3.2 Account 2: "Dogmatic falsificationism"

According to this account of science, the discovery of counter-evidence is actually an asset rather than a liability. For it enables scientists to discard bad theories.

**Falsifying counter-evidence**

$$I_0 + X_{n+1} = \textit{not-}Y_{n+1}$$

**supports counter-theory**

$$I_0 + T = \textit{not-}E$$

**which falsifies original theory**

$$I_0 + T = E$$

As Lakatos writes, "For the dogmatic falsificationist...empirical *counterevidence* is the *one and only* arbiter which may judge a theory."<sup>6</sup> From this perspective the history of scientific progress is a long series of *refutations* that left only the good theories in their wake. However, dogmatic falsificationism suffers from three main problems according to Lakatos.

#### 3.2.1 Problem 1: Substantive theories *versus* observational theories

Dogmatic falsificationism assumes that falsifying evidence is always objective and accurate. Yet in real research, empirical evidence is never objective and accurate. Evidence is always collected and interpreted using an "observational theory" – a theory that specifies how to turn raw data into "observations" on variables  $X$  and  $Y$ . Such observational theories are just as disputable, falsifiable and potentially false as the substantive theories that scientists are trying to test using such 'evidence'. As Lakatos explains,

---

<sup>6</sup>Emphasis in original. Lakatos 1970, p. 96.

"Some empiricists... hold that there is a 'natural demarcation' between [empirical] statements impressed on an empty and passive mind directly by the senses... and [empirical] statements which are suggested by impure, theory-impregnated sensations.... [But actually] there are and can be no sensations unimpregnated by expectations and therefore *there is no natural...demarcation between observational and theoretical propositions.*"<sup>7</sup>

In short, the observational propositions are just as "theoretical" as the theoretical propositions are.

From this perspective, we see that the scientist uses an "observational theory" to turn raw data into "observations" on variables  $X$  and  $Y$ . Then he uses those "observations" to test and potentially falsify a "substantive theory".

---

<sup>7</sup>Emphasis in original. Lakatos 1970, pp. 98–9.

**Observational theory ( $T_{interpr}$ ) turns raw data into falsifying observations**

Datum( $X_{n+1}$ ) + $T_{interpr}$ = Obs( $X_{n+1}$ )				
Datum( $Y_{n+1}$ ) + $T_{interpr}$ = Obs( <b><i>not</i></b> - $Y_{n+1}$ )				
$I_0$	+	$X_{n+1}$	=	<b><i>not</i></b> - $Y_{n+1}$

**falsifying observations support substantive counter-theory**

$I_0$	+	$T_{subst}$	=	<b><i>not</i></b> - $E$
-------	---	-------------	---	-------------------------

**which falsifies original substantive theory**

$I_0$	+	$T_{subst}$	=	$E$
-------	---	-------------	---	-----

Such falsifying observations raise fundamental questions. Was it the raw Datum( $Y_{n+1}$ ) that falsified the substantive theory? Or was it the observational theory ( $T_{interpr}$ ) that falsified the substantive theory? And if it was the observational theory, then might an alternative observational theory ( $T'_{interpr}$ ) have confirmed the substantive theory instead?

**New observational theory ( $T'_{interpr}$ ) turns raw data into supporting observations**

Datum( $Y_{n+1}$ ) + $T'_{interpr}$ = Obs( $Y_{n+1}$ )				
--	--	--	--	--

**supporting observations confirm substantive theory ( $T_{subst}$ )**

$I_0$	+	$X_{n+1}$	=	$Y_{n+1}$
$I_0$	+	$T_{subst}$	=	$E$

The heart of the problem is that there is no such thing as raw data. Empirical "observations" are always "sensed" through some lens that is built according to an *observational* theory – a theory of how lenses should be built to yield accurate observations. Yet observational theories are still *theories*: they can be wrong. So when empirical "observations" seem to refute a substantive theory, it might be because the substantive theory is wrong,



or it might be because the observational theory is wrong. From this perspective, a substantive theory can *never* really be disproven, even by empirical evidence that is anomalous and appears to disprove it. For its author can always dispute the observational theory that generated the falsifying observations and propose an alternative observational theory that avoids such anomalies.

### 3.2.2 Problem 2: Taking anomalies too seriously

Dogmatic falsificationism is vulnerable to the same criticism that plagued verificationism. Facts cannot prove propositions. Even if the scientist can discover some falsifying counter-evidence (e.g.  $X_{n+1}$  is correlated with *not*- $Y_{n+1}$  rather than  $Y_{n+1}$ ), such evidence can never prove the counter-theory that " $T$  causes *not*- $E$ " (in order to falsify the theory that  $T$  causes  $E$ ). For no matter how much falsifying evidence the scientist discovers, there might always be another data point in which  $X_{n+2}$  is correlated with  $Y_{n+2}$  (again), thereby confirming the theory that  $T$  causes  $E$ . As Lakatos writes:

"[T]he truth-value of the 'observational' propositions cannot be indubitably decided: *no factual proposition can ever be proved from an experiment*. Propositions can only be derived from other propositions, they cannot be derived from facts: one cannot prove statements from experiences – 'no more than by thumping the table'....

"If factual propositions are unprovable then they are fallible. If they are fallible then clashes between theories and factual propositions are not 'falsifications' but merely inconsistencies. Our imagination may play a greater role in the formulation of 'theories' than in the formulation of 'factual propositions', but they are both fallible. Thus *we cannot prove theories and we cannot disprove them either*. The demarcation between the soft, unproven 'theories' and the hard, proven 'empirical basis' is non-existent: *all* propositions of science are theoretical and, incurably, fallible."<sup>8</sup>

Because of this problem, dogmatic falsificationism has a tendency to take anomalies too seriously. That is, it produces "false negatives": Observing an apparent anomaly and then discarding the substantive theory as invalid even though it is actually valid.

---

<sup>8</sup>Emphasis in original. Lakatos 1970, p. 99.

### 3.2.3 Problem 3: How many *ceteris paribus* conditions are too many?

When a scientist discovers evidence that contradicts a theory, it may mean that the theory is wrong, or it may simply mean that the theory's author failed to specify its range of applicability properly. In other words, the theory may still be valid under a narrower range of conditions. Lakatos (1970) referred to a theory's range of applicability as the *ceteris paribus* condition. This phrase is Latin for "other things being equal". It means "no other causes were operating in these empirical cases". If the *ceteris paribus* condition is specified too widely, then scientists might discover evidence that contradicts the theory and discard the theory even though it is actually valid under a narrower *ceteris paribus* condition. This is a "false negative" – discarding a theory that is actually valid. To save a theory from this fate, its author can specify a narrower *ceteris paribus* condition. This remedy preserves the opportunity for valid theories to survive, and thus preserves the opportunity for scientific progress to be made. As Lakatos writes, "one can easily argue that *ceteris paribus* clauses are not exceptions, but the rule in science."<sup>9</sup>

This remedy for false negatives creates another problem, however. Suppose that after *ceteris paribus* condition  $C_0$  has been properly specified, another scientist discovers new evidence that contradicts the theory even under condition  $C_0$ . The theory's author can always claim that the *ceteris paribus* condition was specified too widely – and re-specify it more narrowly, as  $C_0 + C_1$ . In other words, the author can always claim:

The new evidence does "not refute [the theory], since [the evidence] may only indicate *other causes* operating simultaneously"<sup>10</sup> under condition  $C_0$ . Those other causes do not operate under narrower conditions  $C_0 + C_1$ , however, so the theory is still valid under those narrower conditions.

A theorist can save his theory from falsifying evidence in this way repeatedly. As Lakatos writes,

"[S]ome scientific theories are normally interpreted as containing a *ceteris paribus* clause: in such cases it is always a specific theory *together* with this clause which

---

<sup>9</sup>Lakatos 1970, p. 102.

<sup>10</sup>Lakatos 1970, p. 102.

may be refuted [by contradictory evidence]. But such a refutation is inconsequential for the *specific* theory under test because, by replacing the *ceteris paribus* clause by a different one, the *specific* theory can always be retained whatever the tests say.”<sup>11</sup>

Lakatos claims that scientists *need* to be able to retain theories in the face of apparently falsifying evidence, because even the best theory does not explain everything. There are always observations that it fails to explain – even observations that contradict it. “[E]xactly the most admired scientific theories simply fail to forbid any observable state of affairs [– even a state that contradicts it].”<sup>12</sup> Hence even the best theories are inevitably confronted with evidence that appears to falsify them. If they are discarded, however, then scientific progress would come to a halt. Therefore scientists need to be able to retain a theory by multiplying the *ceteris paribus* conditions attached to it.

The problem, however, is that the multiplication of *ceteris paribus* conditions eventually renders a theory “undisprovable”.<sup>13</sup> Its author can always dismiss mounting contradictory evidence by claiming: Other causes were operating under those conditions, but the theory is still valid under narrower conditions. The option to make this claim creates a bias toward “false positives” – retaining a theory mistakenly when it really should be discarded because it is contradicted by mounting evidence. In other words, since *ceteris paribus* conditions are the norm in science, theories are ultimately undisprovable – *even bad theories*! And that is the third problem with dogmatic falsificationism. Even bad theories can never really be falsified by “...the dogmatic falsificationist [for whom]...empirical *counterevidence* is the *one and only* arbiter which may judge a theory.”<sup>14</sup>

---

<sup>11</sup>Emphasis in original. Lakatos 1970, pp. 101–2.

<sup>12</sup>Lakatos 1970, p. 100.

<sup>13</sup>Lakatos 1970, p. 102.

<sup>14</sup>Emphasis in original. Lakatos 1970, p. 96.

### 3.2.4 Dogmatic falsificationism fails to explain how scientists ever disproved any theories, let alone how they disproved only the bad theories

We have seen that dogmatic falsificationism suffers from three problems which, taken together, imply that theories are ultimately undisprovable. First, factual counter-evidence may *seem* to disprove a substantive theory, yet that may only be because of the nature of the observational theory used to turn the raw *data* into counter-evidence. If another observational theory is used, it may turn the same raw data into *supporting* evidence that confirms the substantive theory instead. Second, factual counter-evidence can never prove a counter-theory. Third, factual counter-evidence may *seem* to disprove a substantive theory, yet that may only be because the theorist failed to specify the *ceteris paribus* condition narrowly enough. The theory may still be valid under a narrower *ceteris paribus* condition – or not!

Taken together, these problems imply that theories are essentially undisprovable. Yet the history of science reveals that many bad theories *were* disproven and discarded. This historical pattern cannot be explained by dogmatic falsificationism. In particular, it cannot answer two questions.

1. How were *any* theories ever disproven (given that theories are ultimately undisprovable)?
2. How did scientists manage to disprove and discard only the *bad* theories while leaving the *good* theories to survive and constitute scientific progress?

### 3.2.5 Implications for this study

As mentioned earlier, the conventional institutionalists gathered much evidence to show that states divide assets according to opportunities for mutual gains from exchange, thereby falsifying the conventional realist theory that states divide assets according to relative power. Conventional realists responded in all three of the ways anticipated by Lakatos.

First, conventional realists gathered more evidence that verifies their theory that states

divide assets according to relative power. In the 1970s, conventional realists showed that the gains from maintaining international order are distributed according to relative power. In the 1980s, they characterized the conditions under which concerns about relative gains prevent states from cooperating for mutual gain. In the 1990s, they showed that even when states manage to cooperate for mutual gains, the gains are distributed according to relative power. And in some cases the strong states reap all of the gains while the weak states reap none of the gains or incur net losses. All of this evidence verifies the conventional realist theory that states divide assets according to relative power even in peacetime.

Second, conventional realists handled anomalies by narrowing the *ceteris paribus* clauses in their theories.

Third, conventional realists called into question the observational theory that the conventional institutionalists were using to turn the raw data of international relations into evidence of international cooperation. According to the realists, the institutionalists made the mistake of assuming that when states made international commitments and reached international agreements, they actually altered their conduct from what it would have been had they continued acting in their unilateral self-interests (unconstrained by international commitments). This mistake led the institutionalists to overestimate the amount of international cooperation that was actually taking place, according to the realists.

According to Lakatos, the conventional institutionalists can never disprove the realist theory simply by gathering more evidence that appears to falsify it. For the conventional realists will always be able to respond in these three ways. Thus the realist theory is undisprovable by the method of dogmatic falsificationism. If the theory is going to be disproven, then it will have to be disproven by some other method.

These problems lead to the third account of scientific progress in Lakatos' framework. This account focuses on the *conventions* that scientific communities use to render theories disprovable and separate the good theories from the bad theories.

### **3.3 Account 3: "Naive methodological falsificationism"**

According to this account of science, scientists render theories disprovable by adopting two sorts of conventions for the conduct of scientific inquiry. First, they adopt conventions re-

garding the observational theories that are acceptable for collecting and coding empirical evidence for use in testing and disproving substantive theories. Second, they adopt conventions for how to evaluate whether the *ceteris paribus* conditions attached to substantive theories are fulfilled. As long as these conventions are followed, the discovery of empirical evidence that contradicts a substantive theory leads to the rejection of that theory as 'falsified'. This is *methodological* falsificationism. In the next two subsections, I will outline Lakatos' explanation of these two requirements, summarize his critiques of them, and draw the implications for this study. In the third subsection, I will outline Lakatos' broader critique of naive methodological falsificationism and draw further implications for this study.

### 3.3.1 Conventions regarding observational theories

These conventions are motivated by the problem of subjectivity mentioned earlier. There is no empirical observation without an "observational theory" to interpret and make sense of the observations.

"There is an important demarcation between '*passivist*' and '*activist*' theories of knowledge. 'Passivists' hold that true knowledge is Nature's imprint on a perfectly inert mind: mental *activity* can only result in bias and distortion.... [By contrast] 'activists' hold that we cannot read the book of Nature without mental activity, without interpreting [the raw data] in the light of our [observational] expectations or theories.... [*A*]ctivists believe that [such] conceptual frameworks can be developed and also replaced by new, *better* ones...."<sup>15</sup>

Where do these "conceptual frameworks" or "observational theories" come from? And how do scientists decide which ones to keep and which ones to discard?

Building on Karl Popper's work, Lakatos argues that members of a scientific community reach agreement on which *observational* theories to use in gathering and coding empirical evidence to test *substantive* theories. Once agreement is reached, the community regards the selected observational theories as "unfalsifiable" and uses them as "unproblematic background knowledge" in collecting raw data and coding it into empirical evidence for the purpose of testing the substantive theories. As long as a scientist knows the observational

---

<sup>15</sup>Emphasis and quotation marks in original. Lakatos 1970, p. 104.

theories and understands how to use their coding rules to turn raw data into empirical evidence, he can decide whether a particular batch of evidence is "acceptable" for testing a particular substantive theory.

"[The observational theories are] statements decided by agreement.... [T]he truth-value of such statements cannot be proved by facts but, in some cases, may be decided by agreement.... [The scientific community] makes unfalsifiable by *fiat* some...[empirical] statements which are distinguishable by the fact that there exists at the time a 'relevant technique' such that 'anyone who has learned it' will be able to *decide* that the [empirical] statement is 'acceptable'. Such a statement may be called an 'observational' or 'basic' statement.... Indeed, the very selection of all such statements is a matter of a decision.... This decision is then followed by a second kind of decision concerning the separation of the set of *accepted* basic statements from the rest....

"The methodological falsificationist realizes that in the 'experimental techniques' of the scientist, fallible [observational] theories are involved, in the 'light' of which he interprets the [raw empirical] facts. In spite of this he 'applies' these [observational] theories, he regards them in the given context not as theories under test but as *unproblematic background knowledge*

'which we accept (tentatively) as unproblematic while we are testing the [substantive] theory [at hand]'.

He may call these [observational] theories – and the [empirical] statements whose truth-value he decides in their light – 'observational': but this is only a manner of speech which he inherited from [dogmatic] falsificationism. [In fact] the methodological falsificationist *uses our most successful ['observational'] theories as extensions of our senses* and widens the range of ['observational'] theories which can be applied in testing [the substantive theories] far beyond the dogmatic falsificationist's range of strictly observational theories [i.e. those that can be proved by facts].... Calling these statements 'observational' is no more than a manner of saying that...the methodological falsificationist uses [them] uncritically, as 'background knowledge'. *The need for decisions to demarcate the [substantive] theory under test from [such] unproblematic background knowledge is a characteristic feature of this brand of methodological falsificationism....* This consideration shows the conventional element in granting 'observational' status to a[n observational] theory.... [T]hese conventions are institutionalized and endorsed by the scientific community; the list of 'accepted' falsifiers is provided by the verdict of the experimental scientists."<sup>16</sup>

The key point is that once the scientific community selects an 'observational' theory as its preferred method of collecting raw data and coding it into empirical 'evidence', the commu-

---

<sup>16</sup>Emphasis, parentheses and quotations in original. Underlining added. Lakatos 1970, pp. 106–8.

nity "makes [that theory] unfalsifiable by *fiat*."<sup>17</sup> This move solves the problem mentioned earlier: When a substantive theory is contradicted by empirical evidence, its author can always dispute the observational theory that produced the falsifying observations. Such dispute is forbidden once the community has made that observational theory unfalsifiable by fiat.

Now, a substantive theory can be rejected once someone discovers 'evidence' that contradicts it. However, this is a *methodological* falsification, not a *dogmatic* falsification. The methodological falsificationist can only *reject* a substantive theory. He cannot disprove it (as the dogmatic falsificationist claimed to be able to do). Moreover, because the rejection results from the community's choice of conventions to hold each other to – as much as from the raw data itself – the rejection might be mistaken: a false negative. As Lakatos explains,

"This [conventional approach] is how the methodological falsificationist establishes his 'empirical basis' [for rejecting substantive theories].... This 'basis' can hardly be called a 'basis' by justificationist [verificationist] standards: there is nothing proven about it – it denotes 'piles driven into a swamp'. Indeed, if this 'empirical basis' clashes with a [substantive] theory, the theory may be *called* 'falsified', but it is not falsified in the sense that it is disproved. Methodological 'falsification' is very different from dogmatic falsification. If a theory is falsified, it is proven false; if it is [only] 'falsified' [however], it may still be true. If we follow this sort of 'falsification' by the actual 'elimination' of a theory, we may well end up by eliminating a true, and accepting a false, theory.... Yet the methodological falsificationist advises that exactly this is to be done. The methodological falsificationist realizes that...we *must* find a way to eliminate *some* theories. If we do not succeed [in finding a way to eliminate some theories], the growth of science will be nothing but growing chaos."<sup>18</sup>

Thus methodological falsificationism is based on conventions regarding the collection and coding of empirical evidence. The conventions may be questioned. Yet at least they give scientists a rational way to reject theories contradicted by empirical evidence (as opposed to dogmatic falsificationism, whose method of rejecting theories was not rationalizable). It is rational in the sense that it is based on observational theories that scientists agree by convention are rational. Thus it is the inter-subjective rationality of the scientific community that ultimately determines which substantive theories the community will accept or reject.

---

<sup>17</sup>Emphasis in original. Lakatos 1970, p. 106.

<sup>18</sup>Emphasis and quotes in original. Lakatos 1970, p. 108.



### 3.3.2 Conventions regarding how to test *ceteris paribus* clauses

These conventions are motivated by the problem of *ceteris paribus* clauses mentioned earlier. When a theory is confronted by contradictory evidence, its author can always claim that its *ceteris paribus* clause was specified too widely – that other causal factors were operating in that wide range of conditions – but the theory is still valid under a narrower *ceteris paribus* condition where those other factors were not operating.

”[How can scientists] ‘falsify’ a theory which cannot explain anything ‘observable’ without a *ceteris paribus* clause. No finite number of ‘observations’ is enough to ‘falsify’ such a theory [since its *ceteris paribus* clause can always be narrowed down to rule out a particular piece of contradictory evidence].”<sup>19</sup>

The solution is simply to test the *ceteris paribus* clause to see whether other causal factors were operating or not. If not, then the clause is accepted.

The methodological falsificationist solves the problem by making a further...*decision*: when he tests a theory together with a *ceteris paribus* clause and finds this conjunction has been refuted [by the evidence],.... he has to take the crucial decision: whether to also relegate the *ceteris paribus* clause into the pool of ‘unproblematic background knowledge’. He will do so if he finds the *ceteris paribus* clause well corroborated.

”How can one test a *ceteris paribus* clause severely? By assuming that there *are* other influencing factors [i.e. that other things were not equal], by specifying such factors, and by testing these specific assumptions. If many of them are refuted, the *ceteris paribus* clause will be regarded as well-corroborated.”<sup>20</sup>

This is where conventions enter the picture. The scientific community agrees on a set of conventional *methods* for testing *ceteris paribus* clauses (e.g. ruling out alternative rival hypotheses). As long as these conventions are followed, then the discovery of contradictory evidence is assumed to refute a substantive theory. For example, as long as the appropriate control variables are included in a statistical regression, then the main hypothesis is rejected when the coefficient on the main variable is estimated to be statistically *insignificant*.

---

<sup>19</sup>Emphasis and quotes in original. Lakatos 1970, p. 110.

<sup>20</sup>Emphasis and quotes in original. Lakatos 1970, p. 110.

Lakatos argues that such methodological conventions are *strongly committing*. Once a *ceteris paribus* clause is corroborated by the conventionally accepted methods, the scientist becomes committed to rejecting a substantive theory that is contradicted by empirical evidence.

"[T]he decision to 'accept' a *ceteris paribus* clause is a very risky one because of the grave consequences it implies. [An empirical] '*anomaly*' in relation to theory *T*'...becomes a potential falsifier of *T* itself after having decided to relegate the *ceteris paribus* clause into 'unproblematic background knowledge'....

"The methodological falsificationist is in a serious plight when it comes to deciding where to draw the demarcation... between the problematic and unproblematic. The plight is most dramatic when he has to make a decision about *ceteris paribus* clauses [e.g. that they are valid], when he has to promote one of the hundreds of 'anomalous phenomena' into a 'crucial experiment', and decide that in such a case the experiment was 'controlled' [to insure no other causal factors were operating].... [Yet as Popper argues] we cannot make scientific progress unless we have a firm rational strategy or method to guide us when [there is a] clash [between theory and evidence].<sup>21</sup>

In other words, naive methodological falsificationism is committed to making theories falsifiable by adopting the conventions necessary to perform falsification tests. Having adopted these conventions, however, it is bound to comply with them and reject theories that have been falsified.

Lakatos criticizes this "firm strategy" on two grounds. First, it fails to explain the main episodes of scientific progress throughout history – a point to which I will return in a moment. Second, the conventions on which it relies may be arbitrary, and hence the falsifications that they produce may be equally arbitrary.

Methodological falsificationism....recommends risky decisions. But the risks are daring to the point of recklessness and one wonders whether there is no way of lessening them.... [I]s not the firm strategy of th[is] brand of methodological falsificationism...*too firm*? Are not the [conventional] decisions it advocates bound to be *too arbitrary*?"<sup>22</sup>

The firm strategy of rejecting theories that fail the conventional empirical tests can lead to false negatives – thereby blinding scientists to good theories. For example, Goertz and

---

<sup>21</sup>Emphasis and quotes in original. Lakatos 1970, pp. 110–4.

<sup>22</sup>Emphasis and quotes in original. Lakatos 1970, pp. 110–2.

Mahoney (2012) discuss Geddes' decision to reject the hypothesis that labor repression is correlated with economic growth on the grounds that the regression coefficient was statistically insignificant. They show that in fact there *is* a relationship between the two variables in that dataset. None of the countries with low labor repression had high economic growth. Only the countries with moderate or high labor repression had high economic growth. Thus labor repression is a necessary condition for high economic growth among the countries in that dataset. This example illustrates Lakatos' point that a firm strategy of rejecting theories that fail the conventional empirical tests can blind scientists to important causal mechanisms operating in the world.

### 3.3.3 Scientific progress occurs mainly through the confirmation of new theories, not the falsification of existing theories

According to the naive methodological falsificationist, scientific history is a long series of falsifications that left only the valid theories still standing. Each falsification was based on strict conventional rules of theory testing that left no room for appeals. The author of a falsified (substantive) theory was not permitted to dispute the observational theories used by convention to collect the falsifying evidence. Nor was he permitted to save the theory by narrowing down the *ceteris paribus* condition attached to it. Once the original *ceteris paribus* condition was corroborated by the conventional methods, the substantive theory would be rejected immediately and irrevocably upon being contradicted by empirical evidence.

Lakatos argues that this is not an accurate picture of scientific history. In fact the authors of falsified substantive theories *were* permitted to dispute the observational theories used by convention to collect the falsifying evidence.

"[S]tubborn theoreticians frequently challenge experimental verdicts and have them reversed. In the falsificationist conception of scientific 'law and order' we have described there is no place for such successful appeals."<sup>23</sup>

And substantive theories contradicted by empirical evidence were *not* rejected immediately even though their *ceteris paribus* conditions were corroborated.

---

<sup>23</sup>Quotes in original. Lakatos 1970, p. 114.

”[E]ighty-five years elapsed between the acceptance of the perihelion of Mercury as an anomaly [for Newton’s theory] and its acceptance as a falsification of Newton’s theory, in spite of the fact that the *ceteris paribus* clause was reasonably well corroborated.”<sup>24</sup>

Most importantly, scientific history was not a series of confrontations between theory and evidence, where the theory was rejected if falsified by the evidence. Rather, scientific history was a series of confrontations between one theory and another theory, where the theory that explained as much as its rival explained *and more* was accepted.

”[There are] at least two crucial characteristics common to both dogmatic and our [naive] methodological falsificationism which are clearly dissonant with the actual history of science: that

- 1 *a test is – or must be made – a two-cornered fight between theory and experiment so that in the final confrontation only these two face each other; and*
- 2 *the only interesting outcome of such confrontation is (conclusive) falsification: ‘[the only genuine] discoveries are refutations of scientific hypotheses.’*

However, history of science suggests that

- 1’ tests are – at least – three-cornered fights between rival theories and experiment and
- 2’ some of the most interesting experiments result, *prima facie*, in confirmation rather than falsification.”<sup>25</sup>

Lakatos therefore concludes that this brand of methodological falsificationism is ”naive” because it misses the central fact of scientific history. Progress arises not simply from the falsification of existing theories, but from the discovery *and confirmation* of new theories *that outperform the existing theories*.

### 3.3.4 Implications for this study

Conventional Realism and Conventional Institutionalism define progress by the methodologies of verificationism and dogmatic falsificationism. But as Lakatos showed, a theory can never be proven by the verificationist approach or disproven by the dogmatic falsificationist

---

<sup>24</sup>Lakatos 1970, p. 115.

<sup>25</sup>Emphasis in original. Third bracketed phrase in original. Lakatos 1970, p. 115.

approach. For this reason scientists turn to the approach of *methodological* falsificationism. They use *methods* to protect a paradigm from false negatives and false positives (e.g. discarding a valid theory by mistake and accepting an invalid theory by mistake). To protect a paradigm from false negatives, they use the method of *ceteris paribus* conditions. To protect a paradigm from false positives, they use the method of "*observational theories*" to turn raw data into "observations" for use in testing theories. According to the rules of methodological falsificationism, the accumulation of falsifying evidence against a theory requires that scientists discard that theory. However, Lakatos argued, scientists tend to retain observational theories even after falsifying evidence accumulates, because scientific progress is not made *simply* by testing each theory against evidence. It is made by testing one theory *against another theory* in terms of their comparative ability to explain evidence.

Conventional Realism and Conventional Institutionalism employ observational theories. Conventional Realism employs an observational theory that turns raw data on the outbreak of conflicts into observations about shifts in relative power, uncertainty about relative power and commitment problems related to relative power. Conventional Institutionalism employs an observational theory that turns raw data on international cooperation into observations about costly punishment as an enforcement device. Conventional Realism and Conventional Institutionalism have retained their observational theories despite falsifying evidence, as Lakatos anticipated, because no one has yet produced a new theory that outperforms them in explanatory power.

Through history, according to Lakatos, scientific progress occurred when an innovator developed a new observational theory and a new substantive theory that together explained (1) what previous paradigms explained, (2) the anomalies they failed to explain, and (3) novel facts they never claimed to explain. The new paradigm developed in this study, Contractual Realism, is composed of a new observational theory and a new substantive theory. The observational theory turns the raw data on the outbreak of conflicts into observations about legal incompatibility disputes. It also turns the raw data on international cooperation into observations about anchor contracts and anchor based punishment mechanisms.<sup>26</sup> I argue that this combination of a new observational theory and a new substantive theory provides

---

<sup>26</sup>An anchor contract is an agreement whose terms are invariant to shifts in relative power. An anchor based punishment is beneficial for the punisher to impose rather than costly, as with conventional punishments.

greater explanatory power than existing theories.

### 3.4 Account 4: "Sophisticated methodological falsificationism" – the state of the art for measuring scientific progress

This account of science questions the "firm strategy" of decisively rejecting theories simply because they are contradicted by empirical evidence – and replaces it with a more sophisticated strategy. As Lakatos explains, the history of scientific progress is filled with examples of theories that were *not* decisively rejected simply because they were contradicted by empirical evidence. Instead theoreticians amended their (substantive) theories in one of two ways: (1) by narrowing down the *ceteris paribus* conditions attached to their theories, or (2) by proposing alternative observational theories for collecting and coding empirical evidence. In some cases the amendment saved the substantive theory and the theory turned out to be valid. In other cases the amendment saved the substantive theory for a time but eventually the theory turned out to be invalid. Given this pattern of scientific history, the key question is how to distinguish "progressive" amendments that save valid theories from "degenerative" amendments that save invalid theories.

"Why aim at falsification at any price? Why not rather impose certain standards on the theoretical adjustments by which one is allowed to save a theory? Indeed, some such standards have been well-known for centuries, and we find them expressed in age-old wisecracks against *ad hoc* explanations, empty prevarications, face-saving, linguistic tricks....

"Popper agrees with the conventionalists that theories and factual propositions can always be harmonized with the help of auxiliary hypotheses: he agrees that the problem is how to demarcate between scientific and pseudoscientific *adjustments*, between rational and irrational changes of theory. According to Popper, saving a theory with the help of auxiliary hypotheses which satisfy certain well-defined conditions represents scientific progress; but saving a theory with the help of auxiliary hypotheses which do not [satisfy such conditions] represents degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, 'conventionalist stratagems'. But then [if Popper is right] any scientific theory has to be appraised together with its auxiliary hypotheses, initial conditions, etc., and, especially, together with its predecessors so that we may see by what sort of *change* it was brought about. Then, of course, what we

appraise is a *series of theories* rather than isolated *theories*.”<sup>27</sup>

According to Lakatos, this appraisal problem is the key to explaining scientific history and conducting scientific inquiry.

Historically, valid theories were discovered through a process of *progressive* amendment to account for contradictory evidence, while invalid theories were finally recognized as such after their amendments were recognized to be *degenerative*. Prospectively, the central task of scientists who are adjudicating between competing theories is to distinguish between progressive and degenerative amendments.

”Let us take a series of theories,  $T_1, T_2, T_3, \dots$  where each subsequent theory results from adding auxiliary clauses to...the previous theory in order to accommodate some [empirical] anomaly, each theory having at least as much content as the **unrefuted** content of its predecessor. Let us say that such a series of theories is *theoretically progressive* (or ‘constitutes a theoretically progressive problemshift’) if each new theory has some excess empirical content over its predecessor, that is, if it predicts some novel, hitherto unexpected fact. Let us say that a theoretically progressive series of theories is also *empirically progressive* (or ‘constitutes an empirically progressive problemshift’) if some of this excess empirical content is also corroborated, that is, if each new theory leads us to the actual discovery of some *new fact*. Finally, let us call a problemshift *progressive* if it is both theoretically and empirically progressive, and *degenerating* if it is not [both].... Progress is measured by the degree to which a problemshift is progressive, by the degree to which the series of theories leads us to the discovery of novel facts. We regard a theory in the series ‘falsified’ when it is superseded by a theory with higher corroborated content.”<sup>28</sup>

From this perspective the central aim of scientific inquiry is to discover new theories that explain as much as existing theories explained *and more*. Lakatos derives several practical lessons from this understanding of scientific inquiry. In each of the following subsections, I summarize one of these lessons and derive its implication for this study.

### 3.4.1 No falsification of old theories before the development of a new theory

A theory cannot be falsified without first developing a new theory that outperforms it.

---

<sup>27</sup>Emphasis in original. Underlining added. Lakatos 1970, pp. 117–8.

<sup>28</sup>Emphasis in original. Underlining and boldface added. Lakatos 1970, p. 118.

"*There is no falsification before the emergence of a better theory.... This shows that 'crucial counter-evidence' – or 'crucial experiments' can be recognized as such among the scores of anomalies only with hindsight, in the light of some superseding theory.*"<sup>29</sup>

In this way Lakatos resolves the conundrum of the relation between confirmation and falsification through scientific history:

"[Confirmationists] valued 'confirming' instances of a theory; [by contrast] naive [methodological] falsificationists stressed 'refuting' instances; [by contrast] for the [sophisticated] methodological falsificationists it is the – rather rare – corroborating instances of the *excess* information which are the crucial ones; these receive all the attention. [That is to say] we are no longer interested in the thousands of trivial verifying instances nor in the hundreds of readily available anomalies: [rather] the few crucial *excess-verifying instances* are decisive."<sup>30</sup>

In other words, the only way to *falsify* one theory is to *confirm* a new rival theory that explains as much as the falsified theory explained and more. This is how confirmation and falsification are related to each other – intimately and inextricably.<sup>31</sup>

**Implication.** In this study I derive as many observable implications as possible from the theory of Contractual Realism in order to explain as many "novel facts" as possible. Accordingly, the study derives many more implications about hegemonic war than previous theories of war did. The study derives implications about many different kinds of conflict: symmetric and asymmetric, international and domestic, violent and non-violent. And the study derives implications not only about the causes of conflict, but also about the conduct and settlement of conflict. In so far as all of these implications are corroborated by the historical evidence, it lends support to the claim that Contractual Realism offers greater explanatory power than previous theories did. In so far as Contractual Realism explains empirical facts that were *anomalies* for the previous theories, it supports the claim that Contractual Realism refutes the previous theories.

---

<sup>29</sup>Emphasis in original. Lakatos 1970, pp. 119–20.

<sup>30</sup>Emphasis in original. Lakatos 1970, p. 120.

<sup>31</sup>At the same time, such confirmations do not literally prove the new theory. They only demonstrate its explanatory power. "A 'verification' is a corroboration of excess content in the expanding [new] programme. But, of course, a 'verification' does not *verify* a programme: it shows only its heuristic power." (Emphasis in original. Lakatos 1970, p. 137)



### 3.4.2 Theorists cannot wait for falsifications of existing theories to search for new theories

Theoretical innovators cannot wait for the falsification of existing theories before they go searching for new theories. They must first find new theories that outperform the existing theories. For only then, through the lens of the new theory, can they recognize that the anomalies not explained by existing theories were "crucial experiments" or "critical tests".

"Falsification cannot 'compel the theorist to search for a better theory', simply because falsification cannot precede the better theory.... [S]ophisticated falsificationism demands that one should try to look at things from different points of view, to put forward new theories which anticipate novel facts, and to reject theories which have been superseded by more powerful ones."<sup>32</sup>

**Implication.** In this study I "look at things from different points of view" in a few ways. The study breaks down the conventional distinctions between symmetric and asymmetric conflict, international and domestic conflict, violent and non-violent conflict. The study supplies a new observational theory for conceptualizing and measuring stability, security and the root causes of disputes that erupt in conflict. The study supplies a new observational theory for conceptualizing and measuring power, the returns to power, and the returns to conflict. Finally, the study uses these different points of view to "put forward new theories which anticipate novel facts", e.g. to derive observable implications for as many different dependent variables as possible.

### 3.4.3 Critical tests between theories turn on how each theory resolves anomalies

When it is not clear which of two competing theories explains more than the other, the adjudicator must examine how the proponents of each theory went about amending it to account for empirical anomalies. If they amended it in a way that also expanded it to explain new facts, then the amendment was progressive. But if they amended it in a way that did not explain any new facts, then the amendment was degenerative.

---

<sup>32</sup>Quotes in original. Lakatos 1970, p. 122.

"If we put forward a[nother] theory [in a series] to resolve the contradiction between the previous theory [in the series] and a[n empirical] counterexample, [but we do it] in such a way that the new theory, instead of offering a content-increasing (scientific) *explanation*, only offers a content-decreasing (linguistic) *reinterpretation*, [then we must conclude that] the contradiction [between theory and evidence has been] resolved in a merely semantical, unscientific way. *A given fact is explained scientifically only if a new fact is also explained with it.*"<sup>33</sup>

**Implication.** It is not enough for this study to demonstrate that Contractual Realism explains a wide range of historical facts about conflict and settlement. It is also necessary to show that previous theories were amended to handle anomalies in a degenerative way. In chapters 5 and 6, I review existing theories of conflict, settlement and long-run development in Europe. I critique the ways in which theorists amended these theories to account for anomalies. And I show that Contractual Realism can explain these anomalies without needing to be amended in such ways.

#### 3.4.4 Anomalies can be resolved by proposing a new observational theory, but only if it generates excess empirical content

When a substantive theory is confronted by contradictory evidence, its author may dispute the observational theory that was used to collect and code the evidence, but only if he proposes an alternative observational theory that increases the explanatory power of his substantive theory.

"[E]xperiments do not simply overthrow theories... [for] no theory forbids a state of affairs specifiable in advance. It is not that we propose a theory and Nature may shout NO; rather, we propose a maze of theories, and Nature may shout INCONSISTENT.

"The problem is then *shifted* from the old problem of replacing a [substantive] theory refuted by 'facts' to the new problem of how to resolve inconsistencies between closely associated theories [e.g. a substantive theory and an observational theory]. Which of the mutually inconsistent theories should be eliminated? The sophisticated falsificationist can answer that question easily: one had to try to replace first one, then the other, then possibly both, and opt for that new set-up which provides the biggest increase in corroborated content, which provides the most progressive problemshift.

---

<sup>33</sup>Emphasis in original. Lakatos 1970, p. 119.

"Thus we have established an appeal procedure in case the theoretician wishes to question the negative verdict of the experimentalist. The theoretician may demand that the experimentalist specify his 'interpretive [observational] theory', and [the theoretician] may then replace it – to the experimentalist's annoyance – by a better one in the light of which his originally 'refuted' [substantive] theory may receive positive appraisal."<sup>34</sup>

**Implication.** In this study I provide: (1) a new observational theory for measuring stability, security and the root causes of disputes that erupt in conflict; (2) a new observational theory for measuring power, the returns to power, and the returns to conflict; and (3) a new substantive theory of conflict initiation, conduct and settlement. This new set-up is used to explain hegemonic wars and settlements in Parts II through V. This set-up is then used to explain other kinds of conflicts and settlements in Parts VI and VII. The central claim is that this set-up "provides the biggest increase in corroborated content,... the most progressive problemshift" for explaining conflict and settlement through European history.

---

<sup>34</sup>Quotes in original. Lakatos 1970, p. 130.

## CHAPTER 5

### Critical tests (I): Theories of Europe's development at the international level

In the previous chapter I explained the main principles of scientific progress. The core principle is that existing theories cannot be disproven until someone develops a new theory that surpasses them in explanatory power. The developer's task is to show that previous theories could only explain anomalous cases by adding auxiliary hypotheses, while the new theory can explain the anomalous cases without adding auxiliary hypotheses. With this demonstration it becomes clear that the new theory offers the opportunity for a critical test. Either the previous theories provide the explanation of the data in question, or the new theory provides the explanation of it. But not both. That option is logically foreclosed by the demonstration that the new theory explains all that the previous theories explained – and more – based on *different* causal assumptions without needing auxiliary hypotheses.

The purpose of this chapter and the next chapter is to summarize previous theories of Europe's development in these terms. For each theory I will:

1. summarize the theory;
2. list the supporting cases that corroborate it;
3. list the anomalous cases that contradict it (thus requiring the addition of auxiliary hypotheses to save it); and
4. summarize how the new theory offered in this study – Contractual Realism – explains both sets of cases without needing auxiliary hypotheses.

In these two chapters I *summarize* the theories and the evidence that set up the critical tests. I do not provide the historical evidence in detail with references to make the full case. That is the purpose of the case studies offered in later chapters. Here the goal is simply to clarify what the case studies need to demonstrate and what the critical test results will be if the case studies succeed in those demonstrations.<sup>1</sup>

#### 1. Overview of this chapter

Since the Renaissance the great powers of Europe have experienced repeated oscillations between peace and war. Every 50 to 100 years a great power launched a campaign of territorial expansion and tried to establish hegemony over much of Europe. Scholars have labeled these conflicts “hegemonic” wars. These wars had pervasive effects on Europe's development at the international and domestic levels. To fight the wars, states had to develop new technologies, greater fiscal capacity, and more efficient bureaucracies. To settle the wars, states had to develop new institutions at the international and domestic levels. To make the

---

<sup>1</sup>I provide evidence and references at some points simply for clarity, narrative continuity, and the benefit of non-specialists who are less familiar with the historical cases under study

settlements stable, states had to refine the institutions to reflect the lessons learned from instabilities after previous hegemonic wars. One refinement was to create good governance rules at the international level to protect weak states from strong states and at the domestic level to protect weak actors in society from strong actors. A second refinement was to create new diplomatic and contractual structures to enable the great powers to reach shared understandings and maintain them without recourse to the traditional tools of balancing, boycotting and social capital subversion. After all of these developments, Europe was a very different place than it had been in 1500.

In this study I argue that all of the hegemonic wars in Europe since 1500 were caused by the same causal mechanism. This mechanism explains all of the developments just mentioned. I also argue, however, that the dichotomy between war and peace fails to capture how these developments were produced by this causal mechanism. Three additional dichotomies are needed. The first is the dichotomy between stability and instability at the international level. The second is the dichotomy between good governance and misgovernance at the international level. The third is the dichotomy between social capital formation and social capital depletion at the international level. These dichotomies reflect the range of effects produced by this causal mechanism. Previous theories have attempted to explain variation within each dichotomy (e.g. from stability to instability and back, etc). Yet each theory encounters anomalous cases that can only be explained by the addition of auxiliary hypotheses. The new theory proposed in this study, Contractual Realism, explains these anomalies (as well as the cases explained by the previous theories).

In the first section, I critique the leading explanations of what caused hegemonic wars in Europe between 1500 and 2000. In the second section, I critique the leading explanations of the causes of stability and instability. In the third section, I critique the leading explanations of the causes of good governance and misgovernance. In the fourth section, I critique the leading explanations of the causes of social capital formation and depletion. In each section I also summarize how Contractual Realism explains both the cases that the previous theories explained and the anomalies that they left unexplained.

## **2. Explanations of hegemonic wars in Europe between 1550 and 1950**

According to the leading theories, these wars were caused by relative power factors. I label this causal assumption the “relative power axiom”. While each theory explains some of the key facts, it fails to explain others or is contradicted by them. Theorists either ignore these anomalies or explain them away by adding auxiliary hypotheses or amending the coding rules for measuring relative power. The Contractual Realist theory explains these anomalies without such maneuvers (as well as the cases the previous theories explained).

### **2.1 Balance of power theory**

This explanation focused on the role of shifts in territorial power. Whenever a great power increased its territorial holdings by too much, on this view, the other great powers banded together to form a balancing coalition to roll it back. This theory explains the cases of hegemonic war in which the military hostilities began *after* one great power expanded its

territorial holdings. These cases were: Spain's expansion into the Netherlands in the 1580s, Austria's expansion into Germany in the 1620s, and France's expansion into Italy and the southern Netherlands in 1701. The theory fails to explain the cases in which military hostilities began *before* any great power expanded its territorial holdings, however. These anomalies include the Nine Years War that began in 1688, the French Revolutionary Wars that began in 1792 and World War I.

To explain these anomalies, balance of power theorists expanded the rules for coding shifts in relative power. Whereas the original theory defined power shifts in terms of territorial expansion alone, the amended theory defined power shifts to include differences among states in the long run growth of their populations and economic productivity. The amended theory was claimed to explain the three cases of hegemonic war in which military hostilities began before any great power engaged in territorial expansion. Thus the Nine Years War was attributed to the relatively high growth of France's population and productivity in the 1600s. The French Revolutionary Wars were attributed to the relatively high growth of French population and productivity in the 1700s. World War I was attributed to the relatively high growth of German population and productivity in the 1800s.

There are two problems with the amended theory. First, France's long run growth rates had gone past their peak relative to the other great powers long before the outbreak of the Nine Years War. So this case is an anomaly for the theory. Second, France's long run growth rates declined in the 20 years before the outbreak of the French Revolutionary Wars. This case is an anomaly for the theory. The Contractual Realist theory can explain all three cases without recourse to coding rule amendments or auxiliary hypotheses (see section 30 below).

## 2.2 Defensive realism

The previous theory, balance of power theory, was plagued by three kinds of anomaly. First, there were cases where war erupted without any shift in relative power *defined as* relative territorial holdings. Second, there were cases where war erupted without any shift in relative power *defined as* relative long run growth rates. Third, there was a case where war erupted without a shift in either kind of relative power. These anomalies needed to be addressed to save the axiom that hegemonic wars were caused by shifts in relative power. To save it, scholars added an auxiliary hypothesis that there are two types of states: "status quo" states and "revisionist" states. According to the amended theory, each war was caused by two factors *in conjunction*: the emergence of a "revisionist" state and a shift in relative power of some kind. Each war was necessary, on this view, because of a fundamental clash of interests between the "revisionist" state and the "status quo" states. If no revisionist state had emerged, then the status quo states would not have needed to initiate a war or enter one – and there would have been none.

This amended theory seems to resolve the three kinds of anomaly. First, the theory explains the cases where war erupted *after* a state engaged in territorial expansion. Its expansion revealed "revisionist" aims that provoked the "status quo" states into forming a balancing coalition to roll it back. Second, the theory explains the cases where war erupted *before* any state engaged in territorial expansion. One state had a higher long run growth rate than the others, this induced it to develop "revisionist" aims, and the "status quo" states detected those aims before the state could launch a territorial expansion. Third, the theory explains the cases where war erupted without any prior territorial expansion or shift

in long run growth rates. One state developed “revisionist” aims for some other reason. The “status quo” states detected those aims and perceived they *included* territorial expansion. And the status quo states decided to block it before it could get off the ground.

There is one problem with the amended theory. In each hegemonic war, *each* of the opposing states viewed the other state as “revisionist”. And each state had a formal basis in law for doing so: The observation that the other state had violated its rights in law. So it is factually incorrect to say that (1) there were two types of state in each hegemonic war: the revisionist state and the status quo states; and (2) the war was caused by some fundamental clash of interest between the two types. In fact, *each* side had a basis for perceiving itself as the status quo type of state and the opponent as the revisionist type. *Both* sides were status quo types – and both sides were revisionist types! Clearly, the distinction between these two types of state – and the assumption that only one side in each war was status quo – does not provide the traction needed to explain the historical evidence on mutual perceptions and their impact on decisions to wage war. Each of these seven wars is an anomaly for any theory based on that assumption.

### 2.3 Offensive realism

Once scholars introduced the auxiliary hypothesis that there are two types of state – “status quo” states and “revisionist” states – their challenge was in how to link that auxiliary hypothesis to the main hypothesis that wars are caused by shifts in relative power. Defensive realism made the connection by assuming that all states are “status quo” states *until* a shift in relative power induces one state to develop “revisionist” preferences. The next theory in the sequence, offensive realism, made the connection by adopting the opposite assumption. All states have “revisionist” preferences, in the form of a preference for regional hegemony, yet none of them has enough power to act on that preference *until* a shift in relative power gives one state enough power to act on it.

This theory is intended to explain the origin of one great power’s quest for regional hegemony in each case of hegemonic war between 1550 and 1950. However, the theory fails to explain three of the cases: the Wars of Louis XIV that broke out in 1688 and 1701 and the French Revolutionary Wars that broke out in 1792. In each of these cases the theory codes the historical evidence as indicating a massive shift in relative power in France’s favor in the year that the war broke out. The evidence shows no such shift, on any definition of relative power. These cases are thus anomalies for the theory.

### 2.4 Structural realism

Running underneath all of the previous theories was a deeper axiom about the role of relative power in international politics. This axiom holds that relative power constitutes a *structural constraint* on state behavior. It forces states to maintain military forces at all times, to use force to defend their interests if other defenses fail, and above all, to form balancing coalitions to prevent adversaries from gaining too much power. This theory explains the cases of hegemonic war in which a balancing coalition formed and prevented an expansionist state from establishing regional hegemony:

- the Dutch and English coalition to prevent Spanish hegemony in the 1580s;

- the Swedish and French coalition to prevent Austrian hegemony in the 1630s;
- the English-led coalitions to prevent French hegemony from 1689 to 1713; and
- the Allied coalition to prevent German hegemony from 1914 to 1918.

The theory cannot explain the cases of hegemonic war in which a balancing coalition failed to form or failed to prevent the expansionist state from establishing regional hegemony, however. The theory also fails to explain why some states even “bandwagoned” with the expansionist power. The cases of balancing failure and bandwagoning include the hegemony established by France after 1797 and the hegemony established by Germany in 1940. These cases are anomalies for structural realism.

## 2.5 Neoclassical realism

To explain instances of “failure to balance” and “bandwagoning”, scholars added in auxiliary hypothesis to the relative power axiom. They assumed that when one state expanded territorially due to a shift in relative power, the other states’ decisions whether to fight back depended on domestic political factors that affected their perceptions of relative power and responses to shifts in power. This amended theory seems to explain the cases in which states “bandwagoned” against the aggressor rather than “balancing” against it. The problem with the theory, however, is that this auxiliary hypothesis is not actually necessary to explain such cases. The Contractual Realist theory explains such cases without requiring such auxiliary hypotheses (see below).

## 2.6 Game-theoretic implementations of Conventional Realism

Perhaps in response to such anomalies and additions of auxiliary hypotheses, scholars sought a more parsimonious theory that could explain a range of wars as a consequence of relative power factors alone. To do so they focused on the essential puzzle of war. Why would two states fight a war over a dispute if they could reach a negotiated settlement and save the costs of a war? Scholars developed two theories to answer this question. Each theory fails to explain the hegemonic wars in Europe between 1550 and 1950, however.

**The bargaining theory of war** This theory assumes that war results from uncertainty about relative power in a world where power is always shifting. In such a world, one state can bluff that its power has increased by more than it really has, in order to extract more concessions from another state than it could ever win by using force (where “winning” means “reaping a net gain”). The only way for the other state to deter such bluffing is to stand firm and fight back against the first state’s demands at least some of the time. In equilibrium, such bluffing is deterred; but when a state’s power really has increased by enough that it could win by using force, the other state does not know this (due to the uncertainty assumption of the theory), so it stands firm and the result is war.

This theory provides a logical answer to the question of why two states would ever fight a war over a dispute. However, the theory fails to explain long wars (Powell 2006). For if a war were to break out through this mechanism, the uncertainty about relative power would be resolved quickly by the fighting itself. If the demanding state’s power had not really grown by enough to win by force, then that state would sue for peace quickly. Alternatively, if its power really had grown by enough to win by force, then the other state would sue for peace



quickly. Each of the hegemonic wars since 1550 was long and costly, however. Thus all of these cases are anomalies for the theory.

**War as a commitment problem arising from relative power factors** In response to this explanatory puzzle, scholars developed a different answer to the question of why two states would ever fight a war. If one state expected another state's power to grow by enough that it could extract major concessions by force "soon", the first state would have an incentive to wage a war "now" to prevent the second state's power from growing by that much. This theory explains cases of hegemonic war in which the state that initiated military hostilities was declining in relative power and the state it attacked was rising in relative power. However, the theory fails to explain the cases in which the state that initiated hostilities was rising in relative power and the state it attacked was declining. These cases are anomalies for the theory.

## 2.7 Conventional Realism: Its anomalies are due to its faulty core assumption

This review has covered all of the main theories of hegemonic war due to shifts in relative power developed by scholars working in the conventional realist paradigm. Each theory suffered from anomalies and required an auxiliary hypothesis or loosening of the coding rules to save it. From a Lakatosian perspective, these maneuvers were acceptable as long as no one had developed an alternative theory that could explain the evidence these theories had explained *and* the anomalies they failed to explain *without* recourse to such maneuvers. The Contractual Realist theory presented in this study accomplishes this, however. Looking back from this new vantage, the series of theories created by the conventional realists appears to be scientifically degenerative. It is worth listing their anomalies in one place to clarify what they fail to explain:

- balance of power theory fails to explain the cases where war erupted without any prior shift in relative power (where relative power is defined as territorial expansion or differential long run growth rates);
- defensive realism fails to explain why *each* of the opposing states viewed itself as the status quo state and the opponent as the revisionist state;
- offensive realism fails to explain the cases where a state engaged in territorial expansion without any prior shift in its relative power (defined as long run population or productivity growth higher than that of other states);
- structural realism fails to explain the cases where one state expanded territorially but other states failed to form a balancing coalition or even bandwagoned with the aggressor;
- neoclassical realism resorts to auxiliary hypotheses about domestic political factors that shape perceptions of relative power and responses to shifts in power;
- the bargaining theory of war fails to explain long and costly wars;
- the theory of war as a commitment problem related to relative power fails to explain the cases where war was initiated by the state that was rising in relative power, not the state that was declining relative power.

This list of explanatory failures and salvage efforts raises a natural question. What is it about the conventional realist paradigm that undermines its explanatory power?

This paradigm assumes that states divide valuable assets such as territory and political authority according to relative power at all times, not only in war but in peacetime as well. That is why, according to this paradigm, states are so concerned about relative power in peacetime. If relative power shifts by too much too fast in peacetime, then the declining state will become vulnerable to expropriation by the rising state before having a chance to stop it. If there is uncertainty about relative power shifts in peacetime, then one state can bluff that its power has increased by more than it really has in order to extract concessions that it lacks the power to secure by force with a net gain. And other states will cave into such expropriation out of the fear that the expropriator actually has enough power to secure its demands by force if opposed. These concerns are logical *under the assumption* that states divide valuable assets according to relative power in peacetime as well as in wartime. Yet however logical these concerns seem to be, the theories built on these concerns fail to explain the key facts of the hegemonic wars in Europe since 1550. Is there an alternative assumption that is equally logical but better able to explain the key facts?

## 2.8 Contractual Realism: A new core assumption about how states divide valuable assets in peacetime

The conventional realist assumption is that states divide valuable assets according to relative power even in peacetime. In fact, when states divide assets by this method, it provokes costly competitions in raising relative power (Rowe 1989, 1990). For each state has an incentive to invest resources in raising its power to shift the terms of division of the asset in its favor (to gain a greater share of the asset). When other states invest in raising their power too, the result is that relative power remains unchanged. So each state incurs the cost of its investment in power without gaining the expected benefit of a boost in its relative power. The only time that states remain willing to bear such net losses is when they have no alternative method of dividing assets amongst themselves in peacetime. If they have an alternative method, then they have a strong incentive to use it: To avoid the net losses that accrue from such competitions in raising relative power.

States can avoid such competitions by dividing valuable assets amongst themselves according to anchor contracts. These are agreements whose terms are tied to anchors in the real world that do not move when relative power shifts (Rowe 1989, 1990). By dividing assets through anchor contracts, states eliminate the incentive to invest in raising relative power. For even if a state were to make such investments, this would not shift the contract terms in its favor. The terms are tied to anchors that do not move even when relative power shifts. By eliminating the incentive to make investments in power, states avoid costly competitions in raising relative power, thereby saving considerable resources.

This alternative method of dividing assets is only effective, however, when states have a way of enforcing the anchor contracts that does not rely on power based enforcement tools. For if those were the only enforcement tools available, reliance on them would provoke a costly competition in raising relative power. Hence states need an alternative way of enforcing anchor contracts. One alternative is a *justice bargain*. This enforcement method is based on the fact that the real world is changing all the time, so the anchors in the real world move randomly due to forces beyond anyone's control. Each state commits that it will allow the other state to receive a greater share of the asset in question when an anchor moves in its favor – as long as the movement was due to random forces beyond its intentional control.

But each state withholds this random variable benefit if the other state is found guilty of moving the anchor in its favor intentionally. This bargain gives each state an incentive not to move the anchors in its favor intentionally, that is, an incentive to comply with the anchor contracts. This compliance incentive does not require power based enforcement tools to be effective. But it does require a justice system of some kind to distinguish between movements of the anchors that were intentional and movements that were due to random forces beyond anyone's control. Thus, the anchor based method of dividing assets among states is only effective when there is a justice bargain and a justice system to support the anchor contracts.

This theory explains why the legal incompatibility problem is the main cause of conflicts and instability among states. Because the real world is changing all the time, the anchors move randomly due to unintentional factors. Sometimes they move randomly in a way that gives two states the right to the same asset – a legal incompatibility problem. When this happens unintentionally on either state's part, neither state can allow the other state to have the asset, because that would give the other state an incentive to create similar incompatibilities intentionally in the future. When each state stands firm for this reason, its intransigence appears to the other state as evidence that it created the current incompatibility intentionally. In this kind of situation, the justice system can no longer serve its function of distinguishing between intentional and unintentional movements of the anchors. Legal incompatibility problems render justice systems ineffective, and in doing so they leave states without an anchor based method of dividing valuable assets in peacetime. In these situations states revert to power based methods of dividing valuable assets. This generates instability and conflict.

## **2.9 Contractual Realism explains the anomalies left by the previous theories (as well as the cases they explained)**

According to the Contractual Realist theory, each hegemonic war was caused by a legal incompatibility problem that arose between two great powers. Because each problem was a novel one that international institutions had never encountered and were not designed to solve, it provoked an intractable disputes between the two states. Each state perceived that the only way it could defend its rights in the dispute from the opposing state was to switch from anchor based enforcement to power based enforcement. Each state employed military force for this purpose and conquered whatever foreign territory it needed to hold simply to defend its rights in the dispute.

This theory explains the facts about hegemonic war that conventional realism explained as well as the anomalies it did not explain.

Contractual Realism explains what balance of power theory failed to explain: Why some hegemonic wars erupted without any prior shift in relative power of any kind. According to Contractual Realism, conflict is provoked by legal incompatibility problems even without any prior power shifts.

The Contractual Realist theory explains what defensive realism failed to explain: Why each side in each war perceived itself as the status quo state and the opponent as the revisionist state. According to Contractual Realism, there is only one type of state. This type responds to perceived violations of its rights using anchor based enforcement methods if possible, but power based enforcement methods if necessary. When a legal incompatibility problem gives *each* state the perception that its rights have been violated by the other state

– and the problem cannot be solved by anchor based methods – each state responds with military force. This approach explains the fact that *each* of the opposing states in each hegemonic war perceived itself as the status quo state and the opponent as the revisionist state.

Contractual Realism explains what offensive realism failed to explain: Why states developed revisionist aims without any prior shifts in relative power. According to Contractual Realism, each of the opposing sides in each war conquered the foreign territory it needed to hold to defend its rights in the dispute over the legal incompatibility problem. The other side perceived those conquests as revisionist because the conquests were used to violate its rights.

Contractual Realism explains what structural realism failed to explain: Why states failed to balance in some cases and evened bandwagoned in some cases. According to Contractual Realism, the legal incompatibility problem that drove each war was created, itself, by an underlying spillover problem across international borders. One state was engaging in a novel domestic activity that increased its income but simultaneously emitted negative spillover effects that reduced another state's income. Although the activity was within the first state's legal rights, the negative spillover effect violated the second state's legal rights. The first state claimed the right to continue the domestic activity, but the second state claimed the right to a cessation of that activity because of its negative spillovers. Thus both states had the right to the same asset (e.g. the activity in question). This legal incompatibility problem gave each state an incentive to wage war simply to secure its rights, even without any prior shift in relative power (defined in terms of territorial expansion or long-run growth rates).

At the same time, the spillover problem shifted the balance of *short run* productivity power between the opposing states (in five of the seven cases of hegemonic war). This is a different kind of power shift than the kinds identified and theorized by conventional realism. So it is worth defining this type of power shift clearly. I define a state's power as the number of men it has an incentive to devote to military labor given the productivity of its military labor relative to the productivity of its civilian labor.

If a state's military labor productivity *increases* or its civilian labor productivity *decreases*, this increases the number of men it has an incentive to devote to military labor and increases its power as a result. Conversely, if its military labor productivity *decreases* or its civilian labor productivity *increases*, this decreases the number of men it has an incentive to devote to military labor and decreases its power as a result.

Such changes in a state's power cause shifts in relative power among states. These are what I call "shifts in short run productivity power". Such shifts explain a wide range of facts about hegemonic war that are not explained by any of the previous theories using their definitions of relative power. In particular, such shifts explain how and why the territorial holdings of the great powers shifted back and forth widely and repeatedly during each hegemonic war. Such shifts also explain why some states failed to balance or even bandwagoned.

Contractual Realism thus explains what neoclassical realism could only explain by adding auxiliary hypotheses about domestic political factors that affect perceptions of relative power and responses to it. The Contractual Realist theory contains no such auxiliary hypotheses.

The theory assumes that each state's behavior is tightly constrained by its short run productivity power relative to that of other states. This is the *sole* binding constraint on each state's choice of the number of men it will devote to military labor. Different types of spillover problem had different effects on short run productivity power in both the state emitting the spillover effects and the state negatively affected by them. Hence the short run productivity constraint had different effects on each state's choice of the number of men to devote to military labor. This choice, in turn, constrained each state's choices on other policies that depended on military labor. The short run productivity constraint thus explains many state policy choices both across states in each war and across wars. It explains state decisions:

- whether to expand or contract territorially;
- whether to balance or bandwagon against the opposing state(s);
- whether to balance immediately or only after finding an ally;
- whether to form permanent or temporary alliances; and
- whether to form alliances through coercion, compensation or debt-financing.

By explaining these state behaviors without auxiliary hypotheses about domestic political factors, the Contractual Realist theory offers "excess corroborated content" in the Lakatosian sense.

The short run productivity analysis also proves that each of the hegemonic wars between 1550 and 1950 was caused by a legal incompatibility problem *alone*, regardless of whether or not there was a prior shift in relative power (of any kind). In five of the seven cases, the legal incompatibility problem caused a shift in short run productivity power prior to the outbreak of the war. In the other two cases, the legal incompatibility problem did not cause any shift in short run productivity power prior to the outbreak of the war. In all seven cases, war erupted and lasted for years. This fact shows that each state's incentive to initiate and prolong a war arose solely from the legal incompatibility problem.

Contractual Realism explains what the bargaining theory of war failed to explain: Why each hegemonic war was so long and costly. According to Contractual Realism, each of the opposing states perceived that it needed to wage war for long enough to impose more costs on the opponent than it would gain from a compromise settlement, or else it would have an incentive to create such legal incompatibilities intentionally in the future.

Contractual Realism explains what could not be explained by the hypothesis of war as a commitment problem related to relative power: Why some hegemonic wars were initiated by the state that was rising in relative power, not the declining state. Each hegemonic war was initiated by the state that was losing income from the spillover problem. In some cases this state was declining in short run productivity power (due to the spillover problem). In other cases this state was rising in short run productivity power (due to the spillover problem). Either way, the state losing income from the spillover problem perceived that it needed to initiate war simply to stop those income losses and reverse them, regardless of its relative power status. Moreover, the fact that shifts in short run productivity power explain so many other state policies (as listed above) proves that this was the constraint on state policy choices that was binding in the states' own eyes. So it is the form of power shift that should be used to test the hypothesis of war as a commitment problem related to relative power. The observation that the war-initiating state was rising in this form of power in some cases but declining in this form of power in other cases disproves that hypothesis as a general explanation of these wars, in light of the availability of an alternative hypothesis

that explains all seven wars independently of any power shift.

In sum, the Contractual Realist theory explains all that the conventional realist theories explained and all that they failed to explain – and more. It thus offers the opportunity for a critical test.

### **3. Explanations of stability and instability at the international level**

Each hegemonic war was terminated by a peace settlement among the major belligerents. Some of these settlements were stable but others were unstable. To analyze this variation, I define instability as an equilibrium in which states commit unilateral asset seizures in violation of existing agreements on the distribution of assets between states. Stability, by contrast, is an equilibrium without such seizures.

#### **3.1 Previous explanations of stability and instability in Europe since 1500**

The leading explanations focus on the distribution of power among the largest states in the system and its effect on power based enforcement mechanisms.

- **Balance of power theorem.** Stability depends on the maintenance of a balance of power among the largest states in the system. Each state must maintain enough power to deter another state from committing aggression (either alone or with the help of allies). If any state becomes much more powerful than the others, then power based deterrence fails and instability results. This theorem has generated a number of theories:
  - ancient realist theory
  - classical realist theory
  - structural realist theory
  - neoclassical realist theory
  - defensive realist theory
  - offensive realist theory.
- **Preponderance of power theorem.** Stability depends on one state having a preponderance of power over the others. This state must maintain its preponderance in order to deter the others from committing aggression. If any of them gains enough power to get into a balance with it, then power based deterrence fails and instability results. This theorem too has generated a number of theories:
  - leviathan theory
  - power transition theory
  - hegemonic stability theory
  - long cycle theory

- unipolar stability theory
- international hierarchy theory.

The two rival theorems generate the opposite implications for the expected patterns of evidence on hegemonic war settlements. The balance theories imply that a balance of power among states should be associated with stability and a preponderance of power in one state should be associated with instability. Conversely, the preponderance theories imply that a preponderance of power in one state should be associated with stability while a balance of power among states should be associated with instability. Each theory explains some of the evidence on hegemonic war settlements but is contradicted by the other evidence.

Table 30 displays the six hegemonic war settlements in Europe between 1500 and 2000. Three were stable and three were unstable. The first stable settlement, the Peace of Westphalia, was associated with a balance of power among the former belligerents. This case supports the balance theories but contradicts the preponderance theories. It is an anomaly for those theories. The second and third stable settlements were associated with a preponderance of power in one of the former belligerents. These cases support the preponderance theories but contradict the balance theories. They are anomalies for those theories.

Now consider the unstable settlements in table 30. The first unstable settlement, the Peace of Augsburg, had three distinct phases of instability. The first two phases, from 1555 to 1600, were associated with a preponderance of power in one of the former belligerents. These cases contradict the preponderance theories (which said that preponderance yields stability), but support the balance theories (which said that preponderance provokes instability). The third phase, from 1600 to 1618, was associated with a balance of power. This case contradicts the balance theories (which said that balance yields stability), but supports the preponderance theories (which said that balance provokes instability). Thus the first two phases of the Peace of Augsburg are anomalies for the preponderance theories while the third phase is an anomaly for the balance theories.

The second unstable settlement listed in the table, the Peace of Utrecht, was associated with a balance of power between Britain and France – the two main belligerents in the preceding wars over the Protestant succession in England. This case contradicts the balance theories (which said that balance generates stability), but supports the preponderance theories (which said that balance provokes instability). This case is thus an anomaly for the balance theories. Finally, the third unstable settlement, the Peace of Versailles, was associated with the emerging preponderance of Germany. This case contradicts the preponderance theories (which said that preponderance produces stability) but supports the balance theories (which said that preponderance produces instability). Thus the Peace of Utrecht is an anomaly for the balance theories while the Peace of Versailles is an anomaly for the preponderance theories.

In sum, each of the two rival theorems explains some of the patterns of stability and instability but is contradicted by the other patterns.

### **3.2 Contractual Realism explains the anomalies left by these theories (as well as the cases explained by them)**

Table 30 displays the same hegemonic war settlements with additional categories of causes and effects. According to Contractual Realism, the settlements in the first row were stable

because they were governed by anchor contracts and anchor based enforcement. These methods divided valuable assets among states on terms that were invariant to shifts in relative power. So no state had an incentive to raise its relative power to shift the terms in its favor to gain a greater share of the assets. In particular, no state had an incentive to commit unilateral asset seizures to increase its relative power at another state's expense. The result was stability.

The settlements in the second and third rows were unstable because anchor based enforcement was rendered ineffective by legal incompatibility problems, forcing states to revert to power based enforcement. This reversion provoked a competition in raising relative power. States had to commit unilateral asset seizures simply to keep up in the competition, resulting in instability.

The pattern of competition for relative power depended on the type of legal incompatibility problem and the intractable disputes it provoked. In disputes over *economic* exclusionism, each of the opposing sides had an incentive to establish unipolar predominance over the other side. This incentive resulted in wide and decisive swings in relative coalition size that enabled one side to maintain a dominant position over the other for some time (e.g. years). In disputes over *political* exclusionism, by contrast, each side had an incentive to force the other side to capitulate, but then back off rather than impose dominance over it. This incentive resulted in narrow and indecisive swings in relative coalition size that allowed the side that had just capitulated to get back on the offensive quickly and turn the tables. The outcome was an unstable balance: Repeated swings in the balance of power where each period of capitulation lasted a short time (e.g. weeks or months).

Contractual Realism thus explains what the previous rival theorems could not explain about the *unstable* settlements: Why a preponderance of power failed to generate stability in three of them and a balance of power failed to generate stability in the other two. Neither a preponderance of power nor a balance of power can generate stability in the absence of anchor based enforcement. In its absence states must revert to power based enforcement, which is inherently unstable because it provokes competitions in raising relative power and unilateral asset seizures.

Contractual Realism also explains what the previous rival theorems could not explain about the *stable* settlements: Why one of them featured a balance of power while the other two featured a predominance of power. According to Contractual Realism, once anchor based governance became effective in each settlement, it widened the range of relative power that was consistent with stability by a significant margin. This wide range could tolerate either a balance of power or a predominance of power without leading to instability.

By explaining the anomalies left by the previous rival theorems as well as the cases they explained, Contractual Realism offers the opportunity for a critical test.

### **3.3 The dichotomy between stability and instability v. the dichotomy between peace and war**

Instability is defined as unilateral asset seizures. Such seizures may provoke the victim to retaliate with military force in a way that escalates into war – or not. If there is no retaliation, then there is simply instability without war. In this sense, instability is different from war. From another angle, though, instability is a part of war. The opposing sides in a war are not firing their weapons at each other continuously – 24 hours a day – for the entire war. Yet



they do stockpile power based enforcement tools and jockey for power continuously (from the eruption of the war to its final termination). In the course of jockeying for power they commit unilateral asset seizures in the effort to raise relative power at each other's expense. Some of these seizures provoke immediate retaliation and a battle on the spot, but others do not. Thus the time period between the initial eruption of the war and its final termination includes both periods of instability – unilateral asset seizures that do not provoke immediate retaliation – and periods of firing weapons at each other. Moreover, even in the midst of firing their weapons at each other, they also take ground from each other against each other's will. These are unilateral asset seizures *within* battle. From this perspective, a "war" is simply a series of unilateral asset seizures, some at times when the opposing sides are firing their weapons at each other, some not. Thus the periods of firing weapons are a *subset* of the instability periods within the war.

This perspective raises the standard for what constitutes a good explanation of "war". It is not simply a matter of explaining why the opposing sides are firing their weapons at each other (when they are). It is also a matter of explaining why the opposing sides are stockpiling power based enforcement tools, jockeying for power, and committing unilateral asset seizures. One might assume that they engage in these activities simply to be able to fire their weapons at each other. But this is tautological. It begs the question of why they are firing their weapons at each other. A good explanation must explain all four components of a war: stockpiling, jockeying, instability *and* firing. To do this, it helps to look at other periods in European history when there was stockpiling, jockeying and instability, but no firing – the unstable settlement periods listed in table 30.

### **3.4 Similarities between the hegemonic wars and the unstable settlements after some hegemonic wars**

Contractual Realism's explanation of the unstable settlement periods is the same as its explanation of the hegemonic wars. Both were due to legal incompatibility problems that disabled anchor based governance and unleashed competitions in raising relative power. The patterns of cause and effect were the same in both the unstable settlement periods and the hegemonic wars. This is evident in table 30, which displays the hegemonic wars in the same terms as the unstable settlement periods were displayed in table 30. In the hegemonic wars, too, the legal incompatibility problems disabled anchor based governance and required states to revert to power based enforcement. Here too this provoked a competition in raising relative power that required states to commit unilateral asset seizures simply to keep up in the competition. Here too the legal incompatibility problems were of two types. Again the economic disputes caused wide, decisive and long-lived swings in relative coalition power that resulted in one side's preponderance. But the political disputes caused narrow, indecisive and short-lived swings in coalition power that resulted in an unstable balance. These similarities between the hegemonic war periods and the unstable settlement periods are striking and invite investigation and explanation. Contractual Realism provides a way to explain them.

## 4. Explanations of good governance and misgovernance at the international level

The purpose of each peace settlement was to divide the assets that the great powers had just waged war over: territory, political authority, minority rights, trade access, etc. Each settlement divided the contested assets into shares, gave each state the right to its share, and saddled each state with the obligation to respect the other states' shares. Each settlement also created an international institution to process disputes over the shares. The settlements varied in the degree to which strong states respected weak states' rights in the following years. To analyze this variation, I define *good governance* as an equilibrium in which strong actors respect weak actors' rights. Conversely, *misgovernance* is an equilibrium in which strong actors violate weak actors' rights.

The leading explanations of good governance at the international level fall into three categories. Conventional realist theories focus on power based enforcement. Conventional institutionalist theories focus on collective boycotting as an enforcement method. Conventional constructivist theories focus on the internalization of norms as an enforcement method. When applied to hegemonic war settlements in Europe from 1500 to 2000, each set of theories encounters anomalies. Contractual Realism explains these anomalies as well as the cases explained by these theories.

### 4.1 Power based enforcement (I): Theories of force requirements

These theories focus on the level of relative power that is needed to enforce compliance with a peace settlement.

- **Power maintenance theorem.** Each state must *maintain* enough power to punish another state for violations using power based enforcement tools. If any states fail to maintain enough power, then they will be too weak to punish violators of their rights, so they will be vulnerable to violations by stronger states (e.g. the ones that did maintain enough power).
- **Power reduction theorem.** The states that won the war must *reduce* their power by enough to reassure the losing states that the winners will not exploit the power advantage with which they won the war to violate the losers' rights. Then the losers will commit to the settlement and comply with it. Otherwise they will not commit, or commit but not comply, simply out of fear of being taken advantage of by the winners.

These rival theorems generate the opposite implications for paired comparisons between two settlements following two similar wars. According to the power maintenance theory, the settlement in which states maintained comparatively *more* power based enforcement tools is more likely to produce a good governance equilibrium. Conversely, the power reduction theory implies that the settlement in which the winning states maintained comparatively *less* power based enforcement tools is likelier to produce a good governance equilibrium. Neither theory can explain all of the variation in governance outcomes in the hegemonic war settlements between 1550 and 1950.

The power maintenance theory receives only mixed support from the evidence on these settlements. The theory is supported by one paired comparison but contradicted by two

other paired comparisons. The theory receives support from a paired comparison between the Versailles Settlement (1919-39) and the Bretton Woods Settlement (1945-71). The key winning state – the United States – maintained comparatively more power based enforcement tools during the Bretton Woods settlement years. And the rights of weak states were more respected during that settlement than during the Versailles settlement, supporting the theory. The theory is contradicted, however, by a paired comparison between the Utrecht Settlement (1713-45) and the Vienna Settlement (1815-53). The key winning state – Britain – maintained comparatively more power based enforcement tools during the Utrecht settlement years. But the rights of weak states were less respected during that settlement than during the Vienna settlement. These two settlements are thus anomalies for the power maintenance theory. The theory is also contradicted by a paired comparison between the Peace of Augsburg (1555-1618) and the Peace of Westphalia (1648-1672). Each of the opposing camps, Catholics and Protestants, maintained comparatively more power based enforcement tools during the Prague Settlement years. But the rights of weak states were less respected during that settlement than during the Westphalia settlement. Thus these two settlements are also anomalies for the power maintenance theory.

The power reduction theory assumes, by contrast, that good governance in the postwar settlement years depends on the winning states *reducing* their power after military hostilities end to reassure the losing states that their rights will be respected. This theory receives no support from the evidence on hegemonic war settlements between 1550 and 1950. All of these settlements are anomalies for the theory.

The first anomalies are evident in a paired comparison between the Versailles Settlement (1919-39) and the Bretton Woods Settlement (1945-71). The winning states reduced their power significantly during the Versailles Settlement, but less significantly during the Bretton Woods Settlement. So the rights of weak states should have been more respected during the Versailles Settlement years, according to the power reduction theory. In fact the rights of weak states were *less* respected from 1919 to 1939 than they were from 1945 to 1971. Thus both cases are anomalies for the power reduction theory. Defenders of the theory might object that it includes an additional hypothesis about the need for an international institution to make a postwar settlement effective. Yet each of these settlements created an international institution that was intended to make it effective. The theory's main logic fails to explain why the institution failed in the case where the winning states reduced their power significantly, but succeeded in the case where the winning states reduced their power less significantly. The theory can only be saved by adding auxiliary hypotheses to explain these anomalies.

Additional anomalies are evident in the two earlier settlements that produced good governance (in the form of respect for weak states' rights). In each settlement *both* the winning states and the losing states reduced their power so as to reassure each other that their rights – and those of weaker states – would be respected. In the Peace of Westphalia (1648), the winning states, France and Sweden, committed to withdraw their military forces from the Empire and complied with the commitment. The losing state, Austria, made the same commitment and complied with it as well. In the Peace of Vienna (1815), the winning states committed to withdraw their forces from French territory and honored the commitment. The losing state, France, committed to reduce the size of its military forces and honored the commitment as well. The power reduction theory assumes that each settlement produced respect for weak states' rights because the *winning* states reduced their power, and

this reassured the losing states that their rights would be respected so it was safe to comply with the settlement. Yet one could argue just as logically that each settlement succeeded for the opposite reason! The *losing* states reduced their power and this reassured the winning states that it was safe to reduce their power without fear that the resulting vacuum would be exploited by the losing states. Insofar as the success of each settlement is explained by this logic – or by some new logic that explains the mutual reductions in power as effects of a deeper causal factor – each settlement is an anomaly for the power reduction theory.

**Contractual Realism explains the anomalies these theories failed to explain (as well as the cases they explained)** Contractual Realism explains the governance outcomes in all six cases. The three settlements that relied on anchor based enforcement produced respect for weak states’ rights. But the three settlements that relied on power based enforcement failed to produce respect for weak states’ rights.

Contractual Realism explains what the power maintenance theory could not explain: Why settlements in which states maintained comparatively *more* power based enforcement tools produced comparatively *less* respect for weak states’ rights. These settlements were hobbled by legal incompatibility problems and intractable disputes that provoked competitions in raising relative power. Strong states had to violate weak states’ rights to keep up in the competition.

Contractual Realism explains what the power reduction theory could not explain: Why a settlement in which the winning states reduced their power significantly produced comparatively less respect for weak states’ rights. The settlement was hobbled by legal incompatibility problems and intractable disputes that provoked competitions in raising relative power. Strong states again had to violate weak states’ rights to keep up in the competition.

## 4.2 Power based enforcement (II): Theories of collective boycotting

These theories assume that order is enforced by the formation of coalitions that are large enough to impose boycotts on those who violate the order. The threat of forming such a coalition is sufficient, on this account, to deter strong actors from violating weak actors’ rights. These theories explain various types of boycotting behavior. One type is a coalition that imposes an economic market boycott (e.g. international trade sanctions). A second type is a coalition that imposes a labor market boycott (e.g. refusing to employ the violators’ labor). A third type is a coalition that imposes social ostracism on violators (e.g. suspending the norms of civility in their regard). A fourth type is a military coalition to oppose an aggressor that conquered a weak neighbor (e.g. suspending the norms of peaceful coexistence with the aggressor). A fifth type is a collective security agreement – mutual commitments to form military coalitions as necessary to oppose aggressors.

All types of collective boycotting share a common requirement. The boycotting coalition must be large enough that each member’s expected cost of participating in the boycott is *less* than its expected benefit from forcing the violator to return to compliance with the order. Provided this condition is met, each member is better off joining the boycott than standing aside and the coalition’s deterrent threat is credible. In practice this condition means that the boycotting coalition must be larger than the violators’ coalition by a significant margin (that depends on parameter values). If the boycotters cannot enlarge their coalition by this margin – or the violators enlarge their coalition by too much – then the collective boycott

method is ineffective in protecting weak actors from violators. In this sense, this method is based on relative power, regardless of the coalition's means of imposing costs on the violator (be it through military action, trade sanctions, social ostracism or some other means).

When these theories are applied to hegemonic war settlements in Europe since 1550, they encounter only anomalies. The settlements that relied on collective boycotting for enforcement produced misgovernance (e.g. strong actors violated weak actors' rights). Conversely, the settlements that relied on anchor based enforcement instead produced good governance (e.g. strong actors respected weak actors' rights). The Peace of Augsburg (1555-1618) featured a range of collective boycotting behaviors on both the Protestant and Catholic sides, yet failed to produce good governance. The Peace of Utrecht & Rastatt (1713-1745) inspired a series of international coalition formations and disintegrations intended to deter opportunism and aggression – the classical period of balance of power politics in Europe. Yet the settlement failed to produce good governance. The Peace of Versailles (1919-1939) was based on a collective security agreement yet failed to produce good governance. Conversely, the hegemonic war settlements that *did* produce good governance were *not* reliant on collective boycotting. They relied on anchor based enforcement instead. Thus all six settlements are anomalies for the collective boycotting theories.

**Contractual Realism explains all six anomalies** Collective boycotting requires that the coalition of boycotters be larger than the coalition of violators by a significant margin. This condition can be met as long as all of the third parties who are deciding whether to join the boycott are in agreement about what constitutes a violation of the order. Then they can agree it is in their interest to join the boycotting coalition rather than the violators' coalition. However, if there is deep and widespread disagreement about what constitutes a violation, then third parties may be divided about which of the opposing coalitions it is in their interest to join. In this environment neither of the coalitions is likely to gain a relative power advantage sufficient to defeat the other one decisively.

The primary cause of disagreements about what constitutes a violation is legal incompatibility problems. When one law gives one actor the right to an asset but another law gives another actor the right to the same asset, each actor perceives that the other is violating his right. And each one can point to a basis in law for his perception. This problem undermines collective boycotting as an enforcement method, because it prevents either side in the dispute from convincing third parties that it is on the "right" side of the dispute and the opponent is on the "wrong" side. So neither side can form a coalition that is larger than the opponent's coalition by the requisite margin. Each of the hegemonic war settlements that produced misgovernance was plagued by this problem.

Conversely, each of the hegemonic war settlements that produced good governance relied on anchor based methods for contracting, adjudication and enforcement. These methods avoided legal incompatibility problems *and* reliance on boycotting coalitions composed of third-party members (e.g. actors not a party to the dispute). Each settlement was based on anchor contracts – agreements whose terms were tied to anchors in the real world that did not move even when relative power shifted among the signatories. Because the world was always changing, the anchors moved randomly due to forces beyond anyone's control. These movements gave each signatory a "random variable benefit" whenever an anchor moved in his favor. This feature enabled anchor contracts to be enforced without reliance on collective boycotting coalitions. Each signatory agreed that as long as he was allowed to receive his

random variable benefit whenever an anchor moved in his favor *unintentionally* on his part, then he would accept being denied his random variable benefit whenever found guilty of moving an anchor in his own favor *intentionally*. I call this type of enforcement bargain a *justice bargain*, because it requires an adjudication institution that can distinguish between intentional and unintentional movements of the anchors. This enforcement method does not require the formation of boycotting coalitions composed of third-party members. As long as the anchors move randomly, often enough and far enough, the threat of being denied the random variable benefit by *one* opposing actor of equal size is sufficient to deter opportunists from intentionally violating the contract. This is how the successful settlements produced good governance without relying on collective boycotting. The natural, random variability of the world – codified in the anchor contracts – enabled second party enforcement without reliance on the relative power requirements of collective boycotting. By avoiding reliance on such power based enforcement tools, the settlements avoided the competitions in raising relative power that plagued the unsuccessful settlements.

### 4.3 First-party enforcement: Theories of norm internalization

These theories posit that each actor internalizes respect for other actors' rights into his utility function. Then he respects their rights simply because he derives utility from doing so, even without the threat of being punished by any other actor. When applied to hegemonic war settlements, these theories encounter a fundamental anomaly. If one actor respects another actor's rights at time  $T$  because he internalized that respectful behavior into his utility function, then why would he violate the other actor's rights at time  $T + 1$ ? If the respectful behavior was internalized, then he would also be respectful at time  $T + 1$ .

The Peace of Augsburg (1555-1618) included a mutual agreement among Catholics and Protestants to respect the norm of tolerance for religious minorities. The norm was respected in the early years of the settlement era but increasingly disrespected as time wore on. This change in behavior is anomalous for theories of norm internalization. Similarly, the Peace of Utrecht & Rastatt (1713-1745) included an agreement that France would respect the norm of monarchical legitimacy embodied in the Protestant Succession in Britain. France's King Louis XV oscillated repeatedly between respecting the norm at times and disrespecting it at other times. These changes in behavior are anomalous for theories of norm internalization.

### 4.4 Further similarities between the hegemonic wars and the unstable settlements after some hegemonic wars

As mentioned, there was a similarity between the unstable settlements and the hegemonic wars. Each was an equilibrium involving unilateral asset seizures (the definition of instability). There was a further similarity between them as well. Each was an equilibrium in which strong states violated weak states' rights. Again the common cause was legal incompatibility problems that disabled anchor based governance. In both the unstable settlement periods and the hegemonic wars, states reverted from dividing valuable assets according to anchor contracts to dividing them according to relative power. In this new environment strong states violated weak states' rights because relative power was the metric that the strong states used to decide which assets to assert control over.

## 5. Explanations of social capital formation and depletion at the international level

The hegemonic war settlements varied in their effects on social capital. Some produced social capital formation while others produced social capital depletion. To analyze this variation, I employ the standard definition of social capital as “trust”. Since trust is unobservable, however, I use three observable proxies to measure social capital formation: (1) jurisdictional integration, (2) direct dialogue, and (3) mutual expressions of social approval. I use the opposite proxies to measure social capital depletion: (1) jurisdictional separation, (2) indirect subversion, and (3) mutual expressions of social disapproval. Using these measures, three of the settlements produced social capital formation and the other three produced social capital depletion.

The leading explanations of variations in social capital formation at the international level fall into three categories. Conventional realist theories assume that social capital formation is driven primarily by the need to form military coalitions for defensive purposes – to “balance” against a state that gained too much power or threatens to do so. Conventional institutionalist theories assume that social capital formation is driven by two factors: opportunities for mutual gains from exchange through closer cooperation (even without a common enemy), and a supply of institutional designs with the capacity to reduce transaction costs. Conventional constructivist theories assume that social capital formation is driven primarily by the actions of “norm entrepreneurs” who convince others to solve externality problems by internalizing new norms of pro-social behavior.

### 5.1 Conventional realist theories

These theories assume that when states form temporary coalitions to balance against a common enemy, the coalition members experience social capital formation for the purpose of improving the coalition’s effectiveness. They integrate military command structures and jurisdictions to the extent needed for battle effectiveness. They engage in direct dialogue to improve coordination of battle plans. They offer public expressions of social approval to help each other maintain military morale and domestic support.

These theories explain one of the hegemonic war settlements in Europe since 1550. The Bretton Woods settlement included the establishment of the NATO alliance to balance against the Soviet threat. NATO members experienced all three forms of social capital formation: jurisdictional integration, direct dialogue and mutual expressions of social approval. The other five settlements are anomalies for these theories, however. A pair wise comparison of the Peace of Augsburg and the Peace of Westphalia shows a significant external threat to Protestants and Catholics in both cases. According to the conventional realist theories, there should have been social capital formation in both cases. Yet the first settlement yielded social capital depletion while the second yielded social capital formation. Thus the first case is an anomaly for these theories, and it raises the question whether the social capital formation that occurred in the second case was due to a common external threat or some other factor. A pair wise comparison of the Peace of Utrecht and the Peace of Vienna shows a significant external threat during the first settlement years but waning external threat during the second settlement years. These theories predict social capital formation during the first period and social capital depletion during the second period, but

the opposite occurred. Thus both cases are anomalies for these theories. Finally, the level of external threat to the western powers was high during both the Peace of Versailles and the Peace of Bretton Woods. So they should have enjoyed social capital formation in both cases. But there was social capital depletion in the first case, making it an anomaly for the conventional realist theories, and raising again the question whether the social capital formation in the second case was due to a common external threat or some other factor.

Contractual Realism explains what these theories could not explain: Why external threat failed to induce international cooperation and social capital formation in some cases, on one hand, yet these positive outcomes were achieved in the case that *lacked* an external threat, on the other hand. The cases where external threat failed to produce these positive outcomes are explained by the eruption of legal incompatibility problems that prevented cooperation. Conversely, in the case that lacked an external threat, states had a more fundamental incentive to cooperate, establish an integrated jurisdiction, engage in direct dialogue, and express mutual approval. By doing so they could avoid a costly competition in raising relative power *among themselves* – even without any external threat requiring them to do so.

## 5.2 Conventional institutionalist theories

These theories attribute social capital formation to (1) opportunities for mutual gains from exchange through closer cooperation and (2) institutions that enable cooperation at an affordable cost. In the case of a postwar settlement, opposing states have an opportunity to cooperate by settling the dispute that caused the war, thereby reaping the mutual gain of avoiding further war costs. This opportunity was available after each of the hegemonic wars. And each of the six settlements created an international institution that was designed to resolve disputes and facilitate enforcement of the settlement. Yet only three of the settlements produced social capital formation while the other three produced social capital depletion. The latter three cases are anomalies for these theories – and raise doubt about their account of the role of institutions in the three cases that had social capital formation. These theories assume that institutions enforce postwar settlements through one of the three conventional means discussed earlier: military force, collective boycotting or norm internalization. Each of these means was tried in each of the anomalous settlements that produced social capital depletion. This fact casts doubt on the claim that these means explain the success of the other three settlements in enabling social capital formation.

Contractual Realism explains what these theories could not explain: Why the three conventional means of enforcement failed in three of the settlements but appeared to succeed in the other three. According to Contractual Realism, the successful settlements – those that produced social capital formation – were not enforced by any of the three conventional means. They were enforced by anchor based methods. Conversely, the three unsuccessful settlements – those that produced social capital depletion – were unable to rely on anchor based methods because of legal incompatibility problems. Thus Contractual Realism explains the outcomes of all six settlements in a consistent way.



### **5.3 Additional similarities between the hegemonic wars and the unstable settlements after some hegemonic wars**

As mentioned, there were two similarities between the unstable settlements and the hegemonic wars. Each involved unilateral asset seizures and the violation of weak states' rights by strong states. A further similarity was that each involved social capital depletion. The common cause, once again, was legal incompatibility problems that disabled anchor based governance. In both the unstable settlement periods and the hegemonic wars, actors could no longer tell the difference between intentional and unintentional movements disputes. So they had to assume that all disputes were intentional simply to protect themselves from opponents' creating disputes intentionally but bluffing they arose unintentionally (due to random factors beyond anyone's control). Actors blamed each other for disputes that had arisen unintentionally on anyone's part, and as a result, innocent actors were treated as guilty parties by their fellows. In this environment social trust was impossible and the consequence was social capital depletion.

## CHAPTER 6

### **Critical tests (II): Theories of Europe's development at the domestic level**

Since the dark ages, the main regions of Western and Central Europe have experienced repeated oscillations between order and disorder. The eras of disorder were caused by a range of factors: the disintegration of Empires, the eruption of civil wars, the decadence of old rules of order, and the emergence of new religions and forms of political participation. Each period of disorder forced actors to create new institutions that could resolve the problems and restore order. These cycles of regress and progress had pervasive effects on the development of Europe's regions. The most visible effect was the formation of a modern state in each region – an institution with the capacity to collect taxes, enforce order, protect weak actors in society, hold policymakers accountable, and maintain popular legitimacy through consent rather than coercion.

The modern state developed in long stages that seem to correspond neatly with the cycles of order and disorder over the last thousand years. In the 9th and 10th centuries, the Carolingian Empire disintegrated into feudal anarchy. In the 11th and 12th centuries, the kings of England and France solved the problem of anarchy by creating the judicial and administrative monarchy – an institution based on novel techniques of law and tax administration. In the 16th century German rulers solved the problem of anarchy by importing Roman law and creating the beginnings of a central authority in each of their territories. In the 17th century rulers solved the problems of religious strife, civil war and noble rebellion by creating absolutist regimes and the beginnings of modern rational bureaucracy and uniform taxation. Rulers then solved the problem of decadent agrarian feudalism by giving peasants legal protections and access to justice in the 18th century and electoral democracy in the 19th century. This chronology appears to explain how Europe's regions transitioned from

feudal anarchy in the year 1000 to the brink of modernity in the year 1900.

This account of domestic development raises some basic questions, however. Why did strong medieval states like England and France develop parliamentary constitutionalism by the 14th century, while the strong medieval states of northern Italy developed into despotisms by the 14th century? Why did the strong medieval states of northern Italy coalesce into territorially expansive states by 1400, while the strong medieval states of Germany disintegrated into tiny statelets by 1400? Why did it take Germany until the 16th century to solve the problem of feudal anarchy, when England and France had solved it in the 11th and 12th centuries? If French administration was precociously well developed in the medieval period, then why did it descend into corrupt venal officeholding by the mid-18th century, whereas Prussian administration took the opposite path – ascending from corrupt venality in the medieval period to law-governed bureaucracy by the mid-18th-century. Why did Prussian peasants *gain* access to the legal system in the 18th-century while French peasants, who had enjoyed access since the medieval period, *lost* it in the 18th-century? Why did merchant-led government in England remain inclusive and adaptive through the early modern period, while merchant-led government in the Dutch Republic become cartelized and corrupt by the 18th-century? Why did corrupt government officeholders coalesce into a stable ruling cartel in the Dutch Republic, but disintegrate into competing factions locked in an unstable balance of power in France? These questions highlight the fact that European states followed sharply *divergent* paths to modernity.

A simple, linear chronology of long-run institutional development may capture the fundamental difference between the starting point – feudal anarchy in the year 1000 – and the end point – the modern state in the year 1900. But it leaves the diversity of *paths* to modernity unexplained. Scholars have developed a number of theories to explain variations in domestic development across European states since the dark ages. Yet they too leave the above questions unanswered – or provide answers that are plagued by anomalies.

In this study I argue that all of these questions can be answered by understanding a single causal mechanism – the same mechanism that explains developments at the international level. I also argue, however, that the simple dichotomy between feudal disorder and its institutional deficits, on one hand, and modern order and its institutional requirements,

on the other hand, fails to capture how such a variety of developmental paths was generated by this mechanism. I argue that three additional dichotomies are needed – the same dichotomies that were needed to explain how this mechanism generated developments at the international level. There is the dichotomy between domestic stability and instability, the dichotomy between good governance and misgovernance at the domestic level, and the dichotomy between social capital formation and social capital depletion at the domestic level. These dichotomies reflect the range of effects generated by this mechanism. Previous theories have attempted to explain variation within each dichotomy, but encounter anomalous cases that can only be explained by adding auxiliary hypotheses. The new theory proposed in this study, Contractual Realism, explains these anomalies as well as the cases the previous theories explained.

In the first subsection, I critique the leading explanations of the causes of stability and instability. In the second subsection, I critique the leading explanations of the causes of good governance and misgovernance. In the third section, I critique the leading explanations of the causes of social capital formation and depletion. In each subsection I also summarize how Contractual Realism explains both the cases that the previous theories explained and the anomalies they left unexplained.

## 1. Explanations of stability and instability at the domestic level

The leading explanations of stability and instability focus on the domestic distribution of power and its effect on power based enforcement mechanisms.

- **Preponderance of power theorem.** Stability depends on the state having a preponderance of power over all domestic groups. The state must maintain its preponderance in order to deter domestic groups from rivaling its authority and violating each other's rights. If any domestic groups gain enough power to get into a balance with the state, then power based deterrence fails and the result is disorder and instability. This theorem has generated several theories:
  - classical *realpolitik*
  - leviathan theory
  - strong state theory.
- **Balance of power theorem.** Stability depends on the maintenance of a balance of power among elites within the state and beyond it. Each elite must maintain enough

power to deter another elite from violating its rights (either alone or with the help of allies). If any elite becomes much more powerful than the others, then power based deterrence fails and the result is a capture of the state by the dominant elite. This theorem has generated several theories as well:

- executive-as-balancer theory
- double-balance theory
- absolutism-as-collaboration theory

The two rival theorems generate the opposite implications for the expected patterns of evidence on domestic stability and instability. The preponderance theories imply that a preponderance of power in the state should be associated with stability while a balance of power among elites within and beyond the state should be associated with instability. Conversely, the balance theories imply that a balance of power among elites should be associated with stability while a preponderance of power in the state itself should be associated with instability. Each theory explains some of the evidence on domestic stability and instability, but is contradicted by the other evidence. Contractual Realism explains both sets of evidence.

### **1.1 Preponderance of power theorem: Strong state theory**

According to this theory, the main condition for domestic stability is a kinetically strong state – one that has enough military force to defeat domestic rivals for its authority and paramilitary force to protect domestic groups from each other. The theory is supported by some evidence but contradicted by other evidence. Consider the medieval evidence and then the early modern evidence.

The theory seems to find support from the cases of England, France and Germany. In this interpretation, the Norman conquest of 1066 gave England a strong state and increasing stability. The expansion of the Capetian monarchy in the 11th and 12th centuries gave France a strong state and increasing stability. The expansion of the Ottonian monarchy in the 10th and 11th centuries gave Germany a strong state and increasing stability. But the Investiture Conflict and the switch to elective monarchy gave Germany a weak state and instability in the 12th to 15th centuries.

This interpretation is contradicted by several strands of evidence, however. After the Norman conquest of England in 1066, both the king and his magnates maintained large

armies. The monarchy could assemble coalition forces large enough to protect its own authority. Yet it often lacked the capacity or will to protect the regional magnates from each other (Morris 1927). Regional instabilities and violence over land disputes continued for the first 70 years of Norman rule (Dalton 1994; Hudson 2010). Over the next 30 years, the king and the magnates maintained even larger armies and the instabilities worsened. These patterns were reversed in the 1200s and 1300s. The king and magnates reduced their armies to historically low sizes and there was increasing stability at the national and regional levels. This English evidence suggests that a militarily strong state is associated with instability and a militarily weak state is associated with stability. The German evidence supports this conclusion. The German king and magnates maintained much larger armies than their English counterparts in the 1200s and 1300s (Arnold 1985). Yet Germany suffered instability while England enjoyed stability.

To explain these anomalies, proponents of the strong state theory offer several auxiliary hypotheses. One hypothesis claims that England was governable and stable because of its compact size, while Germany was ungovernable and unstable because of its great size. A related hypothesis claims that Germany was unstable because of its monarchy's weakness, which was due to the country's size, ethnic heterogeneity and switch to elective kingship. These hypotheses are contradicted by evidence from Germany itself. Despite its great size, Germany was governable and relatively stable from 924 to 1056 under the Ottonian and Salian dynasties. During the unstable period from 1076 to 1450, each of the regional magnates failed to establish regional stability despite the compact size and ethnic heterogeneity of his region and the hereditary basis of his rule. The hypotheses are also contradicted by evidence from England and Italy. Despite having significant ethnic and regional heterogeneity, England still enjoyed stability. Despite being more ethnically homogeneous and having a stronger state than England, Lombardy suffered instability. All of these facts contradict the auxiliary hypotheses and thus cast doubt on the strong state theory.

Turning to the early modern evidence, the strong state theory seems to find support from the cases of France and Prussia. Each country had a kinetically strong state that subdued the nobility and enforced order among rival noble factions. The main evidence is the success of both states in suppressing noble revolts and factional violence after 1650. While Prussia

remained stable through the 1700s, however, France suffered increasing instability. Although proponents of the strong state theory offer several hypotheses to explain this anomaly, each hypothesis is problematic.

One hypothesis claims that military necessity forced rulers to establish domestic political stability. This hypothesis explains the Prussian case, but fails to explain the French case. French rulers faced dire military necessity – fighting and losing five major wars between 1670 and 1770. If military necessity alone were sufficient to induce domestic political stability, France would have enjoyed stability too. A second hypothesis claims that the sale of government offices to reduce public debt was politically destabilizing. The main evidence for this hypothesis is the French fiscal crises of the 1700s that led to increasing instability. Yet the rulers of Brandenburg-Prussia had a long history of selling or leasing district offices to the nobility in return for loans to the ruler. This problem continued from the Middle Ages into the 1600s, but was solved after 1650. The strong state theory does not explain why Prussia solved this problem but France failed to solve it. A third hypothesis claims that Prussia had more effective political leadership than France. The main evidence is the skill, discipline and service ethic of the three main Prussian rulers from 1640 to 1786. Yet both Prussia and France had reformers who advised the rulers to make changes to reduce instability. The strong state theory does not explain why the Prussian rulers followed the advice and the reforms succeeded, while the French rulers either failed to follow the advice or followed it but the reforms failed.

## **1.2 Balance of power theorem**

### **1.2.1 Executive-as-balancer theory**

This theory focuses on the role of the executive arm of the state in balancing power between competing elites. When one elite group violates the rights of another elite group, on this account, the executive agency can bring its armed force to bear on the second group's side. As long as the executive has the capacity and will to join the side that needs more power to defend itself in a domestic conflict, stability is possible. Without such an executive, instability is likely. The theory implies a correlation between partisan executives and domestic stability.

Applied to the evidence from medieval and early modern Europe, the theory encounters only anomalies. In medieval Germany, rulers at all levels acted as partisan judges and military side-takers. By contrast, English rulers presided over a system of nonpartisan judges from the 12th century on. The outcomes were the opposite of those expected from the theory. Germany experienced instability while England experienced stability. In early modern France, noble estate owners and officeholders viewed judges as partisan. When nobles lost court cases they rejected the verdicts and shopped for another court with a partisan judge who would rule in their favor. Since judges earned fees for issuing rulings, they had an incentive to offer partisan rulings to attract business. This gave the nobles, in turn, an incentive to commit unilateral asset seizures in confidence they could find a partisan judge to rule in their favor. The result was pervasive instability in the distribution of legal rights to offices and estates. Cases were never resolved definitively, rights remained contested, and assets remained encumbered by disputes and rivalries between court jurisdictions. In early modern Prussia, by contrast, noble officeholders viewed the adjudication system as nonpartisan. When they lost cameral (court) cases over the distribution of offices, budgets, projects and responsibilities, they accepted the verdicts. The result was increasing stability in the distribution of rights to these assets. Cases were resolved definitively, bureaucratic rights became articulated at finer levels, and mutual confidence in the jurisdictional hierarchy inside the state bred expectations that work effort would be rewarded.

One might object that this evidence pertains to the distinction between partisan and nonpartisan *judges*, while the theory focuses on the role of the executive as a *military* balancer between rival elites. Empirically, the two roles ran together though. In the cases where judges were partisan, the military arm of the state became a partisan side taker and the result was instability. In the cases where judges were nonpartisan, the military arm of the state remained nonpartisan and the result was stability. Thus both sets of cases are anomalies for the theory.

### 1.2.2 Double-balance theory

This theory focuses on the distribution of economic resources and political power among elites. When the two distributions are aligned, on this account, there is a "double balance"



that enables domestic peace and stability. To achieve this alignment, a subset of the elites need to form a "dominant coalition" that concentrates the society's resources under its control by politically reducing access to the resources by other elites. By concentrating the resources under its control, the dominant coalition ensures that each of its members receives a peacetime benefit stream *greater* than his expected benefit from breaking the peace and starting a private war with another member of the coalition. Conversely, if the dominant coalition loses or relinquishes control of too many of its resources, then each member's expected peacetime benefit will fall *below* his expected benefit from making war on another member. The theory's main implication is that resource concentration and stability will be positively correlated. Societies with a high concentration of resources in a subset of the elite are more likely to experience stability than societies where resources are diffused across all of the elites.

When applied to the medieval evidence, the theory is supported by two cases but contradicted by three other cases. The evidence from Germany supports the theory. A series of conflicts over religion and royal succession in the 12th century resulted in the diffusion of society's resources through the ranks of the military aristocracy until local knights became virtual sovereigns in their castles. The result was instability that lasted until the 15th century. The evidence from the states of Northern Italy contradicts the theory. The tightening of guild restrictions produced a concentration of resources in a dominant coalition within each state. The effect was a proliferation of conflicts among elites within each dominant coalition that yielded ongoing instability. This pattern of instability and resource concentration is an anomaly for the theory.

The remaining cases are phases of medieval English history. The first phase began in 1066 with the Norman conquest. It produced a concentration of resources within a dominant coalition of regional magnates. There was violence and instability for decades after the invasion – far longer than can be explained by the invasion and establishment of the new monarchy in 1066. This pattern of resource concentration and instability is an anomaly for the theory. The second phase began in 1135 with a dispute over the royal succession that descended into civil war. Resources became diffused through the ranks of the military aristocracy until, as in Germany, local knights became virtual sovereigns in their castles. This pattern of

resource diffusion and instability supports the theory. The third phase of medieval English history began in 1165 with the settlement of the civil war. The settlement included new legal provisions that would protect military tenants – the knights – from having their land seized by their own lords – the regional magnates. These provisions resulted in the diffusion of land ownership rights *in peacetime* from the dominant coalition of magnates to the local knights. With more ownership security came stability, and with stability came an incentive for the knightly class to pursue a fundamental investment transformation: Disinvesting from military capital and investing in productive economic capital. This transformation caused a further diffusion of resources through the ranks of knightly society that reinforced stability. This correlation between resource diffusion and increasing stability over time is an anomaly for the theory. The theory predicted that if the dominant coalition relinquished control of too many of its resources, then its members would expect less income from remaining at peace than from making war on each other.

### **1.2.3 Absolutism-as-collaboration theory**

This theory claims that it is misleading to focus solely on the capacity of absolutist regimes to crush noble revolts by military force. A closer examination shows that the regimes obtained the nobility's daily cooperation through compensation rather than coercion. Absolute rulers compensated nobles for their compliance with central authority by granting them positions of status at court, ownership rights to government offices and revenues, and appointments in the military officer corps. On this account, the domestic stability of absolutist states – when they enjoyed it – was due to these mutually beneficial exchanges rather than to coercive threats. This theory fails to explain why Prussia enjoyed stability in the 1700s but France suffered instability, however.

While several hypotheses have been offered to explain this anomaly, each one has problems. One hypothesis claims that domestic instability is caused by religious cleavages. Both France and Prussia had religious cleavages, however. Prussia was able to transcend the cleavage between Lutherans and Calvinists, while France was destabilized by the cleavage between Romanist and Gaulist Catholics. This difference is not explained by the religious cleavage hypothesis. A second hypothesis claims that a religion with a strong element of

social discipline can produce domestic stability. The main supporting evidence is the positive role of Calvinism in both Prussia and the Dutch Republic. Both states developed greater bureaucratic efficiency with the help of Calvinist social discipline. However, Prussia enjoyed political stability while the Dutch Republic suffered political instability. This difference is not explained by the social discipline hypothesis. A third hypothesis claims that the political instabilities in France and the Dutch Republic were due to factional infighting among the nobility in each state. There was also factional infighting within Prussia, however. This hypothesis does not explain why the infighting settled down in Prussia but expanded and became destabilizing in France and the Dutch Republic.

### **1.3 Contractual Realism explains both sets of cases: The cases explained by the previous theories and the anomalous cases**

According to Contractual Realism, the key causal factor was the capacity to resolve legal incompatibility problems. In the cases where this capacity developed, actors could rely on anchor based governance. In the cases where this capacity failed to develop or decayed, actors had to rely on power based governance. This difference explains all of the cases – both those explained by the previous theories and those that were anomalies for those theories.

#### **1.3.1 Contractual Realism explains what the preponderance of power theories could not explain**

Contractual Realism explains what the *strong state theory* could not explain about the medieval period: Why kinetically strong states experienced instability while kinetically weaker states experienced stability. The explanation is summarized in table 30. In the medieval period the states in Germany, Northern Italy and England (up to 1180) were hobbled by legal incompatibility problems that disabled anchor based governance. Disputes became intractable and actors were forced to revert to power based governance. They invested heavily in power based enforcement capital and employed it to defend their rights in the disputes. The drive to defend rights against encroachment unleashed a competition in raising relative power. Actors had to commit unilateral asset seizures against each other simply to keep up in the competition. The outcome was pervasive instability. In contrast, the states in France

and England (after 1180) developed the capacity to resolve legal incompatibility problems and relied on anchor based governance. Assets were distributed among actors on terms that were independent of relative power. This reduced the incentive to compete for power and created an incentive to undergo a fundamental investment transformation. Investment in power based enforcement capital was relatively low in these states compared to the first set of states. Yet stability was relatively high because of the anchor contracts.

Contractual Realism also explains what the *strong state theory* could not explain about the early modern period: Why Prussia enjoyed stability while France suffered instability even though both were kinetically strong states. France was hobbled by a fundamental legal incompatibility problem. The monarchy held that it had the *exclusive* right to make the laws, but the parliaments held they had a right to *participate* in the making of the laws. This incompatibility disabled the system of anchor based governance inherited from the medieval period and forced actors to revert to power based governance. The competition to raise relative power caused instabilities that lasted until 1789. In contrast, Prussia developed the capacity to resolve legal incompatibility problems and relied on anchor based governance that generated stability.

This account explains what the venal officeholding hypothesis could not explain: Why France descended into venal officeholding and domestic instability by the 17th-century, while Prussia emerged from the venal system of the medieval period to establish meritocratic bureaucracy and stability by the 18th-century. In each case both outcomes were driven by a deeper cause. In France, legal incompatibilities provoked a competition in raising relative power that forced the opposing sides to build bigger coalitions simply to keep up in the competition. Each side financed its coalition building efforts in a way that compounded the problem of venal officeholding and made it intractable. The monarchy sold more offices to finance the defense of its exclusive right to make the laws. The officers and magistrates charged petitioners higher taxes and fees to finance the defense of their right to participate in the making of laws. Reform of the venal system was impossible because of the underlying legal incompatibility problem. Each side perceived that if it made any concessions on its rights to the opposing side, that would set a precedent that would require it to make further concessions on similar issues *ad infinitum*. In contrast, Prussia relied on anchor based gov-

ernance that eliminated the need for competitive coalition building and divisive methods of coalition finance.

This account explains what the leadership quality hypothesis could not explain: Why French rulers failed to follow the advice to reform or followed it but the reforms failed, while Prussian rulers followed such advice and the reforms succeeded. Prussian rulers and administrators were, above all, good adjudicators who appreciated the need to enact reforms that would improve adjudication protocols at all levels of governance. These reforms created a high institutional capacity to adjudicate governance disputes at all levels – from the ruler’s chamber down to the provincial district committees. As a result, leaders at all levels were able to shape policies in the common interest, achieve their adoption by consensus, and manage implementation disputes cooperatively. The three great Prussian rulers from 1640 to 1786 were effective leaders, above all, because their administrators and agencies had a high capacity for adjudication. Conversely, French rulers were ineffective because the legal incompatibility problem dividing the monarchy and the parliaments hobbled adjudication at all levels of governance.

This account explains what the military necessity hypothesis could not explain: Why Prussia enjoyed stability but France suffered instability even though both states faced great external military pressure. Military pressure only induces effective leadership when administrators and agencies are unhobbled by legal incompatibility problems.

### **1.3.2 Contractual Realism explains what the balance of power theories could not explain**

Contractual Realism explains what the *executive-as-balancer theory* could not explain about medieval adjudication and stability: Why *partisan* judges and executives were associated with instability in Italy and Germany, while *nonpartisan* judges and executives were associated with stability in England. In each case both outcomes were driven by a deeper cause. In England, the adjudication rules could distinguish between intentional and unintentional disputes, thereby enabling anchor based governance. Once actors appreciated that the system saved them the costs of private wars over land disputes, they demanded that the government supply rules to guaranty the nonpartisanship of judges. In Germany, by contrast, the adju-

dication rules could not distinguish between intentional and unintentional disputes, making anchor based governance impossible. Once actors reverted to power based governance, they demanded partisanship from the judges – who were often armed lords themselves – simply to increase the supply of potential allies in the competition to raise relative power. In Italy, the guild restrictions created legal incompatibility problems that hobbled the adjudication systems. So again actors averted to power based governance and demanded partisan judges in the search for allies in the competition for power.

Contractual Realism explains what the *executive-as-balancer theory* could not explain about early modern adjudication and stability: Why *partisan* judges were associated with instability in France, but *nonpartisan* judges were associated with stability in Prussia. In each case both outcomes were driven by a deeper cause. In France, adjudication was hobbled by legal incompatibility problems, not only between the monarchy and the parliaments, but also between provincial jurisdictions. With judges offering favorable rulings to increase their fee income, this gave officeholders and estate owners an incentive to search for ambiguities in the documents specifying rights to such assets in order to bring a dispute to some judge somewhere and obtain a favorable ruling. In the face of such conduct, the adjudication system could not distinguish between intentional and unintentional disputes. Anchor based governance was impossible. Actors reverted to power based governance and a competition in raising relative power was unleashed. To increase the pool of potential allies in this competition, actors demanded more partisanship from the judges. In Prussia, by contrast, the adjudication rules could distinguish between intentional and unintentional disputes, thereby enabling anchor based governance. Once administrators appreciated that the system saved them the costs of bureaucratic turf wars over offices, budgets, projects and responsibilities, this gave them an incentive to develop rules to guaranty the nonpartisanship of adjudicators in bureaucratic disputes.

Contractual Realism explains what the *double balance theory* could not explain:

- Why a high concentration of resources in the dominant elite coalition was correlated with instability rather than stability;
- Why there was instability in the cases with a high concentration of resources *and* the cases with a low concentration, yet stability in cases with a medium concentration;

and

- Why a dominant coalition's policy of diffusing resources to nonmembers, thereby reducing the resource concentration, led to increasing stability rather than instability.

A deeper causal factor was at work in all of the historical cases. This factor determined whether there was stability or instability *and* whether there was resource concentration or diffusion in each case. The key factor was institutional capacity to resolve legal incompatibility problems. Table 30 displays the effects of this factor in the historical cases. In the cases that *lacked* the capacity to solve legal incompatibility problems, actors reverted to power based governance, invested heavily in power based enforcement capital, and engaged in a competition to raise relative power. They committed unilateral asset seizures simply to keep up in the competition, resulting in pervasive instability. These cases fall into two categories. In cases dominated by disputes over *economic* exclusion-ism, there were wide and decisive swings in relative coalition power that gave one coalition or the other a preponderance of power. These cases exhibited a high concentration of resources in a single dominant coalition. In cases dominated by disputes over *political* exclusion-ism, by contrast, there were narrow and indecisive swings in relative coalition power that yielded an unstable balance of power between the rival coalitions. These cases exhibited a low concentration of resources for two reasons. The resources were diffused across the opposing coalitions; and after each coalition forced the opposing coalition to capitulate, the winning coalition disintegrated into smaller subgroups, further diffusing resources. There was instability in both sets of cases, those with high resource concentration and those with low resource concentration, because there were legal incompatibility problems that forced actors to revert to power based governance and wage competitions for relative power in both sets.

By contrast, in cases where actors developed the capacity to solve legal incompatibility problems they relied on anchor based governance. Resources were divided among actors according to anchor contracts whose terms were independent of relative power. This reduced the incentive to compete for relative power and commit unilateral asset seizures, resulting in stability. Each coalition could allow its resources to be diffused to nonmembers without fear of leaving itself vulnerable to expropriation of its remaining resources. For with resources divided on terms independent of relative power, the resource distribution was protected from

unilateral (involuntary) alteration. In consequence, the resulting resource distribution was not determined by relative power calculations for rival coalition building, but by efficiency factors such as the availability of mutual gains from economic exchange. This explains why these cases avoided the extremes of either an overly concentrated distribution of resources or an overly diffuse distribution.

Finally, Contractual Realism explains what the *absolutism-as-collaboration theory* could not explain: Why Prussia enjoyed stability but France suffered instability, even though both regimes were based on collaboration between the ruling executive and the nobility who dominated the main institutions. The collaboration theory is correct in pointing out that absolute rulers obtained the nobility's daily compliance with central authority through compensation rather than coercion. The relationship between ruler and nobility was contractual: Each side had rights and obligations. However, the collaboration theory does not address the question of contract enforcement in sufficient detail to explain variations in stability across regimes. In particular, the theory does not distinguish between two different enforcement mechanisms that yield two different equilibria – one stable, one unstable. The mechanism of collective boycotting only yields stability when each side has access to enough potential third-party allies to create a coalition large enough to deter the other side from committing violations. In constitutional issues, however, each side has at least 40% of the relevant political power or economic resources. So there are not enough potential third-party allies for either side to have a credible deterrent. This is why constitutional contracts are enforced by anchor based mechanisms rather than collective boycotting mechanisms.

If anchor based enforcement is disabled by legal incompatibility problems, then actors revert to power based enforcement methods, which provoke instability and competitions in raising relative power. In particular, the method of collective boycotting provokes competitions in factional coalition building. This account explains why France suffered instability while Prussia enjoyed stability. Anchor based enforcement remained feasible in Prussia because the adjudication systems were reformed repeatedly to avoid legal incompatibility problems. In France, though, anchor based enforcement was disabled by legal incompatibility problems.

This account explains what the factional infighting hypothesis could not explain: Why



factional infighting settled down in Prussia but worsened in France and the Dutch Republic. In the latter cases, legal incompatibility problems disabled anchor based enforcement, forcing actors to revert to power based enforcement and competitions in factional coalition building. The factional infighting could not settle down as long as the legal incompatibility problems remained unresolved. In Prussia, by contrast, legal incompatibility problems were avoided through repeated reforms of the adjudication systems. This enabled actors to rely on actor based enforcement rather than power based enforcement. Hence they lost the incentive to compete for relative power and the factional infighting settled down.

The account also explains what the social discipline hypothesis could not explain: Why Prussia enjoyed political stability but the Dutch Republic suffered political instability *despite* their common Calvinist discipline. Dutch rulers failed to reform adjudication systems in a way that would overcome legal incompatibility problems. By contrast, Prussian rulers reformed adjudication systems repeatedly to avoid legal incompatibility problems.

This account also explains what the religious cleavage hypothesis could not explain: Why religious cleavages were politically destabilizing in France but not in Prussia. In France the religious cleavage between Romanist and Gaulist Catholics created a legal incompatibility problem: A conflict of laws between papal sovereignty and French sovereignty. In Prussia the religious cleavage between Calvinists and Lutherans did not create this kind of problem, because International Calvinism did not claim sovereignty over domestic church governance in the same way that International Catholicism did. This difference suggests that the church issue in France was not politically destabilizing due to its *religious* aspect, but due to its *legal incompatibility* aspect.

### **1.3.3 Contractual Realism explains both the supporting cases and the anomalous cases**

The Prussian governance system was based on anchor contracts and anchor based enforcement. The system was managed through adjudication protocols at all levels of governance that distinguished between intentional and unintentional causes of dispute. The unintentional disputes were resolved by reforming the anchor contracts and anchor based enforcement rules. This governance method made disputes *tractable*. By contrast, the adjudication

protocols in France and the Dutch Republic were confounded by legal incompatibility problems, resulting in intractable disputes. This difference accounts for all of the variables added by the auxiliary hypotheses to save previous theories.

The intractable disputes in France and the Dutch Republic generated domestic conflicts and competitions in raising relative power. Prussia avoided such conflicts and competitions because its disputes were tractable. This difference explains why factional infighting settled down in Prussia but worsened in France and the Dutch Republic.

## 2. Explanations of good governance and misgovernance at the domestic level

Good governance is an equilibrium in which strong actors respect the rights of weak actors. In the domestic context, the primary arena for good governance is in relations between the state and society. When strong actors *in the state* respect the rights of weak actors *in society*, it is good governance. When strong actors in the state violate the rights of weak actors in society, it is misgovernance. To measure variations in governance through European history, I employ two measures: corruption rates and tax rates. Where corruption rates and tax rates on non-elites were relatively low by the standards of the time, it indicates good governance. Conversely, where corruption rates and tax rates on non-elites were relatively high by the standards of the time, it indicates misgovernance. European states varies significantly on these measures throughout the medieval and early modern periods. The main explanation of these variations in the literature focuses on **legal restraints on the executive**. Four variants of this explanation have been produced. Each variant has anomalies that required auxiliary hypotheses to save it. Contractual Realism explains the patterns of variation in governance through Europe's history without auxiliary hypotheses.

### 2.1 Medieval constitutionalism theory

On this account, the main difference between European states in the course of their development was in whether they eventually became democracies or autocracies in the modern period. Those that developed elected legislatures that imposed legal restraints on the exec-

utive became democracies. Those that failed to develop legislatures and legal restraints on the executive became autocracies. Downing explains this difference as a consequence of two factors that operated centuries earlier: Differences in institutions in the medieval period and external military pressures in the 1600s. This theory seems to explain the English, French and Prussian cases. England had strong constitutionalism in the medieval period and weak pressures for military mobilization in the 1600s, so it developed into a democracy in the modern period. Prussia had weak constitutionalism in the medieval period and strong pressures for military mobilization in the 1600s, so it developed into an autocracy in the modern period. France had strong constitutionalism in the medieval period but strong pressures for military mobilization in the 1600s, so it followed a path to democracy that was more difficult than England's path but easier than Prussia's path.

This interpretation of the evidence is convincing as long as the focus is on the **electoral** dimension of democracy. If the focus is on the **rule of law** dimension of democracy, however, then the theory fails to explain the French and Prussian cases. From 1648 to 1789 Prussia developed a rule of law state that included strong legal restraints on the executive. The result was relatively low tax and corruption rates compared to France. During the same period France suffered a steady decay of the strong constitutionalism and legal restraints on the executive it had inherited from medieval times. The result was high tax and corruption rates in the 1700s compared to Prussia. This difference between Prussia and France is not explained by any difference in pressures for military mobilization in the 1600s: both states experienced the same pressures. The difference is also not explained by any difference in medieval constitutionalism. For if that factor explained it, France would have developed *stronger* legal restraints on the executive than Prussia developed from 1648 to 1789, not weaker.

## 2.2 Local governance and military competition theory

According to this theory, European states developed two different kinds of restraints on the executive by the early modern period – and each is explained by a different factor. On one hand, the rise of *elected parliamentary* restraints on the executive – constitutionalism – is explained by whether a state had participatory local governance in the medieval period or

administrative local governance. The participatory governance led to constitutionalism in the early modern period while the administrative governance led to absolutism. On the other hand, the rise of *legal* restraints on the executive – in the form of a bureaucracy governed by law – is explained by whether a state experienced the onset of external military competition before 1500 or after. If before 1500, then its state agencies would remain structured by patrimonial customs inherited from deep in the Middle Ages. If after 1500, then its state agencies could develop into law governed bureaucracies by incorporating developments in legal and administrative science since 1500.

This theory seems to explain both the similarities and the differences between France and Prussia in the 1700s. Both states had *administrative* local governance centuries earlier, so both states developed into absolutist regimes by 1650. The two states differed in the timing of the onset of external military competition, however. It started before 1500 in France, so state agencies remained mired in patrimonial custom through the 1700s. The competition started after 1500 in Prussia, by contrast, so it developed a comprehensive system of bureaucratic law by 1789.

There are several problems with this interpretation of the historical record. First, the distinction between participatory and administrative local governance is not well defined. In medieval France local governance had both participatory and administrative elements – as it did in England (the main example of the participatory type according to this interpretation). So it is not clear from the local governance data why France turned toward absolutism after 1500 while England turned toward constitutionalism. Second, even if French local governance is coded as “administrative” in the medieval period, the fact remains that France had central and provincial parliaments in both the medieval and early modern periods. Their main function was always to restrain the executive by participating in the processes of making and enforcing the laws that governed society. Thus the question is: Why were the parliaments less able to restrain the executive in the 1700s than earlier in their history? If their failure was due to the nature of local governance in the early to high medieval period – 500 to 1000 years earlier – then why did this factor take so long – until the 1700s – to hobble the parliaments in their traditional function of restraining the executive? Isn’t it possible that the parliaments were hobbled by a more recent and powerful force instead.

Third, the timing of the onset of military competition is not well defined. Prussia is coded as experiencing the onset of competition after 1500. Yet its core medieval territory, Brandenburg, experienced intense military competition before 1500. This pressure induced Brandenburg's rulers to grant local governance rights to nobles who were willing to extend loans to the rulers and join their military coalitions. This pattern of debt based coalition finance, driven by intense military competition, structured the institutions of officeholding in Brandenburg-Prussia in a way that persisted until the late 1600s. It was only after 1713 that the territory switched to law governed bureaucratic structures. This switch cannot be explained by positing that state structures were not shaped by military competition before 1500. They were shaped profoundly by military competition before 1500.

Fourth, French administrative structures were precociously well developed in the medieval period. These structures included the king's right to dismiss some royal officers at will – the defining feature of bureaucratic law according to this theory. Indeed the military fiefdoms of medieval France were governed by the rule of law to a greater extent than the military fiefdoms of medieval Brandenburg. Thus the question is *not*: Why did law based governance structures never get off the ground in medieval or early modern France? The question is: Why did the precociously law based governance structures of medieval France *decline* from 1500 to 1789?

### 2.3 Merchant community theory

According to this theory, legal restraints on the executive emerged from agreements between monarchs and merchant communities. The monarchs agreed to adhere to legal restraints in return for the right to collect tax revenue from the merchants. Such agreements were mutually beneficial because the mercantile tax base was large. The monarchs gained significant increases in tax revenue as long as the merchants received security from government takings through legal restraints on the executive. On this interpretation of the historical record, the main difference between European states in the course of their development was in whether they had a large mercantile community in the late medieval to early modern period. Those that did eventually developed legal restraints on the executive by the 20th century.

If the focus is on *electoral democratic* constraints on the executive, then this theory seems

to explain the cases of England, France, the Netherlands and Germany. All of them but Germany had large mercantile communities in the early modern period. All but Germany developed into electoral democracies by the 20th century. A closer look at the 18th-century evidence reveals anomalies, however. France had a large mercantile community but suffered a steady decay in the legal restraints on its executive that led to increasing tax and corruption rates by 1750 and revolution by 1789. The Dutch Republic also had a large mercantile community but suffered a decay in the legal restraints on its executive that led to increasing corruption rates and revolution by 1787. Prussia had a relatively small merchant community at the time, by contrast, but strengthened the legal restraints on its executive from 1713 to 1789 in a way that led to decreasing tax and corruption rates. All of this evidence contradicts the claim that the larger the merchant community is, the more likely it is that a state will develop legal restraints on the executive.

## 2.4 Social discipline theory

Both Prussia and the Dutch Republic developed greater bureaucratic efficiency with the help of Calvinist social discipline (as mentioned above). However, the Dutch Republic suffered increasing corruption rates in the 1700s while Prussia enjoyed decreasing corruption rates. This difference is not explained by the social discipline theory.

## 2.5 Contractual Realism explains both sets of cases: Those that supported the previous theories and those that were anomalies

Contractual Realism explains what the first two theories could not explain:

- Why intense military competition before 1500 in *both* France and Brandenburg (Prussia) led to the development of legal restraints on the executive in Prussia in the early modern period but not France. (*military competition theory*)
- Why Prussia's *lack* of medieval constitutionalism led to the development of legal restraints on the executive in the early modern period, but France's *legacy* of medieval constitutionalism led to the *decay* of legal restraints on the executive in the early modern period. (*medieval constitutionalism theory: causal factor 1*)
- Why intense military competitions in the 1600s induced the development of legal restraints on the executive in Prussia but not in France. (*medieval constitutionalism*)

*theory: causal factor 2)*

In the medieval period France relied on anchor based governance that created legal links between the monarchy and the nobility. The crown was seen as the supreme source of justice throughout the kingdom in an era when there was no distinction between the judicial, legislative and executive functions of royal authority. The king's provision of law courts to resolve difficult appeal cases and his issuance of royal edicts and laws to resolve difficult policy issues were part of a seamless web of monarchical governance. In response to the stresses of military competition in the medieval period, French kings used their legal links with the nobility to expand their governance rights in the areas of military leadership and taxation. The noble parliaments in Paris and the provinces still retained many governance rights though. The king was seen as having the exclusive right to make laws and issue edicts. The parliaments were seen as having the right to propose amendments before they "registered" his laws and edicts. Whether he had the obligation to accept the amendments was never made clear. This matrix of medieval constitutionalism set the stage for legal incompatibility problems between the crown and the parliaments in the early modern period. By 1650 the two institutions were at legal loggerheads. Such problems prevented mutually beneficial exchanges between them to establish legal restraints on the executive (see below).

Prussia had the opposite experience. In the medieval period, judicial systems throughout Germany lacked the capability to support anchor based governance (see chapter 30). Rights of all kinds were enforced by power based governance. The main source of tensions among the princes and nobilities was disputes over political exclusion-ism. Such disputes fueled military competition in and around Brandenburg (which was later to become Prussia). Due to the military and strategic incentives arising in such disputes (chapter 30), Brandenburg's rulers were forced to alienate governance rights to the provincial nobilities, thereby *destroying* legal links between the ruler and the nobilities and fueling further separatism and disintegration. This process left the provincial nobilities with many governance rights vis-à-vis the ruler, but left the ruler with almost no governance rights vis-à-vis the provincial nobilities. The provincial nobilities became so independent of the ruler that by the 1600s the ruler presided over an institutional *tabula rasa* at the center. The disadvantage of this legal vacuum was that it left the territory of Brandenburg vulnerable to devastation during the Thirty Years

War. The advantage was that when the central ruler began to remedy the problem of decentralization by initiating the construction of central institutions in the 1650s, there were no legal incompatibility problems to obstruct that process. Mutually beneficial agreements to redistribute governance rights between the provinces and the central government were feasible, including agreements to establish legal restraints on the executive.

This account explains why the intense military competition of the 1600s had such different effects on France and Prussia. In both states the military competitions of the 1600s intensified the pressure for domestic reforms. In France, however, it also intensified the underlying legal incompatibility problems, driving the crown and the parliaments apart and preventing mutually beneficial exchanges on governance rights. Under the pressures of military competition in the 1600s, the medieval inheritance of constitutionalism devolved into a *lack* of legal restraints on the executive by the 1700s. This development resulted in relatively high corruption rates and tax rates on non-elites compared to Prussia in the 1700s.

In Prussia, the military competitions of the 1600s did not collide with any fundamental legal incompatibility problems. When the central ruler began to remedy the problem of decentralization in the 1650s, he had to compensate the provincial nobilities for giving up some of their rights, but he had few long-standing rights for which he needed to demand compensation from them. Hence he could focus his demands on the imperatives of bureaucratic rationalization – only demanding concessions from the provincial nobilities on the dimensions of governance that he prioritized for central government effectiveness, and in return, granting all of their demands on the dimensions they prioritized. His approach to reform set the standard for the further reforms of his successors in the 1700s that established bureaucratic-legal restraints on the Prussian executive. Thus, under the pressures of military competition in the 1600s, the medieval legacy of localism allowed for an *increase* in legal restraints on the executive by the 1700s. This development resulted in relatively low corruption rates and tax rates on non-elites compared to France in the 1700s.

Contractual Realism explains what the *merchant community theory* could not explain:

- Why early modern France and the Dutch Republic failed to develop legal restraints on the executive and suffered economic misgovernance despite having large mercantile tax bases; and



- Why early modern Prussia developed legal restraints on the executive and enjoyed relatively better governance despite having a small mercantile tax base.

In France and the Dutch Republic, the ruling executives and taxable elites had an opportunity to reap mutual gains from exchange by trading legal restraints on the executive for higher taxes on the elites. They could not make the necessary mutual concessions because of legal incompatibility problems, however. Hence each state suffered relatively high corruption and tax rates on non-elites and inefficiently low taxes rates on elites. In Prussia, by contrast, legal incompatibility problems were avoided through anchor based governance, so rulers and elites could make the necessary mutual concessions. Hence Prussia had comparatively lower corruption and tax rates on non-elites and higher tax rates on elites. These points are worth explaining in more detail.

In France the legal incompatibility problem was that the monarchy claimed the exclusive right to make the laws but the parliamentary elites claimed the right to participate in lawmaking. When the monarchy demanded to solve fiscal crises by selling more offices and raising new tax revenue from the elites, they feared that if they made any concessions on the issue it would set a precedent that would leave them vulnerable to demands for more revenue and office sales *ad infinitum*. They accused the crown of "monarchical despotism" and demanded that the monarchy commit to limit the sale of offices before they could agree to pay higher taxes. The monarchy, in turn, feared that if it made concessions in response to such demands, that would set a precedent leaving it vulnerable to demands to limit its authority *ad infinitum*. Crown officials accused the parliaments of "usurping" the king's authority to make the laws. Given this deadlock between the two sides, it was impossible to reach an agreement by which the monarchy accepted legal restraints on its authority in return for new tax revenues.

A similar deadlock arose in the Dutch case due to a similar legal incompatibility problem. The regent cartel held that its members had the exclusive right to occupy their offices and exercise the prerogatives of their offices. The patricians excluded from regental offices held that they had the right to participate in determining policies followed by those offices. When the regent cartel demanded to solve fiscal crises – caused in part by their own corruption – by raising new tax revenue from the excluded patricians, they feared any concessions would

set a precedent leaving them vulnerable to demands for more revenue and prerogatives for the cartel *ad infinitum*. They accused the cartel of corruption and demanded that the cartel limit corruption before they could agree to pay higher taxes. The cartel, in turn, feared any concessions would set a precedent leaving it vulnerable to further demands to limit its prerogatives of office *ad infinitum*. Given this deadlock no deal to exchange higher taxes for legal restraints on the executive was possible.

This account explains why rulers in France and the Dutch Republic refused to commit to legal restraints on their executive authority despite the possibility of tapping a large mercantile tax base. In Prussia, by contrast, the bureaucracy relied on anchor based governance methods that avoided legal incompatibility problems. Hence rulers could commit to adhere to bureaucratic-legal restraints on their authority without fear of setting runaway precedents. And nobilities could commit to pay relatively higher taxes than their counterparts in France and the Dutch Republic without fear of setting runaway precedents.

## Part II

# Explaining history

## CHAPTER 7

### Rationality, equilibrium and optimality in long-run history: Two views

#### 1. What constitutes a good explanation of long-run history?

There is a saying among social scientists about how to tell a good explanation from a bad one. If a proposed explanation of some behavior gives you the sense that you would have done the same thing in the same situation, then it is a good explanation. This standard of explanation raises a question for the social scientist who is trying to explain long-run history. Would a 21st century observer *ever* have done the same thing as an historical actor from the 12th or 16th century? There are several reasons to think not. Historical actors had a different type of rationality than the average 21st century observer. Historical actors had a different understanding of the demands of political and economic competition and the requirements for a stable competitive equilibrium. And historical actors had a different way of assessing social optimality and deciding which sub-optimality problems needed to be solved. If the 21st century observer is going to imagine himself in the shoes of the historical actor, the observer needs to know that actor's type of rationality, understanding of competitive equilibrium, and way of assessing social optimality. The task of the social scientist who is explaining long-run history, therefore, is to describe these elements of the historical actor's situation. This is no easy task because there are deep debates among social scientists over the meaning of the concepts of rationality, equilibrium and optimality.

The purpose of this chapter is to untangle the threads of debate over these concepts and draw out the implications for this study. In each section, I will compare two competing views of each concept, explain how methodologists have shown one viewed to be flawed, and describe how the other view is used in this study. In the final section, I will explain

how each debate is an instance of the broader debate in the social sciences between the methodology of *reductionism* and the methodology of *holism*. The aim of the chapter as a whole is to demonstrate the methodological moves that are necessary to gain the benefits of both methodologies while avoiding the costs of both, and on that basis, delineate the type of empirical evidence needed to test a theory of conflict and settlement that explains long-run history. This chapter lays the foundation for the next chapter which addresses the question of how to generalize about history without making the mistake of assuming history is driven by universal laws.

## 2. Two views of rationality

### 2.1 Optimization *versus* rule-following

#### 2.1.1 Optimization

In the field of economics the standard definition of rationality is "the optimal adjustment of means to ends."<sup>1</sup> The rational actor responds to changes in his environment by adjusting his behavior to increase gains and reduce losses. When the marginal benefit from a certain behavior increases, the individual does more of that behavior. When the marginal cost of a behavior increases, the individual does less of it. The simplest way to model this kind of behavior is to assume that the individual engages in *mathematical optimization*. He maximizes his utility subject to a budget (cost) constraint, or he minimizes his cost of producing a good subject to a production technology constraint. As one economic methodologist described it:

"The criterion for economic rationality is thus the logical consistency of the agent's actions with his or her (explicit) knowledge and preferences. And since, even under conditions of subjective uncertainty, that knowledge and those preferences logically imply a best course of action, the agent is rational only when he or she selects that particular best course."<sup>2</sup>

The main value of this kind of model is that it can explain why individuals change their

---

<sup>1</sup>Langlois 1986a, p. 227.

<sup>2</sup>Langlois 1986a, p. 226.

behavior when their material environment changes. The standard explanation goes like this. If the individual did more of behavior *A*, it must have been because the material benefit from *A* went up or the material cost of *A* went down. The main application of this approach is to explain cases where an individual's environment changed and the individual changed his behavior in response. This approach runs into problems, however, in trying to explain cases where an individual's environment changed but his behavior remained the same – or cases where his environment remained the same but his behavior changed.

### 2.1.2 Rule-following

The main alternative to optimization modeling is to assume that individuals engage in *rule-following behavior*:

”[T]he agent is [assumed to be] a programmed robot that follows a simple rule of thumb. Thus it is no longer the logic of the agent's situation [in terms of the changing relative costs of taking different actions], but rather certain rules, asserted *ab ovo*, that determine the agent's behavior.”<sup>3</sup>

This approach explains behavior by tracing the historical sources of the rules that individuals follow and then simply assuming that they follow the same rules today that they followed yesterday (or last week, month, year or decade – whatever the relevant historical range is). This approach can explain cases where an individual's behavior remained the same even after his material environment changed. But it cannot explain cases where his behavior changed after his material environment changed (except by assuming that his behavioral rule had an exception for that circumstance built into it). The standard explanation of such cases is provided by the optimization approach (described above).

### 2.1.3 Relevance of the distinction for theories of state strategy

The distinction between the optimization approach and the rule-following approach is relevant for theories of state strategy. Many of the existing theories assume that the statesmen who make strategy engage in some kind of rule-following behavior. For example, defensive

---

<sup>3</sup>Langlois90b.

realism assumes that states seek security at all times. Offensive realism assumes that states seek regional hegemony at all times. Structural realism assumes that states follow a rule of responding to changes in the international balance of power at all times. Neoclassical realism assumes that states follow a rule of responding to changes in domestic political systems at all times.<sup>4</sup>

According to these theories, these four metrics – security, hegemony, the balance of power and domestic politics – are the core metrics by which to measure state strategy and the forces that drive it. However, as we will see in Parts II through V of this study, these theories fail to explain why and how states *changed* their strategies on these metrics after their material environments *changed*. That is, these theories fail the test of "rational strategy" according to the optimization approach. The theory offered in this study, by contrast, explains a wide range of changes in state strategies and international outcomes as rational responses to observable changes in states' material environment. This suggests that the optimization approach offers a more powerful explanation of state strategies and international outcomes than the rule-following approach. However, the optimization approach also has liabilities that need to be identified, understood and avoided before its explanatory power can be realized.

## **2.2 Optimization models of state strategy: What is the statesman's "means-end" framework, and where does it come from?**

In its simplest form the optimization approach assumes that the individual has a choice between two actions –  $A$  and  $B$  – and he chooses the mix of the two that maximizes his objective function,  $F$ , given the relative cost of the two actions. For example, if the cost of taking action  $A$  goes up, then he chooses less of action  $A$  and more of action  $B$ . If the cost of choosing action  $A$  gets high enough, then he takes none of action  $A$  and only action  $B$ . In this type of model  $F$  is the "end" that the actor pursues and  $A$  and  $B$  are the "means" that he employs to pursue this "end". Thus the model is a "means-end framework" or MEF.

This approach raises some basic methodological questions. How does the academic modeler know which possible "ends" the historical actor considered before deciding to pursue

---

<sup>4</sup>See the literature review chapter of Part III for citations.

"end"  $F$ ? And how does the academic modeler know which "means" the historical actor considered before deciding to employ "means"  $A$  and  $B$ ? One possible answer is for the modeler to assume that the historical actor had *complete knowledge of his entire world* – that he knew as much about it as all subsequent historians would ever learn about it later. If this answer is taken to its logical conclusions, it is tantamount to assuming that the historical actor:

- knew all possible ends that he could value;
- knew all possible actions he might take to pursue them;
- knew the marginal benefit he would receive from taking each one;
- knew the marginal cost he would incur by taking each one; and
- chose the ends that he would value the most and the optimal mix of all of the possible "means" to those "ends".

While this assumption might seem outlandish, it is the approach taken, at least implicitly, by leading theories of state strategy and international outcomes. Consider some examples.

- Niou and Ordeshook (1986) offer a game theoretic model of the balance of power in which each statesman has complete knowledge of the relative power of all states in the international system and the ability to calculate a relative power quotient for every possible coalition of states and every possible configuration of military forces that might emerge in a war.
- Morrow (2007) offers a game theoretic model of the laws of war in which each statesman has a million possible military strategies to choose from before he decides whether to commit to a treaty that would outlaw some of those strategies.
- Fearon (1995) offers a game theoretic model in which the main condition for bilateral peace is that each statesman has complete knowledge of the other state's power (in all its military dimensions) and the ability to calculate his state's optimal strategy for every possible combination of relative military force levels between the two states.



This knowledge and ability enable the opposing statesmen to maintain peace purely through deterrent threats. The theory claims that by implication, the main cause of war is uncertainty about relative power, because it prevents states from maintaining peace in this way.

- Powell (2006) offers a game theoretic model in which the main condition for bilateral peace is that power not shift by too large a margin too quickly. Each statesman has complete knowledge of the other state's power (in all its military dimensions) and the ability to calculate his state's optimal strategy for every possible combination of relative military force levels between the two states throughout a power shift. This knowledge and ability enable the opposing statesmen to maintain peace purely through deterrent threats – as long as power does not shift too much too fast.

Given the centrality of this approach in the leading literature, it should be examined to identify the problems with it and the potential solutions. For analytical purposes let us call the assumption that statesmen have complete knowledge of their world and the capacity to calculate the optimal mix of all possible means to their ends the "global MEF" assumption.

### **2.2.1 The pure optimization approach conflates rationality with skillfulness**

What is wrong with assuming that each actor has complete knowledge of his world and the capacity to optimize his strategy fully – a global MEF? For starters, no individual has such complete knowledge. And even if he did, it would be beyond his cognitive capacity to optimize across so many options simultaneously. More fundamentally, such an actor is not even behaving "rationally" in any realistic sense of the term. For if an actor had complete knowledge of his whole world and optimized across all of his options – a global MEF – his mathematical optimization problem would involve a foot-long objective function. Technically, this would still be explaining his behavior as an optimizing response to changes in his material environment (in so far as changes in the relative costs of his "means" would lead to changes in his behavior to achieve his "ends"). Yet intuitively, this is not really modeling his behavior as a product of optimization at all. Rather it is modeling his behavior as a product of fixed rule-following. For his foot-long objective function is a *single rule that*

*does not change* even when the relative costs of the different means to his ends change.<sup>5</sup>

From this perspective, the problem with the pure optimization approach is that it defines rationality in terms of skillfulness at solving mathematical optimization problems. So if an agent is *boundedly* skillful at this task, then he must be *boundedly* rational – by definition.

”This is an extremely strong conception of rationality, one that arguably confuses two issues: (1) the agent’s inclination to act in his or her best interest and (2) the level of the agent’s knowledge of and ability to achieve that interest.

”The alternative [to this strong conception] would have been to see rationality as a more humble trait, one that, because of its very limitations, could still operate more or less unboundedly in the face of limited cognitive capacity and limited knowledge. Under this alternative formulation, the same agent may be (subjectively) rational even though mistaken, confused or poorly informed about what, to an omniscient observer, would be in his or her best interests. The agent’s limitations may take the form of an inability to solve a complex optimization problem (even if few [real world] problems ever actually present themselves in this form), but such a limitation makes the agent not boundedly *rational* but boundedly *skillful*. Rationality, in this alternative formulation, is a matter of doing the best one can with what one is given, which includes one’s knowledge

---

<sup>5</sup>The economic methodologist put it this way:

”[I]f we are looking at models that represent the agent as solving an optimal-control problem with a foot-long objective function, are we really seeing an appeal to the logic of the situation [in terms of the relative costs of the different actions the agent might take]? We might just as well view this kind of maximizing behavior as a form of behavioralism [rule-following] – with a perhaps implausible psychology behind it. (Langlois 1986a, pp. 232–3)

”Situational analysis [of the relative costs of different actions the agent might take] insists that the agent acts not *optimally* but merely *reasonably* under the circumstances. Indeed, to act optimally – from the point of view of an omniscient observer – is often to act quite *unreasonably*. Behaviorists and other critics have pointed out relentlessly the narrow-mindedness or even straightforward irrationality of an agent actually trying to solve a foot-long Lagrangian as a guide to action.... [Thus] it is not at all clear that we should classify such optimization models as situational analysis [of the relative costs of different actions the agent might take]. Solving a complex optimization problem is an easy task for a computer: it is a matter of following an algorithm, programmed in much the same way one could program satisficing behavior or rule-following. Why not call these behavioralist models – that is, programmed-agent models? Solving a huge optimization problem may be a lousy decision procedure; but it *is* a decision procedure.” (Langlois90b)

and information-processing abilities.<sup>6</sup>

”Th[is alternative] criterion of rationality is the ability to act reasonably, to act appropriately to one’s circumstances, to adapt. [Think of] behavior that meets this alternative criterion of rationality as ‘adaptive’ behavior.”<sup>7</sup>

In other words, it is problematic to define rationality as ”optimization with a global MEF”. Real world actors do not have the knowledge and information-processing abilities to find the optimal solution to anything approaching a global MEF. And even if they did, that would make them highly skilled automatons following a fixed behavioral rule for problem-solving, not adaptive – and adaptable – agents responding to unexpected changes in their environment as best they can with what they’ve got.

If optimization with a global MEF is a problematic definition of rationality, then what are the alternative definitions? The two main alternatives are ”satisficing” rationality and ”adaptive” rationality.

### 2.2.2 ”Satisficing” does not solve the methodological problem

To engage in ”satisficing” behavior, an actor adopts a standard of imperfection in his effort to maximize utility – a standard for what constitutes ”good enough” – and he stops there. One type of standard is the *outcome thermostat*. The actor chooses a standard, say 90%, and as long as his utility does not fall below 90% of its hypothetical maximum, that is good enough. Another type of standard is a *ceiling for input costs*. The actor chooses a ceiling, say 10% of the value of the end he seeks, and he continues to search for the utility maximizing mix of means to that end until his search costs reach 10% of the end’s value.

The problem with satisficing models, however, is that they beg the question of where the standard of imperfection comes from. If the historical actor chooses the standard himself, then how does he make this choice? Through optimization? Then the academic modeler is back where he started – having to assume that the historical actor has complete knowledge of his whole world and makes the globally optimal choice of which standards of imperfection

---

<sup>6</sup>Langlois90b.

<sup>7</sup>Langlois 1986a, p. 230.

to use in his satisficing decisions! In other words, the assumption of satisficing behavior generates a problem of infinite regress. The historical actor might satisfice to certain standards, but the standards must come from somewhere. And if the actor chooses the standards himself, then there must be some more general "means-end" framework that he employs to choose the standards. But how does he make *these* choices? Through optimization or satisficing?<sup>8</sup>

The conclusion must be that the historical actor's MEF *cannot* be the global MEF. It must be some *subset* of the full set of means/ends trade-offs that he faces in the world around him. But what subset? And what is the source of this subset? By whom or what mechanism is it selected or created? The answers fall into two categories: Modeler-sourced MEFs and actor-sourced MEFs. Three main problems arise with modeler-sourced MEFs:

1. the fallacy of rational over-determination; (see Chapter 4.2.3)
2. the fallacy of assuming that historical actors employed *one* efficiency standard in deciding which of the *inefficiencies* in their world to redress, when actually they employed another efficiency standard in making those decisions; (see Chapter 5.1)
3. the determinacy paradox and the inside/outside fallacy. (Chapter 5.2)

Therefore this study focuses on actor-sourced MEFs. The basic idea is that actors adapt to changes in their material environment using whatever MEFs they have at hand. Then the main question is: Where do they get these MEFs?

---

<sup>8</sup>The economic methodologist writes:

"[I]f all that is at stake [in the debate over how to define rationality] is some constraint on information-processing and computational capacity, then one's satisficing alternative quickly collapses into substantive rationality – satisficing is actually the optimal course of action in view of costly computational resources (Baumol and Quandt 1964).... (Langlois 1986a, p. 226)

"If rationality consists only in the optimal adjustment of means to ends, then rationality must presuppose some framework of means and ends within which the optimization is to take place. Where do these frameworks come from? As a logical matter, they cannot themselves be explained as the result of maximizing choice. For if the choice of frameworks were the result of maximizing within some higher framework, the choice of that higher framework would remain unexplained – and so on ad infinitum (Winter 1964: 262-4; Kirzner 1982: 143-5)." (Langlois 1986a, p. 227)

### 2.2.3 The main sources of the "means-end" frameworks that actors use

The main sources of MEFs fall into four categories. The first and second categories are internal to the actor while the third and fourth categories are external to the actor:

1. Hard-wired psychology (evolution or instinct).
2. Soft-wired knowledge (learning).
3. The system constraint (competitive pressure). This is the constraint on each actor's choice of strategies and tactics (ends and means) that is created by competition between actors. When each actor chooses in a rationally adaptive way, there is a joint equilibration process that constrains his choice to be the best response to the other actors' choices (given his knowledge and information-processing abilities). The system constraint is, in effect, a selection mechanism that weeds out maladaptive strategies and tactics.
4. Institutions (contract terms). These are agreements among actors to limit the means and ends (tactics and strategies) that they employ to survive and compete. Institutions include:
  - socially defined values, priorities and taboos;
  - politically determined limits on policies, strategies and the uses of state resources;
  - economically motivated agreements to reap mutual gains from exchange and avoid holdups and cut-throat competition by fixing prices, quantities and other contract terms to be invariant to changes in relative pricing power and bargaining power.<sup>9</sup>

The purpose of defining institutions as contract terms is to emphasize that:

1. institutions define rights and obligations;
2. actors incur the cost of complying with institutional obligations because they receive the benefit of exercising institutional rights in return;
3. disputes over institutions are disputes over contractual rights and obligations;
4. actors calculate whether to settle or stand firm in a dispute by comparing the costs and benefits of settling (given the existing contract terms) to the costs and benefits of

---

<sup>9</sup>See economic literature on fixed price contracts cited in Section 6.1 of this chapter and in the second chapter of Part II.

standing firm (given the conflictive tactics and strategies that would become available).

In this study I focus primarily on the external sources: the system constraint and institutions. I consider the internal source of soft-wired knowledge – learning – only in so far as it relates to the evolution of institutions over time through the process of conflicts and settlements. I abstract away from hard-wired psychology completely on the grounds that it is not a variable over the time periods I consider, so it cannot explain the variation in actors' behavior from stability to conflict to settlement. These variations are explained well by changes in the other three sources of actors' MEFs:

- The meaning of institutional constraints changes when novel contingencies fall between the cracks of existing institutional rules (contract terms).
- When such contingencies cause conflicts and competitions in raising relative power, it induces actors to engage in learning about how to handle the contingency through some combination of conflict and settlement.
- The system constraint changes when the relative costs and benefits of alternative strategies of conflict and settlement change.

In what follows I focus on the relation between the system constraint and institutions. First I discuss how the system constraint can be interpreted most productively. Then I address the main question: Whether to explain state strategies and international outcomes as a product of the system constraint alone, of institutions alone, or of some combination of the system constraint and institutions? And if a combination, then what combination?

### **2.3 The system constraint: A definition and two alternative interpretations**

The system constraint is the constraint on individual choice that emerges from the competitive interaction of two or more individuals. The standard way of characterizing the system constraint is through "single exit modeling". A good example is the standard economic model of a perfectly competitive market. The model is constructed mathematically to insure that each agent has a single exit from his decision problem. Each firm has only one choice for the quantity of output to produce and the price at which to sell that output. Each

consumer has only one choice for the price at which to buy that output. This type of model has a number of methodological advantages:

1. It allows the agents in the model – firms and consumers – to exercise freedom of choice, and yet the model still produces determinate explanations of their behavior.
2. It explains market prices and quantities in a determinate way. That is, it produces clear predictions about the direction of changes in prices and quantities in response to changes in market demand and firm production costs.
3. It explains such changes as a function of observable variables (market demand and firm costs).
4. It does not require the modeler to know very much about the internal motivations and knowledge base of the agents (firms and consumers).
5. It therefore enables theorists to write parsimonious models with the potential for broad generalization (such as in the economic literature on general equilibrium theory).<sup>10</sup>

There are two different ways to interpret the system constraint. According to the first interpretation the system constraint is an *environmental filtering mechanism*. This interpretation takes the emphasis *off* of the agent's psychological process of making decisions.

"[W]e can appeal...to the larger environment that the agent's actions help form.... the system constraint.... [A] knowledge of the agent's environment can act as a substitute for knowledge of the details of his psychology.

"This is to a large extent an anti-psychological method. It doesn't require that we delve too deeply into the motivations of the agent. The constraints imposed by the agent's situation reduce his or her options sufficiently that a light postulate

---

<sup>10</sup>As the economic methodologist put it:

"The practice of restricting the agent's means-end framework in order to produce a determinate outcome is what Spiro Latsis refers to as *single exit modeling*, an approach he finds at the base of the neoclassical [economics] research program (Latsis 1972, 1976a,b). As first introduced to economics by Ludwig von Mises, single exit modeling was an attempt to reconcile a desire for determinate models with a belief in the free will of the economic agent. In a single exit model, the agent's behavior is not formally preprogrammed. Yet the model has determinate results, because we place the agent in a situation with only one reasonable exit. The agent is free to do as he or she likes; but, by analyzing the logic of the situation [in terms of the relative costs of the agent's options], we can determine the unique course of action a reasonable person would take." (Langlois 1986a, pp. 230–1)

of reasonable conduct [adaptive rationality] is adequate to secure a determinate outcome.”<sup>11</sup>

In the second interpretation the system constraint is a *mathematical optimization problem* that has a single exit. This interpretation puts the emphasis *on* the agent’s psychological process of making decisions.

”The *actual* decision processes of the agent are replaced by an optimization problem defined solely by the agent’s goals and the situation’s constraints.... If we are considering textbook Marshallian economics [e.g. models of firm and consumer behavior in perfectly competitive markets], it may be easy to persuade ourselves that not much psychology is at work, that we are studying the reasonable response of agents to a[n environmental] problem at the margin. But if we are looking at models that represent the agent as solving an optimal-control problem with a foot-long objective function, are we really seeing an appeal to the logic of the situation [in terms of the relative costs of the agent’s options]? We might just as well view this kind of maximizing behavior as a form of behavioralism [fixed rule-following] – with a perhaps implausible psychology behind it.”<sup>12</sup>

In other words, the more complicated is the situation being modeled, and the more complicated is the model, the less likely it is that the model accurately captures the *actual* psychological process by which the agents make their decisions. And therefore the greater is the risk of committing the fallacy of rational over-determination.

”The [explanatory] power of this [optimization] criterion is that it supplies a unique outcome as rational: There is only one rational exit. But this apparent uniqueness is in fact illusory. It is bought at the expense of an arbitrary specification of the means-ends framework in which the optimization is to take place.”<sup>13</sup>

---

<sup>11</sup>Langlois 1986a, pp. 237, 230–1.

<sup>12</sup>Langlois (1986a, pp. 232–3). Emphasis added. See pp. 230–5 for debates among economic methodologists as to whether the “representative agent” in economic models has “too much” psychology or “too little” psychology.

<sup>13</sup>Langlois 1986a, p. 230.



In other words, even if the particular optimization problem specified in the model *would* give rise to such behavior by actors in the real world, this does not necessarily prove that actors who behaved in that way in the real world were driven by that particular rationale. They may have been driven by some other optimization problem or filtering mechanism that had the *same single exit*.

In sum, it is beyond doubt that the forces of competition and strategic interaction constrain individual behavior. The standard way of characterizing this constraint is the mathematical technique of "single exit modeling". Yet such a model always has two different interpretations – one emphasizing the features of the actor's environment that constrain him, the other emphasizing the psychological process by which he responds to those features of the environment. The first interpretation leaves room for a realistic "mental model" of the actor's psychology. The second interpretation runs the risk of permitting an unrealistic mental model of it, and therefore the risk of characterizing a causal mechanism *different* from the one that actually drove the actor's behavior.<sup>14</sup>

In conclusion it can be useful to characterize the system constraint using a single exit model *provided*:

- such models are interpreted as environmental filtering mechanisms with the emphasis on observable cost and benefit features of the actor's environment, rather than as optimization problems with the emphasis on unobservable, arbitrarily specified psychological processes; and
- the modeler allows for the role of institutional constraints in shaping behavior as well.

First consider how the system constraint functions as an environmental filtering mechanism. Then consider how the system constraint is related to institutional constraints.

### **2.3.1 How the system constraint selects adaptive behaviors and weeds out maladaptive behaviors**

There are two ways to interpret the process by which the system constraint affects individual thinking and behavior. The first interpretation might be called "mathematical marginalism".

---

<sup>14</sup>See North and Munger (1994) for an analysis of what they call "mental models".

The second interpretation might be called "myopic marginalism".

The first interpretation is the "Marshallian" conception (named after Alfred Marshall who, with Stanley Jevons, invented the mathematical model of perfect competition among firms). In this interpretation, the individual firm's cost and benefit schedules are simply scaled down versions of the market supply and demand schedules. In this sense each firm is assumed to have complete knowledge of the entire market's supply and demand conditions. The firm then engages in *mathematical marginalism*. It maximizes profit by producing output up to the point that marginal benefit equals marginal cost. That is to say, "The *actual* decision processes of the agent are *replaced by an optimization problem* defined solely by the agent's goals and the situation's constraints."<sup>15</sup>

The assumption that firms engage in mathematical marginalism may be realistic for perfectly competitive markets for simple goods (like agricultural commodities). The assumption is less and less realistic as the market in question becomes more complicated, however. Taken to its logical conclusions, mathematical marginalism is totally unrealistic and the other interpretation is needed.

"In general-equilibrium theory, which seeks to reduce economic phenomena to psychological states (utility functions) and natural givens (endowments and technological possibilities), the cognitive demands on the [representative] agent are phenomenal. If instead we permit the agent to respond not to the entire economic universe (or an unreasonably large part of it), but to a manageable subset, then the [cognitive] demands [on the agent] are much attenuated, and what we need to know about his internal landscape [of knowledge about his world] is reduced."<sup>16</sup>

In other words, there are two methodological reasons why we should *not* assume that the individual actor's means-end framework is a scaled-down version of the aggregate behavior of all actors in the system. First, we can avoid the unrealistic assumption that the individual agent has complete knowledge of his world and superhuman cognitive abilities. Second, we do not have to figure out all that the agent knows about his world to explain his behavior.

---

<sup>15</sup>Emphasis added. Langlois 1986a, p. 232.

<sup>16</sup>Langlois 1986a, pp. 236–7.

This naturally leads to the second interpretation of the system constraint. Here the actor is assumed to engage in *myopic marginalism* rather than mathematical marginalism.

”Acting ‘on the margin’, after all, often means nothing more than acting in a boundedly rational way – *acting appropriately to the situation one faces on the margin rather than reacting to the total picture.*

”[V]ery often the interesting explananda involve large-numbers [of actors in] situations with more or less tight [system] constraints. And in those situations we can use a more simplified ideal type [of individual agent in our models]. But this does not automatically mean we are limited to Marshallian theories in which we draw conclusions about aggregate outcomes simply by scaling up the behavior of any single representative individual ([or] scaling down aggregate behavior...to get a representative firm....)

”[T]he message of Alchian’s [1950] article...is not so much that we can or should eliminate marginalism [itself] in the sense of eliminating any particular motivational assumption.... [Rather] the aspect of marginalism that Alchian calls most seriously into question is its compositional structure – its assumption that aggregate outcomes are just individual outcomes writ large [as in the Marshallian view of the system constraint as a force driving agents toward mathematical marginalism].”<sup>17</sup>

In other words, the system can still impose a constraint on the individual’s behavior even if no individual knows the “total picture” of the system constraining him. In the case of competition among firms, each firm can still be constrained to choose a certain output level and selling price even if it does not have complete knowledge of the entire market’s supply and demand conditions. In the case of military security competition among states, each state can still be constrained in its strategies even if it does not have complete knowledge of the international system’s conditions of relative power, economic productivity, political legitimacy, etc.

---

<sup>17</sup>The sentence in parentheses is a footnote in the original. Italics added for emphasis. Internal quotation marks in original. Langlois (1986a, p. 241).

In sum, it is beyond doubt that the system constraint affects the individual's thinking and actions. However, it is unclear whether to assume the individual knows a lot or a little about the system. On one interpretation – mathematical marginalism – he knows everything that an omniscient observer would know and he has superhuman capacities to calculate the optimal means to his ends. On another interpretation – myopic marginalism – he knows much less and adapts on the margin without knowing the total picture. This interpretation enables the theorist to make more realistic assumptions about the actor's actual "mental model", and thereby increase the chances that the theory will characterize the actual causal mechanism that drove the actor's behavior.

"If...we permit the agents to respond not to the entire economic universe (or an unreasonably large part of it), but to a manageable subset, then the [cognitive] demands [on the agent] are much attenuated, and what we need to know about his internal [psychological] landscape is reduced....

"[A]s a subjective matter, agents do not typically conceive of their problem-situations as taking in the whole economic universe.... [Therefore the theorist can] take as exogenous the existence of various *social institutions*.... [I]nstitutions have an informational-support function. They are, in effect, interpersonal stores of coordinative knowledge; as such, they serve to restrict at once the dimensions of the agent's problem-situation and the extent of the cognitive demands placed upon the agent. This is...'institutional individualism'. The problem-situations that we take as exogenous are not fully arbitrary – they have...an objective correlative in various 'distinct social yet not psychological entities (called institutions, customs, traditions, societies, etc.)'." <sup>18</sup>

In other words, agents are constrained by the system even though they do not have a "total picture" of it or the capacity to respond optimally to a "big picture" of it. Institutions fill the gaps in agents' knowledge and cognitive capacities by specifying the means-end frameworks (tactics and strategies) they can use in adapting rationally to changes in their environment.

---

<sup>18</sup>Italics in original. Internal quotations are from [145]–Agassi". Langlois 1986a, p. 237.

### 2.3.2 The perfectly knowledgeable, optimizing statesman: Does his MEF have too much psychology in it or too little?

In most models of war, the underlying assumption is that international security competition imposes a "system constraint" on state strategies. It is critical to interrogate these models to determine whether they interpret the system constraint through the lense of maximizing marginalism or myopic marginalism.

- Do these models make unrealistic assumptions about statesmen's knowledge of the world and capacities to calculate the optimal means to their ends?
- If so, do these models run the risk of characterizing causal mechanisms different from the ones that actually drove statesmen's decisions whether to initiate war or remain at peace?

Ultimately, the eruption of a war is due to whatever *blocked or eliminated* the conditions for peace. Therefore any model of the causes of war must be based on a model of the conditions for peace.

- Do existing models of peace make unrealistic assumptions about statesmen's knowledge of the world and capacities to calculate the optimal means to their ends?
- If a model assumes that peace is a non-cooperative equilibrium generated solely by the system constraint of pure security competition, then how much would each statesman need to know about the world – and how great would his cognitive capacities need to be – to support that equilibrium?
- Does the historical evidence lend support to such an explanation of peace and stability?
- Is there an alternative explanation of peace and stability in which it is generated not only by the system constraint but also by institutional rules? And if so, does this explanation receive more or less empirical support than the first explanation?<sup>19</sup>

---

<sup>19</sup>See Langlois (1986a, pp. 230–5) for debates among economic methodologists as to whether the "representative agent" in economic models has "too much" psychology or "too little" psychology.

To answer these questions, let us explore an analogy that highlights the way "single exit" modeling is practiced and interpreted in formal theories of war and peace.

## 2.4 Mathematical marginalism *versus* myopic marginalism in theories of costly conflict: An analogy

Let us start with a simple "single exit" model. This model is inspired by one of the original examples that Latsis offered in his seminal work on single exit modeling. So it is worth quoting his setup and comments on the system constraint.

"Consider another type of explanandum: [Suppose] a spectator at a football game left the stadium by gate  $K$ . Let us slightly elaborate on this spectator's physical environment and suppose that [before he leaves] he is seated near gate  $K$  and that the stadium is packed with thousands of spectators.... [T]he physical and social situation in which the spectator finds himself seems to be compelling [him] towards a specific kind of overt activity. Though a host of logical possibilities exist, the environment seems *physically and situationally constraining*. Deviations from this particular course of action [of leaving through gate  $K$ ] raise questions either about the internal makeup of the agent in question or about unknown situational factors which happened to be relevant to the agent's situational picture. But as the situational picture (including his articulable preferences) becomes more complete one would expect deviations [from this particular course of action] to be almost exclusively attributable to the agent's internal environment."<sup>20</sup>

In other words, the tighter the system constraint is, the fewer options the agent has for how to behave, and the less impact his internal psychology has on his choice among them. If he does deviate from these options for some reason, it must have been for an internal psychological reason.

Following in the Latsis tradition of methodological analysis, let us now imagine that a soccer fan is sitting in an *empty* stadium and wondering what is the fastest way to get back to his car. He thinks, "I'll just head for the nearest exit." He gets up and walks to the

---

<sup>20</sup>Latsis 1980.

interior concourse of the stadium and looks to the right and the left for the nearest exit. Discovering that he is half way between two exits, he picks one of the exits randomly and takes it.

Now imagine that the same soccer fan is sitting in a *full* stadium and wondering what is the fastest way to get back to his car. He thinks, "I'll just head for the nearest exit." He gets up and walks to the interior concourse of the stadium and looks to the right and the left for the nearest exit. Discovering that he is halfway between two exits, he wonders, "Which crowd is smaller: the crowd to my right or the crowd to my left?" He notices that the crowd is pretty big in each direction. He can't literally count the size of each crowd person by person. So he just eyeballs it. Guessing that the crowd to his right is smaller, he takes the exit to his right.

Now imagine that the same soccer fan is sitting in a full stadium with his friends and they all belong to the local team's fan club. The opposing team's fan club is sitting in the next section over and starts to get upset at the last call the referee made. They start hurling insults and shaking their fists at our soccer fan and his fellow club members. After the game is over everyone heads for the cars. A member of one of the clubs breaks the window of a car owned by a member of the other club. A melee of window smashing breaks out and turns into a riot. What caused it? To answer this question, consider another hypothetical example that starts out similarly but ends without any violence.

The same two teams are playing in the same stadium a year later. The opposing fan clubs are sitting in the same two sections. The visiting team's fan club starts to get upset at the last call the referee made. They start hurling insults and shaking their fists at our soccer fan and his fellow club members. After the game is over each member of each club wonders what is the fastest way to get back to his car. He takes the nearest exit and gets back to his car to discover that no one has broken any windows. No violence breaks out.

How would a theorist of international relations explain this peaceful outcome? Well, it depends on what methodology he uses.

**Methodology 1.** The theorist assumes that the "means-end framework" of the soccer fans can only come from one source: the system constraint of pure security competition. And the system constraint is interpreted as a force driving agents toward *mathematical*

marginalism. The theorist writes a model in which each member of each club faces the following optimization problem. "How fast would I have to run back to my car to protect its windows from every single member of the opposing club?" In this model, each member of each club is assumed to have complete knowledge of where each member of the opposing club is seated, and each member calculates exactly how fast he would have to run back to his car to protect it from all of the opposing club members. In the non-cooperative equilibrium, each member of each club runs back to his car just fast enough to protect it from all of the opposing club members. No violence breaks out. And no fan runs any faster than he needs to run to prevent violence from breaking out. In short, using this methodology, the academic theorist would explain the *absence* of violence by looking at the distribution of relative (positional) power.

How would this theorist explain the *previous* case – the one that ended in violence – using this methodology? He would assume that something prevented each member of each club from running back to his car fast enough to protect it from all of the opposing club members. Perhaps one member's car was just too far from his seat and he couldn't run fast enough. Or perhaps one member's seat was so well positioned that he could get to an opposing fan's car before that fan got to it. Or perhaps one member knocked down an opposing member to prevent him from getting to his car fast enough to protect it. In short, the theorist would explain the *outbreak* of violence by looking at the distribution of relative (positional) power.

Now let's take a step back and evaluate this methodology for theory building. The actors' means-end framework comes only from the system constraint of pure security competition and this constraint is interpreted using mathematical marginalism. Using this methodology, the theorist can only explain the actors' decisions based on one factor – the distribution of relative positional power – because MEFs are assumed to come from only one source – the system constraint of pure security competition. As a result the theorist must make very strong assumptions about the knowledge and calculation abilities of the actors. The result is a theory of "rational" behavior, technically speaking. And it is a parsimonious theory. It explains both cases – the game that ended in violence and the game that ended without violence – based on the same factor: the distribution of relative positional power. However, it is not a realistic theory. And its apparent success in explaining the difference between the



two cases might lead the theorist to stop looking for evidence that the difference is better explained by some other factor. So it leaves the evaluator of this methodology with two questions. Do soccer fans in the real world actually behave this way? Doubtful. Is there an alternative methodology that might produce a more realistic model of fan behavior that leads to violence in some cases but no violence in other cases? Yes.

**Methodology 2.** The theorist allows the soccer fans' means-end framework to come from two sources: the system constraint and institutional constraints. And the system constraint is interpreted as an inducement to *myopic* marginalism (not mathematical marginalism). Before the theorist writes any model, he collects more evidence on the two cases – the first game that ended in violence and the second game that ended without violence. He discovers that in the first game the referee made a bad call and *both* fan clubs thought it was a bad call. (Say, one player committed a hard foul; the fouled player got up and spit in the first player's face; and the referee gave red cards to both players.) In the second game, by contrast, the referee made a difficult call and the visiting team's fan club got upset and started hurling insults and shaking their fists. But their club president spread the word to all of them that it may have been a difficult call, but it was the right call in the end, so no one should take the issue any further than hurling insults and shaking fists. His club members then followed his lead and remained non-violent.

With this empirical evidence in hand, the theorist writes a simple model to explain whether a soccer game ends in violence or without violence. The main condition for a soccer game to end without violence is that each of the opposing fan clubs agrees that the referee made the right calls. As long as this condition is met, the game ends without violence. Conversely, the main condition for a soccer game to end in violence is that *both* of the opposing fan clubs feel the referee made the wrong calls. When this condition is met, the game ends in violence.

Now let's take a step back and evaluate *this* methodology for theory building. The actors' means-end framework may come from two sources – the system constraint and institutional constraints – and the system constraint is interpreted using myopic marginalism. Using this methodology, the theorist can make more realistic assumptions about the knowledge and information-processing abilities of the actors. Yet he can still produce a parsimonious

theory. It is parsimonious in three respects. First, it explains their decisions based on a single factor: whether the referees' calls were consistent or inconsistent with the rules of soccer. Second, it attributes the two forms of soccer fan behavior – violence after the first game and no violence after the second game – to myopic marginalism. Third, it assumes that each actor has two MEFs to choose from, and it explains what causes him to choose one or the other.

**MEF 1**, for the situation where the referees' calls were consistent with the rules of soccer, is sourced from institutional constraints.

**MEF 2**, for the situation where the referees' calls were inconsistent with the rules of soccer, is sourced from the non-cooperative system constraint of pure security competition.

This point is worth explaining. In the game that ended without violence, the visiting team's fan club may have gotten disgruntled after the difficult call. But when the club president called them into line, the club members followed their noses (myopically) and obeyed the *institutional rule* he was invoking: As long as the referee makes the right call, you *ignore* the relative power of the opposing club members, you don't jockey for a relative power advantage, and you don't get violent. In the other game that ended in violence, the members of both fan clubs used this same institutional MEF in the same myopic way *up until* the referee made the bad call. Then they switched to a different MEF that emerged from the system constraint of pure security competition: paying attention to the relative power of the opposing club members, jockeying for a relative power advantage, and getting violent.

Is this a theory of "rational" behavior? Yes, as long as we mean adaptive rationality: "the ability to act reasonably, to act appropriately to one's circumstances, to adapt....doing the best one can with what one is given, which includes one's knowledge and information-processing abilities."<sup>21</sup> Is it a realistic theory? It seems more realistic than the first theory.<sup>22</sup>

---

<sup>21</sup>Langlois 1986a, p. 230; **Langlois90b**.

<sup>22</sup>Do soccer fans in the real world actually behave this way? Most soccer games end without any violence, even when the referee makes a bad call and both sides feel disgruntled. But if there ever arose a case where

Is it more parsimonious than the first theory? This depends on how one measures a theory's parsimony. If one measures it simply by the number of causal assumptions in the theory, then this theory is less parsimonious than the first theory. But there is another way to measure a theory's parsimony: the number of its observable implications (corroborated by evidence) divided by the number of causal assumptions it makes. On this measure this theory is more parsimonious than the first theory. It has one or two more assumptions but it has many more observable implications.

In the next few sections I outline the conclusions that emerge from this analogy.

#### **2.4.1 The statesman posited by standard models of costly conflict: Is he too knowledgeable and capable?**

Let us think for a moment about the *presidents* of the two soccer fan clubs. How much would each president need to know about his world – and how great would his cognitive capacities need to be – for him to know whether to instruct his fan club members to remain non-violent or break into violence? It depends whether he is using methodology 1 or methodology 2 (from above). Using methodology 1, he would need to know where each member of each club is sitting, where each member's car is located, and how fast each member can run. He would also need to know the size of the crowds in the stadium concourse that each member would need to run through to get to his car or an opposing fan's car. And he would need to be able to calculate – using all this information – whether each member could get to his car fast enough to protect it. Using methodology 2, by contrast, he would only need to know whether the referee's calls were consistent or inconsistent with the rules of soccer.

This thought experiment raises questions about game theory models of war based on relative power factors. Such theories assume that statesmen have complete knowledge of their world and a full capacity to figure such information into their decisions whether to initiate war or remain at peace.<sup>23</sup> If this assumption were true, it would be phenomenal. Each statesman would have complete knowledge of the relative power of all states in the system,

---

the referee made a bad call, both sides felt disgruntled, and it led to violence, it was probably for the reason posited in this theory rather than the reason assumed in the first theory.

<sup>23</sup>See Section 3.3 above for a summary of the leading theories that make this assumption.

and he would be able to calculate a relative power quotient for every possible coalition of states and combination of relative military force levels that might emerge in a war. (This is analogous to the fan club president who has all of the knowledge and capabilities listed in the previous paragraph under methodology 1.) Is this a realistic assumption about statesmen? It might be realistic if wars were tidy. But wars are messy and unpredictable. Actors at all levels are plagued by structural uncertainties. They cannot assign probabilities to different potential outcomes. They do not even know what the full set of possible outcomes is. If each statesman had to meet this standard of rationality, then (as in general equilibrium theory) 'the cognitive demands on the representative agent would be phenomenal'.<sup>24</sup>

There is always a relative power quotient between any two actors (regardless of whether anyone knows what it is). So a theory based on relative power quotients might seem to apply generally to *every* situation in which there are two or more actors. This is why theories of war based on relative power factors appear to be *general* theories. Yet such theories make exceptionally strong assumptions about statesmen's knowledge of relative power and ability to calculate relative power quotients and optimal strategy responses. Hence the generality of power-based theories is an illusion.

"The [explanatory] power of this [optimization] criterion is that it supplies a unique outcome as rational: There is only one rational exit. But this apparent uniqueness is in fact illusory. It is bought at the expense of an arbitrary specification of the means-ends framework in which the optimization is to take place."<sup>25</sup>

#### 2.4.2 Tidy conflicts *versus* messy conflicts: Implications for theories of war and peace

Real wars are not tidy enough for statesmen to have such complete knowledge and calculation abilities. In a real war every actor is maximizing on every margin he can think of – and

---

<sup>24</sup>"In general-equilibrium theory, which seeks to reduce economic phenomena to psychological states (utility functions) and natural givens (endowments and technological possibilities), the cognitive demands on the [representative] agent are phenomenal." Langlois (1986a, pp. 236–7)

<sup>25</sup>Langlois 1986a, p. 230.

constantly figuring out new margins to maximize on. There are few limits on the strategies and tactics that actors use (compared to the many limits in peacetime). In this sense the world of war is entirely different from the world of peace. In the world of war mathematical marginalism is impossible. Myopic marginalism is the best that any actor can do. This is not to say that actors never adapt rationally during war. They do – and there are many models of rational adaptation during wartime. But it is *myopically* rational adaptation: "Acting 'on the margin', after all, often means nothing more than acting in a boundedly rational way – *acting appropriately to the situation one faces on the margin rather than reacting to the total picture.*"<sup>26</sup>

### 2.4.3 Institutions solve the problem of messy conflicts and competitions in raising relative power

For actors engaging in myopic marginalism – following their noses and doing the best they can with what they are given – the world of war is a forbidding place. This is why adaptively rational actors avoid the world of war. They settle for peace by agreeing to *limits* on their rights to maximize on certain margins. In leaving the world of war, they enter a world of peace and stability in which their strategies, while constrained by relative power at the broadest level, are constrained at most other levels by contract terms instead (e.g. treaties, norms, laws, agreements, institutions, etc). Within these levels, actors' main solution to the problem of messy conflicts and competitions in raising relative power is contracts whose terms are independent of changes in relative power (e.g. anchor term contracts). With this solution actors can take their eyes off relative power and stop jockeying for power all the time. From this perspective, the main cause of the descent back into a world of war is not relative power factors. It is unresolvable disputes over the contract terms that bind actors in peaceful coexistence.

---

<sup>26</sup>Italics added for emphasis. Internal quotation marks in original. Langlois (1986a, p. 241)

#### **2.4.4 Statesmen's MEFs *evolve* in response to specific conflicts and settlements:**

##### **So the rationality assumption must have historical fidelity**

If the main cause of conflict is disputes over contract terms, then a theory of conflict must draw the connection between (1) specific disputes over particular contract terms and (2) the incentive to revert from the world of peace back to the world of war. If the theory is to aspire to any kind of generality, it must discern the common features among different disputes that ended in war – even though each dispute erupted over a different contract term governing inter-state relations. That is, the theory must show that many different disputes over a variety of contract terms caused war by a single causal mechanism that links disputes to the incentive for war.

This standard for generalization is more difficult to meet than the standard for theories based on relative power factors. For, as mentioned, there is always a relative power quotient between any two states. So a theory based on relative power may seem to apply automatically to *every* situation in which there are two or more states. By contrast, in a comparison of two different contract disputes between two different pairs of states, it may not be so evident what the two disputes have in common.

In this study, I assume that there *is* a general mechanism that links many different disputes over a variety of contract terms to the incentive for war. The mechanism is the problem of legal incompatibilities. A legal incompatibility is a situation where one contract term gives one actor the right to an asset, but another contract term gives another actor the right to the same asset. The connection between each particular dispute and the general incentive to revert to war is straightforward. If either of the two states in the dispute gives up its right to the asset in question, then it will give the other state an incentive to create more legal incompatibilities in the future for its own gain. So each state has an incentive to wage war on the other state until its war costs have become greater than its expected gains from winning a share of the disputed right in any peace settlement.

To prove that this incentive was the cause of each war, it is necessary to examine the facts of each dispute in detail and show that each statesman weighed the costs and benefits of two alternatives:

(A) letting the other state have the right in question, *versus*

(B) waging war on the other state to induce it to give up the right in question.

Notice that the statesman's choice between these two options is couched in terms of the "means-end" framework of rationality defined earlier (MEF). Thus each statesman has a choice between two actions: *A* and *B*. If the cost of choosing action *A* gets high enough, then he takes none of action *A* and only action *B*. That is, if the cost of letting the other state have the right in question gets high enough, then the statesman wages war instead.

Historically, each statesman's valuation of these options depended on the specific circumstances that rendered one contract term incompatible with another contract term, thereby generating a dispute. In this sense each statesman's MEF was specific to the dispute at hand. There was no single, general MEF in common among all disputes. This conclusion contrasts sharply with the single, general MEF in common among all outbreaks of war in the leading theories of costly conflict based on relative power.<sup>27</sup>

This contrast carries a major methodological implication. Statesmen's mental models of the means and ends available to them *evolve* in response to specific conflicts and settlements. Therefore the modeling assumptions in a theory of war due to contract disputes must have historical fidelity to the subjective rationality of the statesmen involved in each war (just like an explanation of the outbreak of violence after a soccer game must have historical fidelity to the facts about the referees' calls during the game – and whether those calls were consistent or inconsistent with the rules of soccer on the books that year). This point is worth examining more closely.

When a dispute over contract terms arises, each actor faces a choice whether (A) to adhere to the limits on his tactics and strategies – means and ends – entailed by existing contract terms, or (B) to ignore those limits and switch to some alternative (conflictive) means-end framework. If the later theorist wishes to explain why the actor chose one option or the other, the theorist needs to know (A) what the existing contract terms were and (B) what the alternative (conflictive) means-end framework would be.

---

<sup>27</sup>Fearon 1995; Powell 2006.

A similar point applies to the choice whether to settle an ongoing conflict over contract terms. When each actor decides whether to settle, he faces a choice between (1) continuing to engage in conflict using the conflictive MEF and (2) agreeing to a new set of contract terms that resolves the legal incompatibility which caused the conflict in the first place. The theorist who wishes to explain why actors chose one option or the other needs to know (1) what the conflictive MEF was and (2) what the new contract terms would be.

After each settlement, the new contract terms become the "background" MEF that will be in statesmen's minds whenever the *next* dispute over contract terms might arise. In short, the statesmen's MEFs evolve over time. Therefore a general theory of the causes of conflict and the conditions for stable settlement must be based on the specific MEF's in statesmen's minds at the time of each conflict and settlement. That is, the theory must have historical fidelity to the particular contract terms binding states in their relations to each other at each point in time.

This way of striking a balance between realism and parsimony is very different from the way taken by the standard game theoretic accounts of costly conflict. The contrast points to a deeper methodological question: How can theorists generalize about the causal mechanisms that govern human behavior?

### **3. Two views of competitive equilibrium**

In the 1970s and 1980s, the discipline of economics experienced a sea change in its understanding of market competition. Prior to that point, economists viewed all private restrictions on market trade as anticompetitive. If a group of firms fixed prices or set production quotas to hike prices above the free market level, it was seen as anticompetitive "collusion". If a firm required consumers to sign long-term contracts to limit their option to buy from other firms, it was seen as anticompetitive "exclusive dealing". If a firm required suppliers to sign long-term contracts to limit their option to sell to other firms, it was seen as anticompetitive "market foreclosure". In the 1970s and 1980s, however, economists discovered that there might be efficiency rationales for such horizontal and vertical restrictions on market exchange, at least in oligopolistic industries where firms make large sunk investments. In



such industries unrestricted competition would leave firms vulnerable to economic losses due to the problems of cut-throat competition and holdups (which I will explain below). To avoid such losses, firms would not invest in these industries in the first place. And their non-investment would constitute a social welfare loss for both firms and consumers. If such problems could be solved through horizontal and vertical restrictions, however, it would enable firms to invest in such industries, so these restrictions would be procompetitive and welfare enhancing.

This new understanding of market competition generates a number of methodological implications for scholars who are seeking to provide rationalist explanations of competition, conflict and cooperation. The next few sections draw out these implications.

### **3.1 Explaining cooperation among firms: The "perfect competition" standard**

The perfect competition model assumes that once a firm has made an investment in productive capacity in an industry, it can redeploy the investment to another industry costlessly if necessary. If the firm becomes so inefficient in the first industry that it must drop out of competition there, it can redeploy its capacity investment to another industry at no cost. The model also assumes that a firm's input suppliers and customers have the unrestricted right to bargain with it over price, search for other firms that offer a better price, and switch firms to obtain a better price. In effect, suppliers and consumers extract price concessions from the current firm by threatening to switch to another firm for a better price. This threat induces each firm to lower its production costs in order to compete for customers by offering them lower prices, and for suppliers by offering them higher prices. In other words, the threat of losing customers and suppliers provokes an unrestricted competition among firms that induces each firm to increase its productive power relative to other firms.

This unrestricted competition in raising relative power is efficiency and welfare enhancing. More efficient firms gain a larger market share while less efficient firms gain a smaller market share (and an incentive to become more efficient).<sup>28</sup> If a firm becomes so inefficient that it

---

<sup>28</sup>This assumes that each firm's production function is characterized by strictly increasing marginal cost. If each firm's production function has constant marginal cost, then the most efficient firm gains the entire market and the rest of the firms gain no share of the market and drop out of competition.

must drop out of competition in that industry, it can redeploy its productive capacity to another industry costlessly. So it would not lose any of the value of its capacity investment. Hence it would not be deterred from entering any industry in the first place for fear of such a loss. Thus the unrestricted right of consumers and suppliers to search for a better price, bargain over prices and switch firms to obtain a better price – and the unrestricted competition in raising relative power that it provokes – are *efficiency and welfare enhancing*. The spot market economy becomes more efficient and all agents are made better off – not only the consumers and the suppliers but the firms as well.

Using this perfect competition standard, how would an economist explain vertical and horizontal restrictions on the rights of suppliers and consumers to search, bargain and switch firms? He would assume that firms were engaging in anticompetitive practices. He would explain horizontal restrictions (such as price-fixing or production quotas) as attempts at collusion, and vertical restrictions (such as long-term, fixed price contracts) as attempts at market foreclosure or exclusive dealing. He would conclude that such practices were reducing efficiency by preventing unrestricted competition among firms to raise their relative productive power.

### **3.2 Explaining cooperation among firms: The "transaction cost economics" standard**

#### **3.2.1 The holdup problem**

The situation is different in industries characterized by *imperfect competition*, where each firm must make a sunk investment in production capacity to enter the industry. A sunk investment is one whose value cannot be recouped completely if the firm must leave that market or transaction partner. Once the firm has made such an investment, it is vulnerable to being "held up" by its input suppliers or customers. The supplier might demand that the firm raise its buying price and threaten to sell the inputs to a different firm if the demand is not met. If the demanded price is too high, then the firm is threatened with being forced out of the industry and losing the value of its sunk investment.<sup>29</sup> Similarly, the consumer

---

<sup>29</sup>By agreeing to that price, the firm's average total cost would rise above its average total revenue, thereby saddling the firm with net losses and requiring it to exit the industry.

might demand that the firm lower its selling price and threaten to buy the outputs from a different firm if the demand is not met. If the demanded price is too low, then again the firm is threatened with being forced out of the industry and losing the value of its sunk investment.<sup>30</sup>

To avoid losing their sunk investments in this way, firms will either not invest in such industries in the first place or require long-term, fixed-price contracts with their input suppliers and customers. If such vertical restrictions are prohibited (based on the misconception that they are anticompetitive and efficiency reducing), then firms will not invest in such industries in the first place. Their non-investment will constitute a social welfare loss for all agents: firms, input suppliers and consumers.

The holdup problem is caused by two factors: (1) firms have made sunk investments and (2) suppliers and consumers have an unrestricted right to bargain with the firm over price, search among firms for a better price, and switch to another firm to obtain a better price. In effect, when suppliers and consumers have this unrestricted right, it provokes a competition in raising their relative power to extract price concessions from the current firm by threatening to reduce the value of its sunk investment so low that it must go out of business. In other words, it provokes a competition in raising the relative power to commit holdups. In the limit, suppliers and consumers will maximize their power to commit holdups on all available margins of behavior – what Oliver Williamson calls "opportunism with guile".<sup>31</sup> This kind of competition is *efficiency and welfare reducing*, not enhancing as in the perfect competition model. It deters firms from investing in the industry in the first place, thereby causing a social welfare loss for firms, suppliers and consumers.

What is the solution to this inefficiency? Since the first cause of the problem (sunk investments) is unavoidable, the solution is for firms to be allowed to require long-term, fixed price contracts with their suppliers and customers. These contracts *restrict the power of suppliers and customers to affect market prices through search, bargaining and switching*. Such vertical restrictions prevent them from maximizing their power on all available margins,

---

<sup>30</sup>By agreeing to that price, the firm's average total revenue would fall below its average total cost, thereby saddling it with net losses and requiring it to exit the industry.

<sup>31</sup>Williamson 1985.

thereby avoiding inefficient competitions in raising relative pricing power.

The concept of "transaction costs" is often interpreted to mean contract enforcement costs – the costs of detecting contract violations and punishing the violators. Yet it also includes search costs,<sup>32</sup> renegotiation costs,<sup>33</sup> and switching costs.<sup>34</sup> Thus, in settings where unrestricted search, bargaining and switching cause hold up problems, it is a transaction cost problem. In such settings vertical restrictions are efficiency and welfare enhancing. This is how an economist would explain the use of vertical restrictions by firms if he were employing the transaction cost economics standard.<sup>35</sup>

### 3.2.2 The cutthroat competition problem

The problem of cutthroat competition is similar in its causes and its solution. Briefly, once a firm has made a sunk investment in production for one industry, it is vulnerable to "cutthroat competition" from other firms in that industry in the event that market demand falls too low. In this event the other firms might cut their selling prices by so much to attract consumers that the firm would have only two bad options if it wants to continue competing in the industry. On one hand, it could cut its own selling price to retain consumers, but at the expense of running net losses and being forced out of business eventually. In this case it would lose the value of its sunk investment. On the other hand, it could leave its selling price high enough to reap a net gain on each sale, but at the expense of losing so many customers to the other firms that it is forced out of business eventually. Again it would lose the value of its sunk investment.

To avoid losing sunk investments in this way, firms will either not invest in such industries in the first place or form horizontal agreements with each other to restrict prices or quantities (e.g. price floors or production ceilings). The purpose of such agreements is to insure that each firm makes non-negative profits. If such horizontal restrictions are prohibited (based

---

<sup>32</sup>Barzel 1997.

<sup>33</sup>Fudenberg and Tirole 1990.

<sup>34</sup>P. Klemperer 1995.

<sup>35</sup>See Williamson (1975), Klein, Crawford, and Alchian (1978), Grossman and Hart (1986), and Klein (1996) for the holdup problem. See M. Katz (1989) for a review of the economic literature on vertical market relationships and contracting arrangements.

on the misconception that they are collusive, anticompetitive and efficiency reducing), then firms will not invest in such industries in the first place. Their non-investment will be a social welfare loss for firms and consumers alike.

The cutthroat competition problem is caused by two factors: (1) firms have made sunk investments and (2) consumers have an unrestricted right to search for a better price, bargain for a better price, and switch firms to obtain a better price. When consumers have this unrestricted right, it provokes a competition among firms to attract customers by threatening to reduce the value of another firm's sunk investment so low that it must go out of business. That is, it provokes a competition in raising relative pricing power through cutthroat competition. This kind of competition is *efficiency and welfare reducing*, not enhancing as in the perfect competition model. It deters firms from investing in the industry in the first place, thereby causing a social welfare loss for firms and consumers.

The solution to this inefficiency is to allow firms to form long-term horizontal contracts with each other and long-term vertical contracts with customers. The horizontal contracts *restrict the power of competing firms to affect market prices* by searching for new customers and offering them lower prices. The vertical contracts *restrict the power of consumers to affect market prices* by searching and bargaining for better prices and threatening to switch firms for a better price. Such restrictions prevent firms and consumers from maximizing their power on all available margins, thereby avoiding inefficient competitions in raising relative pricing power. This is how an economist would explain the use of horizontal and vertical restrictions by firms if he were employing the transaction cost economics standard.<sup>36</sup>

### 3.2.3 The analogy between firms and states

The economist who is trying to explain cooperation among firms has a choice whether to use the perfect competition standard or the transaction cost economics standard. The same choice arises in efforts to explain cooperation among states.

---

<sup>36</sup>See Bittlingmayer (1982), Bittlingmayer (1985), Telser (1987), Pirrong (1989), Pirrong (1992), and Sjostrom (1989) for the problem of cut-throat competition. See Pirrong (1989) for a comparative analysis that identifies how industry structure determines the combination of horizontal and vertical restrictions that are needed to avoid cutthroat competition. Economists also note that firms may adopt strategies of product differentiation to retain customers, rather than adopting such restrictions or lowering prices to an unsustainable level.

### 3.3 Explaining inter-state cooperation: The "perfect competition" standard versus the "transaction cost economics" standard

When two states form a free trade agreement, how might an economist explain it? It depends on which of the two efficiency standards he uses: the perfect competition standard or the transaction economics standard. Similarly, when two states engage in a tariff war, how might an economist explain it? Again it depends on which of the two efficiency standards he uses. Consider each standard in turn.

#### 3.3.1 The "perfect competition" standard

Suppose that two states agree to eliminate tariffs and form a free trade agreement. How would the economist explain their cooperation using the perfect competition standard? He would assume the two states understood their interests in the economic efficiency gains from unrestricted international competition *among firms*, and they formed the free trade agreement to realize these gains. That is, the states eliminated the tariffs because they constituted anticompetitive restrictions on international competition *among firms*, and the states enacted free trade to create unrestricted competition *among firms*. The economist would conclude that the states were *reaping efficiency and social welfare gains as a result of agreeing to allow unrestricted competition among firms*.

Now suppose the two states get into a trade dispute and descend into a new tariff war. How would an economist explain the new tariffs using the perfect competition standard? He would assume the two states simply failed to understand their interests in the economic efficiency gains from unrestricted international competition among firms. He would assume the tariffs constitute anticompetitive restrictions on international competition among firms. Each state is protecting its domestic firms from competition from foreign firms, thereby allowing the domestic firms to gain a greater share of the domestic market than they would be able to gain if their share were determined by their productive efficiency relative to foreign firms. The economist would conclude that this outcome is inefficient from a social welfare perspective, *because it constitutes an anticompetitive restriction on competition among firms*.

### 3.3.2 The "transaction cost economics" standard

Suppose again that two states agree to eliminate tariffs and form a free trade agreement. How would the economist explain their cooperation using the transaction cost economics standard? He would assume the tariffs had been a consequence of unrestricted policy competition *among states* to raise their relative power to affect market price through holdups, that is, through threats to impose tariffs on imports to devalue the opposing state's investments in productive capacity. The unrestricted right to commit holdups was deterring productive economic investment and reducing industrial efficiency and social welfare. These efficiency and welfare losses were due to unrestricted policy competition among states to raise their relative power to affect prices through tariffs. The economist would assume the two states understood their interests in the economic efficiency gains from restricting policy competition *among states*, and they formed the free trade agreement to realize these gains. That is, the states eliminated the tariffs because they had resulted from unrestricted policy competition *among states*, and the states reached a free trade agreement to impose restrictions on policy competition *among states*. The economist would conclude that the states were *reaping efficiency and social welfare gains as a result of agreeing to adopt restrictions on competition among states*.

Now suppose the two states get into a dispute over the terms of their free trade agreement and descend into a new tariff war. How would an economist explain the new tariffs using the transaction cost economics standard? He would assume the states' prior free trade agreement was an agreement to prohibit holdups (as described in the previous paragraph). He would assume the agreement was an *incomplete contract* – one that did not cover all possible contingencies that might arise in the future. Then he would examine the facts surrounding the initiation of the tariff war to search for evidence that it arose from a dispute over the contract terms.

### 3.3.3 Comparison between the perfect competition standard and the transaction cost economics standard

These are two completely different explanations of the eruption of costly conflict. The first explanation focuses on the *inefficiency* of costly conflict from the perspective of the

perfect competition standard. The second explanation focuses on the *contractual causes* of costly conflict from the perspective of transaction cost economics. This difference is highly relevant to international relations theorists who seek to provide rationalist explanations of costly conflict. To understand its relevance, it is first necessary to draw an analogy between competition among firms and competition among states. This analogy will illuminate a number of methodological issues.<sup>37</sup>

### **3.4 The analogy between market competition among firms and security competition among states**

#### **3.4.1 Assumptions of the model of market competition among firms**

1. Each firm's market share is determined by its productive efficiency relative to other firms. The more efficient firms gain a larger market share; the less efficient firms gain a smaller market share.<sup>38</sup>
2. A firm's relative efficiency determines its cost of gaining a larger market share. A more efficient firm can offer inframarginal consumers the same product at a lower production cost and a lower selling price, thereby gaining a larger market share; a less efficient firm cannot match the offer, thereby having to settle for a smaller market share.
3. Each firm searches for a larger market share by: (1) searching across consumers for information about their willingness to pay and (2) offering consumers a price discount to induce purchasing behavior that reveals the information.
4. The unrestricted right of firms and consumers to search, bargain and switch transaction partners induces an efficient outcome:
  - (a) All pareto-improving informational and market exchanges are consummated.
  - (b) The firms' relative market shares reflect their relative productive efficiency.
  - (c) Cut-throat competition does not occur because, by assumption, no firm makes any sunk investment. Each firm's production function has strictly increasing average cost. Hence even large drops in demand never saddle a firm with net losses that

---

<sup>37</sup>In technical terms it is not just an analogy but a mathematical isomorphism.

<sup>38</sup>This assumes that each firm's production function is characterized by strictly increasing marginal cost. If each firm's function has constant marginal cost, then the most efficient firm gains the entire market and the rest of the firms gain no share of the market and drop out of competition.



would require it to go out of business. So no firms are deterred from entering the market in the first place for fear of losing the value of a sunk investment.

5. If a less efficient firm gains a greater market share than a more efficient firm, how would it be explained using the perfect competition standard of efficiency?
6. The natural explanation would be that it is due to some kind of *restriction* on the right to search for information, bargain over prices and switch firms to obtain a better price (e.g. horizontal or vertical restrictions such as collusion, exclusive dealing, market foreclosure, etc.).
7. The solution to the inefficiency and the social welfare loss is to eliminate the restriction on competition.

### **3.4.2 Assumptions of the model of security competition among states**

1. Each state's share of the international system's assets is determined by its relative power. The more powerful states gain a larger share of the system's assets; the less powerful states gain a smaller share.
2. A state's relative power determines its cost of taking a larger share of the system's assets by force (e.g. by kinetic security competition). For the state with a relatively large population, economy and technology base, the opportunity cost of transferring civilian workers into the military sector to take another state's assets by force is relatively low. Conversely, for the state with a small population, economy and technology base, the opportunity cost of transferring the same number of civilian workers into the military sector to take another state's assets by force is relatively high. Hence the more powerful states can take their neighbors' assets by force at a lower cost, thereby gaining a larger asset share; the less powerful states cannot match the strategy, so they must settle for a smaller asset share.
3. Each state's share of the system's valuable (non-military) assets is determined by its power relative to other states. In other words, the international distribution of assets is determined by the international distribution of power. Each state searches for a larger asset share by: (1) searching across other states for information about their power and (2) offering them a peace dividend in return for disclosing information about their

power. Whenever relative power shifts between two states, the shift induces a peaceful transfer of assets between the two states in order to bring the international distribution of assets into alignment with the international distribution of power.

- (a) A state could bluff that its power had increased when it really had not in order to gain more than it could ever gain based on its actual power; or a state could bluff that its power had not decreased when it really had an order to lose less than it would lose based on its actual power.
- (b) To deter such bluffing, each state threatens to switch from peace to war against the other state if it refuses to disclose accurate, verified information about its power. In other words, each state offers the other state a peace dividend (or "discount" on the cost of the asset transfer) in return for disclosing accurate, verified information about its power.

In a perfectly competitive security equilibrium with no restrictions on such informational exchanges, all states disclose their actual power information fully and verifiably, and no conflicts are needed to deter bluffing about power. Whenever relative power shifts between two states, each state reveals its new power information to the other state fully and verifiably, the weaker state gives a share of its assets to the stronger state without conflict, and both states receive a peace dividend.

4. These informational exchanges are pareto-improving for all states in the sense that they avoid costly conflicts due to bluffing about power. In this sense this world of perfect security competition is efficient. Specifically, when states have an unrestricted right to search for information about power, bargain over its disclosure, and switch from peace to war to induce full disclosure, the resulting equilibrium is efficient:
  - (a) All pareto-improving informational exchanges are consummated.
  - (b) The states' relative asset shares reflect their relative power.
  - (c) Costly conflicts do not occur because, by assumption, each state's right to search, bargain, and switch from peace to war induces full information disclosure by all states. So no conflicts are needed to deter any state from bluffing about its power.
5. If costly conflicts erupt in order to deter bluffing about power, how would it be explained using the perfect competition standard of efficiency?
6. The natural explanation would be that it is due to some kind of restriction on states'

right to search for information about relative power, bargain over its disclosure, and switch from peace to war to induce full disclosure.

7. The solution to the inefficiency, that is, the solution to the problem of costly conflict, is to eliminate the restriction on states' right to search, bargain and switch.

With this analogy between market competition and security competition in mind, we can now examine the rival explanations of costly conflict mentioned above. Again, the first explanation focuses on the *inefficiency* of costly conflict from the perspective of the perfect competition standard. The second explanation focuses on the *contractual causes* of costly conflict from the perspective of transaction cost economics.

### **3.5 "Rationalist" explanations of costly conflict: What kind of rationality are statesmen assumed to exercise?**

#### **3.5.1 The conventional puzzle of costly conflict – and its conventional resolution**

In his seminal article, Fearon (1995) framed the puzzle of costly conflict as a problem of economic inefficiency. He claimed that any rationalist account of costly conflict must explain why the states that engaged in conflict failed to adopt the efficient policy of remaining at peace and saving the conflict costs. His strategy to resolve this puzzle was to posit a world of perfect competition among states based on relative power. In this world states divide valuable assets according to relative power on the margin at all times, not only in war but in peace as well. Hence each state has an incentive to bluff that its power is greater than it really is in order to gain more than it could ever gain based on its actual power level. The only way for one state to deter the other state from bluffing in this way is to initiate war against it periodically in response to at least some of its claims that its power has increased.

This resolution of the puzzle of costly conflict is based on a counterfactual assumption that seems simple enough. If each state had accurate information about the relative power of other states, there would never be any conflicts of this kind. A state would only demand a greater share of the system's assets if its power had increased by enough that it could seize the share by force *with a net gain* (if its demand were rejected). No state would ever demand a greater share of the international system's assets backed by a bluff that its power

had increased by this much when it really had not. Assuming that each state maximizes its income at all times, this is a world of perfect international security competition. Valuable assets would be distributed among states according to their relative power at all times – and redistributed as necessary whenever their relative power changed.

### 3.5.2 The conventional resolution of the puzzle is based on the perfect competition standard of efficiency

Fearon's analysis produces a clear implication.<sup>39</sup> Since each state seeks to maximize its income, each state has an incentive to increase *both* states' information about relative power in order to avoid costly conflicts and thereby save the conflict costs. In the perfectly competitive security equilibrium, accurate and verified information about relative power would be generated and costly conflicts would be avoided. Empirically, there is both historical and contemporary evidence to suggest that states go to some lengths to generate and exchange accurate information about relative power. Thus Fearon's analysis explains costly conflict by assuming that states simply fail to understand their interests in the mutual efficiency gains from exchanging accurate information about relative power. This explanation is analogous to the economist's explanation of tariff wars based on the perfect competition standard of efficiency – that states simply fail to understand their interests in the mutual efficiency gains from perfect competition among firms.

His analysis thus raises a critical question. Why would states ever *fail* to generate accurate information about relative power given their mutual interest in doing so to avoid costly conflicts? Fearon does not answer this question. He simply *assumes* that states fail to understand their interests in the mutual efficiency gains from exchanging accurate information about relative power. To be precise, he assumes that states *do not have the unrestricted option* to search for information about relative power and bargain for its verified disclosure by threatening to switch from peace to war if it is not disclosed.

---

<sup>39</sup>Although it is not clear whether he explored it.

### 3.5.3 Is the conventional resolution really a *rationalist* explanation of costly conflict?

What kind of rationality are states exercising in this model of security competition? Wouldn't a rational state exercise its option to search for information about relative power and bargain for its verifiable disclosure by threatening to switch from peace to war if it is not disclosed? By assuming *a priori* that states fail to exercise this option, Fearon seems to be assuming that states are irrational. In this sense his model does not seem to provide a "rationalist" explanation of costly conflict.<sup>40</sup>

Is there an alternative way to explain costly conflict based on the other efficiency standard – the transaction cost economics standard? Yes, and it is directly analogous to the explanation of tariff wars based on the transaction cost economics standard provided above. The explanation starts by assuming that peace between two states is a sunk investment for each state. The investment's full value cannot be recouped completely in the event that one state switches from peace to war (thereby forcing the other state to switch as well). The main purpose of a peace agreement, therefore, is to impose restrictions on inter-state policy competition in order to prevent holdups. This kind of peace agreement is an *incomplete contract*, however. It does not cover all possible contingencies that might arise in the future. Once a war has erupted, therefore, the researcher's task is to examine the historical record to search for evidence that the conflict arose from a dispute over the terms of the peace contract.

## 4. Two views of social optimality: the "welfare economics" view versus the "political economics" view

### 4.1 The "determinacy paradox"

Traditionally, when economists wrote models of markets, they assumed the economic policies governing the markets were made by a "benevolent social planner" – a person whose only

---

<sup>40</sup>The other main rationalist explanation of costly conflict in the literature (Powell 2006) suffers from a similar criticism. See the literature review chapter in Part II of this study for a full critique on these grounds.

goal was to maximize social welfare. In the early 1980s, however, economists began to write a different kind of market model, one in which economic policy was made by a self-interested policymaker whose only goal was to maximize his own "private" welfare.<sup>41</sup> Economists called them "endogenous policy models" to connote that policy determination was endogenous to the agents in the model rather than exogenous through the academic modeler's "benevolent social planner". Within a few years economists realized that these models generate a paradox. The economist who wrote the model could still make social welfare comparisons between different equilibria from the perspective of a benevolent social planner. For example, an equilibrium with no trade tariffs is socially optimal, while an equilibrium with high trade tariffs is socially suboptimal. But such social welfare comparisons are meaningless, because they would not have the same meaning to the endogenous policy maker in the model. For example, the academic modeler could identify a trade tariff as socially suboptimal from the perspective of a benevolent social planner. Yet the endogenous policymaker would identify the same tariff as privately optimal from the perspective of his own self-interest and that of the industry protected by the tariff. So the academic modeler's claim that the tariff was suboptimal would be incomprehensible to the endogenous policymaker. The claim would have no meaning to him. This is the "determinacy paradox": The academic modeler can still make social welfare determinations, but his determinations are not comprehensible to the real world policymakers he might wish to advise in order to improve policy. So it is not meaningful for the academic modeler to make such determinations in the first place.<sup>42</sup>

Years after the determinacy paradox was identified, some law professors reframed it as the "inside/outside fallacy". The heart of the problem, in their view, was that many law professors were diagnosing bad policies as a consequence of the self-interested motivations of policymakers, but then advising the same policymakers to adopt better policies under the assumption that they would be public-spirited enough to follow the advice.<sup>43</sup> The message of the "inside/outside fallacy" is that academics are being inconsistent when they go "back and forth" between an "outside" perspective that assumes policymakers have self-interested

---

<sup>41</sup>Hillman82; Magee82.

<sup>42</sup>See Bhagwati, Brecher, and Srinivasan (1984) for the original insight. See T.N. Srinivasan (1997) for a special issue of the journal *Economics and Politics* containing further analysis.

<sup>43</sup>Posner13.

motivations and an "inside" perspective that assumes policymakers have public-spirited motivations.

## 4.2 Implications of the determinacy paradox for explanatory theories of costly conflict

What exactly does an endogenous policy model of tariff-setting explain? It certainly explains why the policy maker adopts a socially suboptimal trade policy. However, it does not explain why a policy maker *who would have wanted to adopt a socially optimal trade policy if possible* failed to do so. For in a model of endogenous trade policy, the policymaker does not want to adopt a socially optimal trade policy in the first place. His utility function does not include society's interest (in the way the utility function of a benevolent social planner includes it). His utility function only includes his own private interest.

This clarification is illuminating for "rationalist" explanations of costly conflict. What exactly does the standard rationalist explanation of costly conflict explain? It certainly explains why the two states engage in a costly conflict that makes both states worse off than they would have been remaining at peace. However, it does not explain why states *that would have wanted to avoid this inefficiency if possible* failed to do so. For, in the standard rationalist explanations of costly conflict (Fearon 1995; Powell 2006), neither state has the policy instruments and goals it would rationally want to have to avoid this inefficiency.<sup>44</sup>

## 4.3 The true puzzle of costly conflict

According to the existing literature, the puzzle of costly conflict is: Why would two rational states ever go to war with each other when they would be better off settling their disputes peacefully and saving the costs of war. What exactly is it that makes the occurrence of costly conflict a *puzzle*? What makes it a puzzle is *not* simply that each state incurs war costs and would have been better off settling its disputes with the other state peacefully. This

---

<sup>44</sup>In the model provided by Fearon (1995), one state would rationally want to threaten the other state with war if it fails to disclose accurate, verified information on its power level. In the model provided by Powell (2006), one state would rationally want to threaten the other state with war if it fails to agree to halt or reverse the increase in its own power. See Section 1.1.5 above and the literature review chapter in Part II for a full critique along these lines.

observation alone is not necessarily puzzling if each state's policymaker is self-interested rather than being interested in the welfare of his state as a whole. In fact, even if the policymaker is interested in the welfare of his state as a whole, the occurrence of costly conflict still may not be puzzling. For, if the academic modeler does not endow the policymaker in the model with the policy instruments and goals he needs to maximize his state's social welfare by settling disputes peacefully, then there is no puzzle. Such a hypothetical policymaker failed to maximize his state's social welfare simply because he was denied (by the modeler) the policy instruments and goals he would have wanted to use to do so.

A real puzzle would arise only if the hypothetical policymaker in the model was interested in maximizing his state's welfare by settling disputes peacefully *and* he had the policy instruments and goals he would have wanted to use to do so – yet he *still* failed to do so. *This* would be a real puzzle. From this perspective, the academic modeler who is trying to explain costly conflict must focus on these key questions:

1. What obstacles prevented the policymakers from avoiding costly conflict?
2. Were the policymakers aware of these obstacles?
3. Did they want to overcome these obstacles?
4. Did they make efforts to overcome these obstacles?
5. Did their efforts fail, and if so, why?

Thus the true puzzle of costly conflict is: Why did rational policymakers who saw the obstacles to peaceful settlement, wanted to overcome them, and tried to overcome them, still fail to overcome them? The rationalist explanations of costly conflict in the existing literature do not resolve this puzzle.<sup>45</sup> For each of those models explains costly conflict as a consequence of an obstacle to peaceful settlement that the hypothetical policymaker in the model was not aware of, did not want to overcome, and made no effort to overcome.

---

<sup>45</sup>Fearon 1995; Powell 2006.



#### 4.4 Reframing the "determinacy paradox" and the "inside/outside fallacy" for positive (explanatory) theory building

The message of the determinacy paradox is that it is meaningless for the academic modeler to make welfare assessments of a hypothetical (modelled) world in which the policymakers themselves have no interest in heeding those assessments in their policy decisions. Similarly, it is meaningless for the academic modeler of costly conflict to make the welfare assessment that costly conflict is *inefficient* in his hypothetical (modelled) world, when the policymakers in that world are not aware of the putative obstacle to efficiency and have no interest in overcoming it. Such a model may explain why policymakers in such a world engage in costly conflict. But it has not explained why policymakers *who would have wanted to avoid that inefficiency* failed to do so. Therefore, it has not resolved the true puzzle of costly conflict (stated above).

The implication is that the theorist of costly conflict must examine the historical evidence in detail to determine the obstacles to peaceful settlement *that the historical actors themselves perceived that they faced and wanted to overcome*. And the theorist must explain why the actors failed to overcome those obstacles. This is the practical implication of the determinacy paradox for positive explanatory theory. When it comes to explaining the real world causes of inefficient outcomes in the historical record, only the inefficiencies which the historical actors themselves perceived and tried to overcome, but failed to overcome, are valid objects of explanation. That is, if the theorist wishes to claim to be resolving a true puzzle of inefficient behavior in the historical record. In short, it is the historical actors' subjective rationality – and their failure to realize its promise in a crisis – which must be explained, not the ideal type of rationality that the academic modeler imports into their world from his ideal world of welfare economics. This is the message of the determinacy paradox for explanatory theorists of costly conflict.

### 5. Reductionism and holism in the explanation of long-run history

If there is a single theme that runs through this chapter, it is the contrast between reductionist methodologies and holist methodologies. A reductionist methodology assumes that

the whole system is *reducible* to the sum of its parts. Hence a model of the parts alone can explain systemic outcomes. By contrast, a holist methodology assumes that the whole system is *greater* than the sum of its parts. Hence a model of the parts alone cannot explain systemic outcomes. Many of the methodological debates of the last few decades can be seen as attempts to determine the proper roles of reductionism and holism in a scientifically progressive research program. Each section of this chapter illuminated one aspect of these debates and drew the lessons for how to construct a theory of conflict and settlement that can explain long-run history.

Section 2 examined competing methodological standards for defining the rationality of actors in the system and modeling their understanding of the system as a whole. One standard defines actors' rationality in terms of their skillfulness at solving mathematical optimization problems with unbounded calculation capacity and knowledge of the system's structure. This standard assumes that a model of the actors alone can explain systemic outcomes. The other standard defines actors' rationality in terms of their adaptability to unpredictable changes in their material environment despite having bounded knowledge and calculation abilities. This standard assumes that a model of the actors alone cannot explain systemic outcomes.

The two standards yield contrasting views of the role of rationality in conflict and the relation between security competition and institutions in enabling peace and stability. The first standard assumes that peace and stability result from the actor's knowledge of all possible strategies among himself, his potential allies, and his potential opponents, and his capacity to calculate his optimal deterrence strategy in light of them. The second standard assumes that peace and stability result from mutual adherence to contracts and institutions that make up for actors' limits in knowledge of their options and capacity to calculate optimal strategies. This contrast challenges theorists to interrogate the historical evidence to determine whether conflicts were ended by settlements that expanded the actors' strategic options or limited them, and new conflicts erupted from factors that blocked strategic knowledge and calculation or from contract disputes.

Section 3 examined competing methodological standards for gauging efficiency in a competitive equilibrium and explaining the roots of efficient equilibria. One standard conceptu-

alizes efficiency as arising from the actors alone and explains efficiency as a natural consequence of their unrestricted competition. The other standard assumes that such unrestricted competition generates inefficient outcomes and the main source of efficient outcomes is the emergence of system-level restrictions on the actors' competition. In so far as the forms of these restrictions cannot be explained by the properties of the actors alone, the system is greater than the sum of its parts.

The two standards yield contrasting explanations of the conditions for peace and stability and the causes of conflicts and competitions in raising relative power. The first standard assumes that peace results from unrestricted security competition that generates efficiency-maximizing exchanges of information about relative power and exchanges of commitments to limit the size of power shifts. War is caused by factors that interfere with such unrestricted security competition. By contrast, the second standard assumes that unrestricted security competition causes inefficient competitions in raising relative power and conflicts. Peace and stability can only be achieved through contractual agreements to restrict security competition. The main cause of conflicts, in turn, is disputes over the terms of these contracts. These differing explanations of war challenge theorists to interrogate the historical evidence to determine whether historical statesmen were driven to conflict by (1) an inability to exchange information about relative power and commitments to limit power or (2) disputes over the contract terms delimiting their peaceful coexistence.

Section 4 examined competing methodological standards for conducting welfare assessments of systemic outcomes and conceptualizing the role and point of view of the assessor. One standard conceives the assessor as standing above and beyond the system in a way that enables totally objective welfare assessments – and seamless conveyance of these assessments to actors in the system in order to induce welfare improving behavioral changes. The other standard conceives the assessor as standing within the system just like any of its other actors, and able to influence policies and behaviors in the same way they do – yet unable to conduct totally objective welfare assessments because he is embedded within the system.

The two standards yield contrasting methods for explaining costly conflicts as somehow rational. The first standard allows the theorist to posit an obstacle to the peaceful settlement of disputes that statesmen were not aware of, did not want to overcome, and made no effort

to overcome. By contrast, the second standard requires the theorist to posit only obstacles to the peaceful settlement of disputes that statesmen were aware of, wanted to overcome and made efforts to overcome – but still failed to overcome. Only such obstacles pose a true puzzle of costly conflict (e.g. how could rational statesmen who wanted and tried to overcome the obstacle still fail?). These contrasting methods challenge theorists to interrogate the historical evidence to determine whether historical statesmen were driven to conflict by (1) obstacles to peaceful settlement that they were unaware of, did not want to overcome, and made no efforts to overcome; or (2) obstacles to peaceful settlement that they were aware of, wanted to overcome and made efforts to overcome – but still failed to overcome.

In the next chapter I will examine competing methodological standards for conceptualizing causal mechanisms. One standard assumes that human systems can be modeled in the same way that physicists model the interactions of atoms and photons – by assuming that human behavior is governed by invariant laws of nature more than any other factor. From this perspective the system is reducible to the sum of its parts. The other standard assumes that human systems are better modeled in the way that evolutionary biologists model the survival fitness of species, the extinction of existing species and the emergence of new species – by assuming that human behavior is governed in part by structural factors but in even larger part by the force of contingency. While invariant laws of nature establish a wide channel within which the course of history must run, the actual path it takes within that channel cannot be explained by reference to such invariant laws. From this perspective the system is greater than the sum of its parts.

The next chapter will also examine competing methodological standards for coding empirical evidence for the purpose of causal inference. One standard assumes that actors and cases are homogeneous and causal mechanisms are stationary over time, making the system reducible to the sum of its parts. The other standard assumes that the properties of actors, cases and causal mechanisms emerge and change over time, rendering the system greater than the sum of its parts.

## CHAPTER 8

### Historical explanation *versus* statistical explanation

#### 1. The historical method of explanation

##### 1.1 The historical method answers questions the statistical method cannot answer

Among economists there is a line of debate between those who see the field as a kind of social physics and those who see it as a kind of evolutionary biology that focuses on a uniquely intentional creature.<sup>1</sup> The physics approach assumes that actors' behavior is governed by *invariant laws of nature* more than any other factor. It also assumes that actors are homogeneous across geographical space and time. The critics of the physics approach claim that none of these assumptions is valid, and therefore it is scientifically counterproductive to write models of human behavior in the way that physicists write models of atoms and photons. Actors are not homogeneous in the way that atoms and photons are. Their behavior is not governed primarily by invariant laws of nature. While there may be invariant laws that establish a wide *channel* within which human behavior must travel, that channel is so wide that even a complete understanding of such laws would not supply explanations for most human actions. The impossibility of law-like explanation does not preclude all understanding and explanation, however. It simply means that the physics approach cannot achieve it and another approach is needed.<sup>2</sup>

One knowledgeable observer described the difference between the physics approach and the evolutionary biology approach in this way – noting that the physics approach is what

---

<sup>1</sup>See Mirowski (1988) and Mirowski (1989) for the debate. See Epstein et al. (1996), Epstein (2006), and Epstein (2014) for the evolutionary approach.

<sup>2</sup>I will return to the issue of the "channel" set by invariant laws and the "path" through that channel set by other factors.

most people think of as *the* scientific method.

"The stereotype of the 'scientific method' has no place for [attempts to explain] irreducible history. Nature's laws are [taken to be] defined by their invariance in space and time. The techniques of controlled experiment, and reduction of natural complexity to a minimal set of general causes, presuppose that all times can be treated alike and adequately simulated in a laboratory....

"But suppose you want to know why dinosaurs died, or why mollusks flourished while [the species] *Wiwaxia* perished? The....restricted techniques of the 'scientific method' cannot get to the heart of this singular event.... *The resolution of history must be rooted in the reconstruction of past events themselves – in their own terms – based on narrative evidence of their own unique phenomena.* No law guaranteed the demise of *Wiwaxia*, but some complex set of events conspired to assure this result – and we may be able to recover the causes of it if, by good fortune, sufficient evidence lies recorded in our spotty geological record."<sup>3</sup>

In other words, there are a whole set of causal mechanisms whose operation simply cannot be proven by laboratory experiments using randomized treatments or large-*n* statistical studies assuming that actors are homogeneous across cases and causal mechanisms are stationary across time.

## 1.2 Causal mechanisms that are undetectable by statistical analysis

There are several reasons why a causal mechanism might be invisible to statistical analysis. First, the causal mechanism itself might create a pattern of empirical evidence that cannot be seen from the angle taken by standard regression analysis, but only from a different angle. An excellent example is given by Goertz and Mahoney (2012) in their study of the difference between the "statistical culture" and the "set-theoretic culture" of causal inference.

"Geddes (1991; 2003)...presents bivariate data on the relationship between GDP per capita growth and labor repression using a larger sample of....32 developing countries.... As the flat regression line...illustrates, there is no linear relationship [between the two variables] with these data (the slope is .09 and  $R^2 = .003$ ). Hence, when Geddes looks at the scatter plot,...she *sees* no relationship between labor repression and economic growth.

"However, the data do suggest the possibility of an important relationship between labor repression and growth when viewed from a set-theoretic perspective. The data have an empty region, a fact we call attention to by emphasizing the

---

<sup>3</sup>Gould89.

region with no observations in the upper left quadrant.... This empty space is what one would expect to see if the following hypothesis were true: labor repression is a necessary condition for high economic growth. Thus, while the data do not support a linear relationship hypothesis, they do seem consistent with a necessary condition [hypothesis]. All cases of exceptional growth have at least moderate levels of labor repression.... Thus, at least moderate levels of labor repression may be necessary for sustained high growth in this population. Qualitative scholars who have a set-theoretic causal model in mind cannot help but notice this aspect of the data.”<sup>4</sup>

A second reason why a causal mechanism may be invisible to statistical analysis is that the data may simply be unavailable. For example, most regression studies of war initiation use a dataset on wars that erupted between 1816 and 2017. If a particular causal mechanism generated wars mostly before 1816, therefore, the data needed to test that hypothesis would not be available.

A third and deeper reason why a causal mechanism may be invisible to statistical analysis is that the relevant universe of cases may be undefinable, so a representative (random) sample of the universe of cases cannot be conducted. For example, Gould (1989) discusses the causal mechanism by which a set of 16 species went extinct millions of years ago and draws fundamental conclusions regarding the nature of history itself. The empirical evidence for his study came from a set of fossils that paleontologists discovered in the *Burgess Shale*, a rock formation in Canada.<sup>5</sup> It is unclear how many more fossils of extinct species exist to be discovered, if any. Hence, it is unclear what the universe of cases would be if one wanted to conduct a representative (random) sample of *all* of the species that ever went extinct.<sup>6</sup>

In sum, there are some causal mechanisms whose operation cannot be proven through large-*n* statistical studies. If one wants to prove their operation and explain how they caused certain outcomes in history, one needs a different method of explanation.

---

<sup>4</sup>Emphasis in original. Goertz and Mahoney 2012, pp. 179–81.

<sup>5</sup>Gould’s book is a survey and extension of the conclusions reached by a number of paleontologists who collected and analyzed the evidence from the Burgess Shale.

<sup>6</sup>One could draw a random sample of the cases of extinction for which evidence already exists. But then one’s conclusions might be distorted by sample truncation bias – and corrections would be impossible because the structure of the truncation would not be knowable.

### 1.3 The logic of the historical method of explanation

"Historical explanations take the form of narrative: E, the phenomenon to be explained, arose because D came before, preceded by C, B, and A. if any of these earlier stages had not occurred, or had transpired in a different way, then E would not exist (or would be present in a substantially altered form, E', requiring a different explanation). Thus, E makes sense and can be explained rigorously as the outcome of A through D. But no law of nature enjoined E; any variant E' arising from an altered set of antecedents, would have been equally explicable, though massively different in form and effect.

"I am not speaking of randomness (for E had to arise, as a consequence of A through D), but of the central principle of all history – *contingency*. A historical explanation does not rest on direct deductions from laws of nature, but on an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result. This final result is therefore dependent, or contingent, upon everything that came before – the unerasable and determining signature of history."<sup>7</sup>

Notice that the impact of contingency on history runs very deep. It is not only the *final* outcome E that is unpredictable by any law of nature. Each of the *preceding* steps in the sequence of events that caused that outcome is equally unpredictable by any law of nature. Thus the sequence from A to E is simply a "slice of time" from a much longer sequence of events running from the very beginning to the present day – *all* of it determined by contingency *more than any other factor*. This truth of history does not rule out the impact of other factors that operated in a more systematic or 'scientific' way. It simply means that such factors could only establish a wide channel within which the course of history had to run. They can never explain what particular path history took within that channel. For this a specifically *historical* explanation is necessary.

Given the very nature of history itself, all that the later theorist can do by way of explaining history is:

1. to propose a set of antecedent conditions that were individually necessary and together sufficient to generate outcome E, and
2. defend that hypothesis against competing hypotheses that propose other sets of antecedent conditions, by
3. deriving as many observable implications as possible from the competing hypotheses, and

---

<sup>7</sup>Gould89.



4. determining which implications are corroborated or refuted by the historical evidence.

Thus in any debate over the true cause of an historical event, the governing question is: What was the *set* of causal conditions that were individually necessary and together sufficient to cause that event.

## 1.4 Generalizing an historical explanation

Methodologists have shown that the concept of *causation* can be defined in two ways.<sup>8</sup>

1. **Constant conjunction.** "[C]auses are always followed by their effects."<sup>9</sup> "We may define a cause to be an object followed by another [object], and where all the objects, similar to the first, are followed by objects similar to the second."<sup>10</sup>
2. **Counterfactual.** "[I]f the first object had not been, the second [object] never would have existed."<sup>11</sup>

The first definition applies well to explanations that take the form of invariant laws of nature (or invariant laws of human behavior, if any).<sup>12</sup> By contrast, the second definition applies well to explanations of unique historical events. "When the focus is on individual events, the counterfactual account [of causation] is a natural choice..."<sup>13</sup>

The second definition, based on counterfactuals, forms the foundation of the set-theoretic method of causal inference for historical events. According to this method, the causal hypothesis is initially stated in the form of a counterfactual: If not *X*, then not *Y*. Later the same hypothesis is stated in the form of a *set of necessary conditions*, *X*, that must be present for *Y* to occur.

"The process through which qualitative researchers generalize counterfactuals suggests a...causal inference circle that begins with definition 2:

1. One starts with Hume's definition 2, which stresses the counterfactual.

---

<sup>8</sup>Goertz and Mahoney 2012, pp. 75–6.

<sup>9</sup>Goertz and Mahoney 2012, p. 76.

<sup>10</sup>**Hume75.**

<sup>11</sup>**Hume75.**

<sup>12</sup>Goertz and Mahoney 2012, p. 76.

<sup>13</sup>Goertz and Mahoney 2012, p. 80.

2. One interprets this definition in terms of logic: if  $X$  had not occurred, then  $Y$  would not have occurred, i.e. if not  $X_i$ , then not  $Y_i$ .
3. One generalizes the individual case counterfactual to all cases, i.e. if not  $X_i$ , then not  $Y_i$  for all  $i$ .
4. One converts this counterfactual into a general statement, using definition 1, about a necessary cause; that is,  $X$  is necessary for  $Y$ .
5. If  $X$  is present in case  $i$  then  $X$  is a cause of  $Y$ .

In this circle, the key move is [in step 4 with] the conversion of the individual case counterfactual into a regularity statement about a necessary cause. In effect, the analyst stays with definition 2 throughout the circle, bringing in definition 1 to produce a generalization across cases. The retention of definition 2 is accomplished by assuming that th[is] definition [of causation] can be directly extended to many cases, thus allowing for the generalization [from the case(s) studied to a wider set of cases].”<sup>14</sup>

In other words, this approach claims that a causal inference from one or a few historical cases can be ”converted” into a ”generalization” about causation in other cases. This kind of generalization is quite different from a statistical generalization, however. A statistical generalization is based ultimately on a *quantitative* correlation between  $X$  and  $Y$ . By contrast a set-theoretic generalization is based on the *logic of the causal mechanism* by which the presence of  $X$  causes  $Y$ , while the absence of  $X$  causes not- $Y$  instead.<sup>15</sup>

The linkage between the counterfactual, on one hand, and the set of necessary conditions, on the other hand, is central to this method of historical explanation. In effect, the researcher starts by undertaking a minute study of the historical facts of an event in search of the causal factors which led to the event  $E$ , and without which the event would not have occurred. After locating these factors in the historical record, the researcher converts them into a statement that they were the necessary conditions for that event to occur.

## 1.5 The rational choice foundations of the historical method of explanation

In the first stage of this sort of analysis, the researcher must ask a key question for each factor that appears like it *might* have been a necessary condition for event  $E$  to occur. If

---

<sup>14</sup>Goertz and Mahoney 2012, pp. 80–1.

<sup>15</sup>Statistical inference also includes a counterfactual logic, in so far as a regression analysis of many cases includes some cases with high values of  $Y$  and other cases with low values of  $Y$ . But the resulting counterfactuals are *numerical* counterfactuals based on the variation in the data set, not *necessary condition* counterfactuals based on the logic of the causal mechanism being tested.

that factor had *not* occurred, then would the final event E *not* have occurred either? This is where the rationality assumption discussed in the previous chapter comes into play, because the answer to that question often takes the form:

If A had not occurred in this historical case, then the actor would *reasonably have adapted* to his new situation by choosing to do E' rather than E. Indeed, in *other* historical cases where A did not occur, the actor did choose to do E' rather than E.

Thus *rational choice* comparisons *across* cases can increase the probability that a set-theoretic inference is valid in *each* case – even if there are not enough cases to perform a large-*n* statistical study. A related technique for increasing confidence in a set-theoretic inference is to derive as many observable implications as possible from the causal hypothesis.<sup>16</sup> When the two techniques are combined, it yields a form of proof of causation that is diametrically opposite in nature to the form produced by statistical regression – a proof that arises from the *diversity* of the dataset rather than its *uniformity*.

## 1.6 The historical method gains its power from the *diversity* of the cases, rather than the *uniformity* and *quantity* of the cases (as in the statistical method)

”[H]istorical science is not worse, more restricted, or less capable of achieving firm conclusions [simply] because experiment, prediction, and subsumption under invariant laws of nature do not represent its usual working methods. The sciences of history use a *different mode* of explanation, rooted in the *comparative and observational* richness of our data. We cannot see a past event directly, but [even hard] science is usually [also] based on inference, not unvarnished observation (you don’t see electrons, gravity or black holes either).

”The firm requirement for all science – whether stereotypical or historical – lies in secure testability, not direct observation. We must be able to determine whether our hypotheses are definitely wrong or probably correct (we leave assertions of certainty to preachers and politicians). *History’s richness drives us to different methods of testing, but testability is our criterion as well. We work with our strengths of rich and diverse data recording the *consequences* of past events; we do not bewail our inability to see the past directly....*

”The great 19th-century philosopher of science William Whewell devised the word consilience, meaning ‘jumping together’, to designate the confidence gained when many independent sources ‘conspire’ to indicate a particular historical pattern.

---

<sup>16</sup>King, Keohane, and Verba (1994) provide a rationale for combining the two techniques.

He calls the strategy of coordinating disparate results from multifarious sources *consilience of induction*.”<sup>17</sup>

Notice the role played by “the *consequences* of past events”. These consequences provide a rich and diverse dataset on which to test the *observable implications* of a scientific hypothesis about what *caused* those past events in the first place. In other words, the bulk of the convincing evidence that A, B, C and D caused E might *not* come from direct observation of A, B, C, and D. Rather it might come from the “strengths of rich and diverse data recording the consequences of” event E. This approach marks quite a profound shift in perspective from the statistical world view in which the only variables that can legitimately go in a regression equation designed to explain event E are variables that occurred *prior* to event E.

The historical method does not rule out the use of direct evidence on A, B, C, and D. However, it does not limit the researcher to using only such evidence – like the statistical method does.<sup>18</sup> Perhaps this is why Gould calls historical explanation “coordinating interpretation.... coordinating disparate results from multifarious sources.” The method situates event E in the context of its hypothesized causes (A, B, C, and D) *and* its multifarious consequences which are implied by those hypothesized causes. This rich contextual approach is fitting for a world in which *every* sequence of events A through E is only “a slice of time from a much longer sequence of events...all of it determined by contingency *more than any other factor*.”<sup>19</sup> In this kind of world, where the evidence is “so abundant and so diverse”, the best that a researcher can hope to do is to paint a picture of the “before, during and after” periods – a coordinating interpretation – that organizes the disparate evidence into a more coherent whole than any other competing causal hypothesis.

”We search for repeated pattern, shown by evidence *so abundant and so diverse* that no other coordinating interpretation could stand, even though any item, taken separately, would not provide conclusive proof.” **Gould89**

---

<sup>17</sup>**Gould89.**

<sup>18</sup>Of course there are statistical regression studies in which the dependent variable is one among many observable implications of the causal hypothesis. My point is simply that even in such a regression, the only variables that can legitimately go on the right-hand side of the equation are variables that occurred prior to the dependent variable on the left-hand side.

<sup>19</sup>Quoted from above.

While an analysis that accomplishes this feat may be the best that a researcher can do, it may still provide a better explanation than competing hypotheses (especially if their observable implications are refuted by the evidence) – and in this sense it may provide the best explanation.

William Whewell, the philosopher of science quoted above, called this method "consilience of induction". I employ this method in this study to compare:

1. seven historical cases of conflict and settlement between great powers,
2. five historical cases of conflict and settlement between a great power and a much smaller state, and
3. eleven historical cases of variation in the domestic institutions that govern domestic conflicts and settlements.

I find that "a particular historical pattern" is repeated across all 23 cases. In each case there is an oscillation between a world of law and a world of power. The transition from a world of law to a world of power is driven by the breakdown of anchor contracts and anchor based enforcement methods. The transition back to a world of law is driven by the formation of new anchor contracts and anchor based enforcement methods.

This "repeated pattern" becomes visible only by viewing the historical facts from the perspective of the historical actors themselves. In this respect this study is similar to studies of the extinction of existing species and the emergence of new species. The criteria for a species' *fitness* for survival are determined by its own immediate environment – a niche unique to its own time and place. This does not mean that the creature's survival or demise was totally arbitrary, uncomparable to any other such event. It simply means that the standards of fitness and efficiency that determined its survival or demise were *historically specific* and *historically bounded*. Similarly, in comparisons across the niches of human behavior, one is comparing subjectivities. In each niche behavior is understandable in terms of its own inner logic of rational adaptation to the immediate events, opportunities and constraints in that niche. Each niche is comparable to the other niches, but only by *not* drawing any aspect of the comparison too closely.

## 2. Causal mechanisms and coding rules

### 2.1 The "physics" view of causation *versus* the "generativist" view

As mentioned earlier, economists are divided between those who see the field as a kind of social physics and those who see it as a kind of evolutionary biology.<sup>20</sup> The physics approach assumes that actors are homogeneous across time and space and their behavior is governed by invariant laws of nature more than any other factor. A major alternative method of model building is the "generativist" approach (Epstein 1996, 2006, 2014). The generativist method allows for much greater heterogeneity across actors, transformation of actors across time, and evolution of causal mechanisms over time. The core assumption is that the features of individual actors, systemic outcomes and causal mechanisms *emerge* over time in a way that is not predictable from the properties of the individuals or the system at any one point in time. The features are *generated* as *emergent properties* of the individuals and the system. This method is well-suited to model a world created by "an unpredictable sequence of antecedent states"<sup>21</sup> – a world in which regularities certainly happen (regularly as it were) but at the end of the day contingency rules.

### 2.2 A generativist view of the standard statistical coding rules

The generativist perspective bears important implications for the way that historical evidence is coded in a theory of conflict and settlement. If it really is true that the features of actors and outcomes are generated as emergent properties over time, then each case of conflict and settlement will be more or less unique, so a completely uniform set of "cases" of conflict and settlement will never be observed in the historical record. And each actor will be more or less unique, so a completely homogeneous set of "representative agents" will never be observed in the historical record. Conventional coding rules assume uniformity of cases and homogeneity of actors. What happens when the real world's diversity of actors, cases and causal mechanisms are "shoehorned" into a dataset that assumes homogeneous actors and cases, and then "fed" into a regression equation that assumes a single, stationary causal

---

<sup>20</sup>See Mirowski (1988) and Mirowski (1989).

<sup>21</sup>Gould89.

mechanism generated the data? Three main things happen:

1. the problem of "mongrel regressions" that are completely uninformative about causation;
2. the fallacy of rational overdetermination (e.g. there are multiple rational causes of the same outcome, yet each one alone would have been sufficient to cause the outcome, so it is unclear which one *actually* caused it);
3. the problem of equifinality, which renders a regression uninformative as to which of many different causal mechanisms actually generated the outcome in question.

Faced with the realization that historical life is so frustratingly diverse, the theorist has two options. He can give up on the goal of generalization in theory building. Or he can adopt a different way of coding historical evidence based on a different epistemology (e.g. theory of how we know what we know).

### 2.3 Diversity-driven data collection, analysis and inference

A central claim of case study researchers is that real world causation yields such a diversity of causes and effects that conventional statistical methods are inadequate to understand it. A standard regression analysis often fails to model cause and effect in a way that captures the actual causal mechanism that generated the data. This critique has given rise to an alternative "diversity-oriented" approach to causal inference. This approach identifies the shortcomings of conventional regression and addresses them.<sup>22</sup>

"[The] aspects of conventional quantitative analysis that interfere with its use as a discovery tool [are]:

- (1) its dependence on populations [of cases] that are constituted prior to data collection and analysis;
- (2) its heavily variable-oriented discourse [about cases], a framework that is antithetical to the analysis of cases as interpretable configurations of aspects, and
- (3) its additive-linear view of causation, an understanding that depends upon strong homogenizing assumptions about cases, which, in turn, make this approach insensitive to causal complexity."<sup>23</sup>

The diversity-oriented method offers alternative conceptions on all three points:

---

<sup>22</sup>Ragin (2000) explains various critiques of conventional statistical methods, summarizes the diversity-oriented approach, and provides citations. He emphasizes that conventional statistical methods serve well in many applications, but are often misapplied to inference problems for which they are not well-suited.

<sup>23</sup>Underlining and paragraph breaks added for clarity of comparison. Ragin 2000, p. 14.

”[1] [R]esearchers conceive populations as meaningful sets of cases that often must be formed and conceptualized in the course of investigation [rather than being constituted in advance of data collection and analysis].

[2] They view cases as configurations of aspects and seek to understand them at the level of the specific instance [rather than viewing cases as embodiments of ”variables” that are assumed to be separable in causal impacts].

[3] They see causation as conjunctural and plural – causes may combine in different and sometimes contradictory ways to generate the same outcome [rather than seeing causation as being linear-additive, unitary and stationary over time].”<sup>24</sup>

The diversity-oriented method is well-suited for the study of conflicts and settlements over time. The course of history is determined largely by contingency and therefore generates a diverse plurality of cases and causal conditions with more or less overlapping aspects. If the goal is to construct a rational choice theory of conflict due to legal incompatibilities, one must analyze specific instances of legal incompatibility in their unique historical contexts. Each case is discovered to be a configuration of many aspects, some of which are unique to the case while others are similar across cases. Through a combination of induction from individual cases, comparison across cases, and deduction from weak assumptions about adaptive rationality (MEFs), a population of cases emerges. While each case instantiates the main causal mechanism linking legal incompatibility to the eruption of conflict and the transition from a world of law to a world of power, it is not clear whether this causal mechanism could ever be captured by a statistical regression – at least not with these cases.

There are several reasons for this. First, there are not enough cases (only 23). Second, the collection of cases would be difficult to reduce to a conventional rectangular dataset using the ”variable-oriented discourse” about cases required by statistical methodology. Third, the causal mechanism is conjunctural and plural in nature. It operates through a ”diverse and multifarious” set of causal factors in each case. When compared across cases, the causal factors have enough in common to constitute a single general mechanism – legal incompatibility – but not enough in common to form a rectangular dataset based on ”strong homogenizing assumptions about cases”.

---

<sup>24</sup>Underlining and paragraph breaks added for clarity of comparison. Ragin 2000, p. 14.



## 2.4 A formal comparison of the standard regression method and the diversity-driven method for analyzing conflicts and settlements

As mentioned earlier, actors' understanding of the means and ends available to them evolve over time as a result of conflicts and settlements.<sup>25</sup> This evolutionary process is worth considering in more detail to clarify the extent of the difference between the diversity-oriented approach to drawing causal inferences about conflict and settlement and the conventional regression approach.

Suppose that a specific legal incompatibility causes a dispute and the dispute erupts into a conflict. Later the conflict is terminated by means of a settlement that is designed to resolve the legal incompatibility that caused the conflict. The settlement terms then shape actors' understanding of the means and ends available to them when the next dispute arises. For concreteness, consider some historical examples.

1. During and after the American Civil War, the states and the Congress enacted the 13th, 14th and 15th amendments to the Constitution. These settlements shaped actors' understanding of the means and ends available to them when disputes arose over school desegregation and voting rights in the 1950s and 1960s. In turn, the Voting Rights Acts of the 1960s shaped actors' understanding of the means and ends available to them when the Justice Department and the Supreme Court were deciding whether to continue federal oversight of state voting practices in the 2010s.
2. In 1978 the state of California passed proposition 13 in order to limit increases in property taxes. This settlement shaped actors' understanding of the means and ends available to them in the California state budget crises of the 1990s that led to the recall of Governor Gray Davis.
3. After the United States evicted the Iraqi military from Kuwait in 1991, the southern Iraqi Shiites rose up against the Sunni-led regime in Baghdad and the US faced a policy dilemma over whether to support the uprising. After the US decided not to support it, the Sunni regime suppressed the uprising, and the international settlement between the Sunni regime and the US left the southern Iraqi Shiites disgruntled. This settlement shaped the Shiites' understanding of the means and ends available to them in dealing with the Sunnis a decade later after the United States invaded Iraq in 2003.<sup>26</sup>

These examples point to a general pattern. When each settlement shapes actors' percep-

---

<sup>25</sup>This point was discussed previously in Section X of this chapter.

<sup>26</sup>See Koppel (2000c, 2006a, 2006b) for analysis of five cases of war in the 19th century due to the concatenation of a conflict and a settlement that led to another conflict. In each case Britain went to war with a smaller state outside of Europe because the settlement of one dispute between the two states shaped actors' understanding of the means and ends available to them in a later dispute. This analysis appears in Part VI of the present study.

tions, options and actions in the next dispute, it creates complex linkages between structural factors and contingent factors in the causation of conflict. This complexity confounds the conventional regression framework for representing empirical evidence and analyzing it for causal inference. To demonstrate this point, I now present a formal comparison between the diversity-oriented framework and the conventional regression framework.

Imagine a sequence of conflicts and settlements over time. In each time period  $t$  there is a conflict caused by a legal incompatibility *and* a settlement whose terms are designed to resolve the legal incompatibility. The conflict and the settlement are represented together as event  $E_t$ . This event is caused by a combination of structural factors  $S_t$  and random, historically contingent factors  $R_t$  which include the legal incompatibility. That is, although the legal incompatibility is an "unpredictable antecedent" of conflict, it is a necessary condition of conflict. Without it the conflict would not have occurred. Thus

$$E_t = S_t + R_t \quad (8.1)$$

Since the settlement terms in event  $E_t$  are designed to resolve the legal incompatibility – itself a contingent factor – the settlement terms are determined partly by historical contingency.

Now imagine that the settlement component of event  $E_t$  becomes part of the framework of means and ends that actors perceive is available to them in the next time period  $t + 1$ . Thus the settlement of event  $E_t$  – shaped partly by *contingent* factors – becomes one of the *structural* factors that shape the causation of the next conflict and settlement  $E_{t+1}$ . In this way the *contingent* factors  $R_t$  that helped to cause event  $E_t$  will become part of the subsequent *structural* factors  $S_{t+1}$  that cause the next conflict and settlement  $E_{t+1}$ . As this happens repeatedly over time, more and more of the *random contingent* factors will enter into the *structural* causes of each subsequent conflict and settlement. The table below captures this pattern and compares it to the conventional regression framework. All subscripts in the table refer to the time period  $t$ .

A number of points emerge from this comparison.

1. In the regression framework, the error terms ( $\epsilon_t$ ) are usually interpreted to refer to causal factors that are not measured or even identified, but which affect the outcome

Table 8.1: Conflicts and settlements over time: Comparing the diversity-oriented framework and the conventional regression framework

diversity-oriented framework			
period	event	structural causes	contingent (random) causes
$t$	$E_t$	$S_t$	$R_t$
1	$E_1$	$= S_1$	$+ R_1$
2	$E_2$	$= S_2 + R_1$	$+ R_2$
3	$E_3$	$= S_3 + R_2 + R_1$	$+ R_3$
4	$E_4$	$= S_4 + R_3 + R_2 + R_1$	$+ R_4$
5	$E_5$	$= S_5 + R_4 + R_3 + R_2 + R_1$	$+ R_5$
regression framework			
period	event	structural causes	contingent (random) causes
1	$E_1$	$= \beta_0 + \beta_X X_1 + \beta_Y Y_1 + \beta_Z Z_1$	$+ \epsilon_1$
2	$E_2$	$= \beta_0 + \beta_X X_2 + \beta_Y Y_2 + \beta_Z Z_2 + \beta_E E_1$	$+ \epsilon_2$
3	$E_3$	$= \beta_0 + \beta_X X_3 + \beta_Y Y_3 + \beta_Z Z_3 + \beta_E E_2$	$+ \epsilon_3$
4	$E_4$	$= \beta_0 + \beta_X X_4 + \beta_Y Y_4 + \beta_Z Z_4 + \beta_E E_3$	$+ \epsilon_4$
5	$E_5$	$= \beta_0 + \beta_X X_5 + \beta_Y Y_5 + \beta_Z Z_5 + \beta_E E_4$	$+ \epsilon_5$

event ( $E_t$ ) nonetheless. By contrast, in the diversity-oriented framework, the contingent factors ( $R_t$ ) are identified, measured and included in the logic of the causal mechanism hypothesized in the theory.

2. In the regression framework, the error terms ( $\epsilon_t$ ) are usually interpreted to refer to causal factors that operate in the same way across all time periods.<sup>27</sup> By contrast, in the diversity-oriented framework, the contingent factors ( $R_t$ ) represent legal incompatibilities that cause conflict through any number of different causal channels.
3. In the regression framework, the main explanatory factors have the same causal impact on the outcome event in each time period. The main explanatory factors are the structural variables which enter the regression equation in the same way in each time period ( $X$ ,  $Y$  and  $Z$ ). Their causal impact is captured by the regression coefficients which are assumed to be constant across all time periods ( $\beta_X$ ,  $\beta_Y$ ,  $\beta_Z$ ). By contrast, in the diversity-oriented framework, the main explanatory factors do not have the same

<sup>27</sup>"In the usual time-series setting, the disturbances are assumed to be homoscedastic, but correlated across observations...." **Green90**

impact on the outcome event in each time period. They have "plural and conjunctural" impacts. The main explanatory factors are the legal incompatibilities captured in the random contingent factors ( $R_t$ ). These factors enter the regression equation in a different way in each time period. Their causal impact can only be identified and converted into a theory of causal mechanism through individual case studies followed by comparison across cases.

4. In the regression framework, the main goal of including the lagged dependent variable is to capture the impact of the *structural* causes of previous outcomes on current outcomes, not the *random* causes. The lagged dependent variable ( $E_{t-1}$ ) impacts the current outcome event ( $E_t$ ) mainly through its embodiment of the structural factors in the previous time period ( $X_{t-1}, Y_{t-1}, Z_{t-1}$ ). Although the lagged dependent variable also embodies the error term from the previous time period ( $\epsilon_{t-1}$ ), this constitutes a *problem* that *interferes* with the statistical estimation of the main causal factors in the model ( $\beta_X, \beta_Y, \beta_Z$ ).<sup>28</sup>
5. By contrast, in the diversity-oriented framework, the main goal of including previous settlement designs in the explanation of current conflicts is to capture the impact of the *random* causes of previous events on current events. The designs of previous settlements impact the causation of current conflicts mainly through their embodiment of the contingent factors in previous time periods (e.g. the legal incompatibilities captured by  $R_{t-1}, R_{t-2}, R_{t-3}$ , etc.).
6. In the regression framework, the impact of previous periods on the present period is usually assumed to *diminish* as they become more distant in the past.<sup>29</sup> By contrast, in the diversity-oriented framework, the impact of the past constitutes an *increasing* proportion of the total set of factors that can cause conflict in each period ( $E_t$ ). In other words, the more past events there were, the *more* likely it is that at least one of them will "come back to haunt" the contracting parties in the present period through some kind of legal incompatibility. Moreover, since the impact of the past on the present operates through the random contingent factors – which are so "diverse and multifarious" – the more past there is, the *less* likely it is that history's impact will run through a single, invariant causal mechanism that operates in the same way through history through just a few causal variables (like  $X, Y$  and  $Z$ ).

In sum, the diversity-oriented approach reflects the true nature of history – "an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result"<sup>30</sup> – in a way that the conventional regression approach does not.

---

<sup>28</sup>The problem is that the lagged regressor ( $E_{t-1}$ ) is correlated with the error term that operates in the same way across all time periods ( $\epsilon_t$ ). This biases the estimates of  $\beta_X, \beta_Y$  and  $\beta_Z$ . "If the regression contains any lagged values of the dependent variable, least squares will longer be unbiased or consistent." **Green90**

<sup>29</sup>"[E]ach disturbance,  $\epsilon_t$ , embodies the entire past history of the  $u$ 's, with *the most recent observations receiving greater weight than those in the distant past.*" **Green90**

<sup>30</sup>**Gould89.**

## **2.5 How to strike a balance between scientific objectivity and historical subjectivity: Fuzzy set coding rules**

Seen from this perspective, history produces an incredibly diverse set of cases of conflict and settlement. The main methodology for describing variation across such a diverse set of cases is the *fuzzy set* method.<sup>31</sup> According to this approach, a particular actor, event or case may belong entirely to a particular set, partly to it, or not at all to it. For example, the competing claims in a particular dispute may belong entirely to the set "legal incompatibility", partly to it, or not at all. This study claims, based on empirical evidence, that each of the interstate disputes analyzed here belongs to the set "legal incompatibility" at least partly; and it belongs by enough that the problem of legal incompatibility is what caused the dispute to descend into conflict. This study also claims based on empirical evidence that the domestic disputes which caused domestic conflicts belonged to the set "legal incompatibility" at least partly; and they belonged by enough that the problem of legal incompatibility is what caused the disputes to descend into conflicts.

While these claims can be assessed in detail in the case study chapters, it may be useful to summarize the historical evidence for these claims at this point to establish their plausibility at the outset.

## **2.6 A fuzzy set view of the legal incompatibility disputes that caused conflict in European history**

The tables below summarize the evidence. The first table summarizes the evidence on interstate wars between peer competitors caused by legal incompatibilities. The second table summarizes the evidence on interstate wars between Britain and a much smaller state outside of Europe caused by legal incompatibilities.

---

<sup>31</sup>Ragin 2000.

Table 8.2: Conflicts caused by legal incompatibilities (I): Inter-state wars between peer competitors

case	state 1's right	state 2's right
1 Wars of Dutch Independence (1568-1648)		
2 Thirty Years War (1618-1648)	Kingdom of Bohemia had right to replace domestic government officers who violated their religious rights guaranteed by law	Kingdom of Austria had right to suppress rebels against the authority of its king who also sat on throne of Kingdom of Bohemia
3 Nine Years War (1688-1697)	English Parliament had right to replace sitting king by parliamentary vote	French king had right to rule in France by pan-European law of divine right, and right to defend English king's rule by same law
4 War of the Spanish Succession (1701-1714)	Austria's House of Habsburg had right of succession to Spanish throne on one legal interpretation of inheritance laws	France's House of Bourbon had right of succession to Spanish throne on another legal interpretation of inheritance laws
5 French Revolutionary Wars (1792-1815)	France had right to reform domestic government and monarchy, even if it meant limiting king's authority by parliamentary vote	Austria's king had right to rule in Austria by pan-European law of divine right, and right to defend French king's authority from domestic encroachment by same law
6 World War I (1914-1918)	South Slavs in Boznia, Herzegovina and Serbia had right to be united in one state	Austrian Empire had right to suppress South Slav rebels in its provinces of Boznia and Herzegovina

Table 8.3: Conflicts caused by legal incompatibilities (II): Inter-state wars between Britain and a much smaller state outside of Europe

case	state 1's right	state 2's right
1 First Anglo-Chinese War (1838-1842)	British trading vessels had right to dock at Chinese ports for general trading purposes	Chinese government had right to interdict opium trading conducted through British trading vessels
2 First Anglo-Afghan War (1838-1842)	British government had right to withdraw diplomatic exchanges of commitments as long as it had a valid excuse	Afghan government had right to switch alliance to Russia if British government reneged on its diplomatic commitments
3 Second Anglo-Chinese War (1856-1860)	English government had right to protect Chinese trading vessels from pirates, by agreement with Chinese government	Chinese government had right to board Chinese trading vessels and British protection vessels to interdict opium trading
4 Second Anglo-Afghan War (1878-1881)	British government had right to have a diplomatic envoy in Kabul	Afghan government had right to close Afghan territory to British diplomatic envoys, on grounds that their entry led to war last time around (1838-42)
5 British Invasion of Egypt (1882)	British government had right to control administration in key ministries of Egyptian government to ensure it repaid its debts to British bondholders	Egyptian public had right to prevent Egyptian government from cutting public spending while simultaneously overpaying British working in its ministries

## CHAPTER 9

### Inference strategies for path dependent histories

#### 1. A new theory and a critical test

The main goal of this study is to develop a new theory of the causes of conflict and the conditions for stable settlement in European history and test the theory against a range of historical evidence. The theory differs significantly from previous theories of conflict and settlement in European history. Therefore a related goal of this study is to conduct a critical test between the new theory and the previous theories. The method for conducting the critical test is to characterize the new theory game-theoretically, derive as many of its observable implications as possible, and assess whether the implications are corroborated by the historical evidence. The test is completed by counting the number of corroborated implications generated by the theory relative to the number of corroborated implications generated by the previous theories. All of these tasks are undertaken in Parts II through VII of the study. The value of the new theory is also assessed by identifying the shortcomings of the previous theories. Their logical and empirical shortcomings are identified in the literature review chapters at the beginning of each Part of the study. Their methodological shortcomings are identified in this chapter and the next three chapters.

The overall purpose of this chapter is to clarify the methodology of causal inference employed in this study and explain why it has a comparative advantage over the methodologies employed in previous studies. The discussion is framed in terms of the two "cultures" of causal inference that now dominate the academic discipline of political science. When it comes to the question of research design for causal inference, political scientists are divided into two cultures: the *statistical* culture and the *set-theoretic* culture.<sup>1</sup> The two cultures

---

<sup>1</sup>Goertz and Mahoney 2012



differ profoundly on the main issues of research design for causal inference. Hence the two cultures use different standards to assess whether a causal inference is valid as a matter of science. Consequently members of one culture may regard a causal inference as valid while members of the other culture regard it as invalid. It is essential, therefore, to clarify which culture of research design this study employs and why that culture has a comparative advantage in solving the inference problems faced in this study. Then it will be clear which standards are appropriate to use in assessing the validity of the causal inferences drawn in this study.

## **2. Choosing a research culture based on its comparative advantage for the inference problems to be solved**

In academic political science, the gold standard of causal inference is the combination of a statistical analysis and a game theory model. The function of the statistical analysis is to prove that there is a *correlation* between the supposed cause and the observed effect after controlling for potentially confounding variables. The function of the game theory model is to show how variations in the causal variable actually *cause* rational actors to change their behavior in the ways measured by the effect variable. This proves that the statistical analysis captured more than simply a statistical correlation: It captured a *causal* relationship between the hypothesized cause and the observed effect.

Under ideal conditions this approach yields valid causal inferences. Hence this has become the conventional gold standard in the field. The ideal conditions required by this approach are not always satisfied though. Many advances in statistical methodology are intended to correct for departures from the ideal conditions. Some departures are so serious that it is difficult or impossible to correct for them, however. In these situations the alternative methods offered by the set-theoretic culture of research design may have a comparative advantage. Indeed, in some situations conventional statistical analysis may be completely unable to detect a causal relationship *even when it really exists*.<sup>2</sup> In such cases the set-

---

<sup>2</sup>See Goertz and Mahoney (2012). They reprint Geddes's regression analysis showing that there is no *statistical* relationship between labor repression and economic growth. Then they show that there is a *set-theoretic* relationship between the two variables.

theoretic approach may be *required* to produce valid causal inferences. Each of the following subsections considers a serious challenge to conventional statistical inference and discusses how that challenge can be met through set-theoretic inference.

## 2.1 "Percentage of the variation explained": Two views of the problem of measuring variation on the dependent variable for explanatory purposes

In most statistical studies of the causes of war, the dataset includes a large number of cases and the study claims to yield generalizations about the causes of war.<sup>3</sup> The study produces generalizations by assuming that the  $n$  cases in the dataset are *representative* instances of a single causal mechanism that drives war initiation. In such studies "war" is defined as an event in which military hostilities erupted and at least 1000 casualties occurred. Assuming this definition of war is valid, these studies do indeed produce generalizations about what causes war. While it seems obvious to define war in this way, however, this approach raises questions about what it really means to "explain variation" in a study of war.

Consider an illustrative example. Suppose that a dataset contains 200 cases. A war erupted in 100 of the cases; but no war erupted in the other 100 cases. Suppose that a statistical regression shows that 90 of the 100 cases of war are explained by factor  $X$ . Then one could say that factor  $X$  explains 90% of the variation. Now suppose that in each of these 90 cases there were exactly 1000 battle deaths. But in each of the 10 cases that are *not* explained by factor  $X$ , there were 1 million battle deaths. So the total number of battle deaths in all 100 cases of war was 10,090,000. While factor  $X$  explains 90,000 of these deaths, factor  $X$  leaves 10 million of these deaths unexplained. From this perspective, it is no longer possible to say that factor  $X$  explains 90% of the variation. Rather it explains less than 1% of the variation.

Let us put the same point another way in order to highlight the question of what it means to explain variation in a study of war. The 90 cases explained by factor  $X$  represent 90% of the variation from "peaceful" state behavior to "warring" state behavior. Yet they represent less than 1% of the variation from "not imposing war costs on the other state" to "imposing war costs on the other state". So the question is: Which type of variation needs

---

<sup>3</sup>cites.

to be explained in order to understand what causes states to engage in war behavior? Is it sufficient to explain what causes two states to *get into* a war with each other? Or is it necessary to explain what causes two states to *impose war costs* on each other? If an answer to first question is sufficient, then it is okay *not* to distinguish in the causal analysis between a war that only caused 1000 battle deaths and a war that caused 1 million battle deaths. But if an answer to the second question is necessary, then this distinction is required. That is, the more costly wars need to be given more weight in the causal analysis than the less costly wars.

The conclusion from this example is obvious. An explanation of the costliest wars tells us more about why states decide to impose costs on each other through war than an explanation of the least costly wars. Therefore, if we define war as "two states imposing costs on each other through military means", then a theory that explains the costliest wars tells us more about what causes war in general than a theory of cheap wars. To put the point in terms of the example offered above, a theory that explains those 10 wars – each of which caused 1 million battle deaths – tells us more about what causes war in general than a theory that only explains the other 90 wars – each of which only caused 1,000 battle deaths. For this reason this study focuses on explaining what caused the costliest wars in European history.

Since the number of these wars is small – less than ten – statistical inference is not possible. For inference problems involving a small number of cases, in-depth case studies may offer a comparative advantage over statistical analysis in producing causal explanations.<sup>4</sup> Hence this study employs in-depth case studies. The most rigorous method for drawing inferences from case studies is the set-theoretic method. According to this method, the theorist proposes a set of causal conditions that are individually necessary and together sufficient to cause the outcome in question. To yield a valid causal inference, the theorist must provide a collection of empirical evidence which proves that (1) when the causal conditions were present, the outcome occurred, but (2) when the causal conditions were absent, the outcome did not occur. Additional evidence of causation can be secured by deriving as many observable implications of the causal hypothesis as possible and locating evidence that corroborates the implications. This study employs the set-theoretic method.

---

<sup>4</sup>cite.

## 2.2 "Proving the operation of a causal mechanism in the data": Two views of the problem of causal heterogeneity

As mentioned, the gold standard in causal inference is a statistical analysis supported by a game theory model. This method of inference assumes *causal homogeneity*: A single causal mechanism explains all of the cases in the dataset that are relatively close to the regression line. And that is the causal mechanism delineated in the game theory model. Often theorists simply assume causal homogeneity even though closer examination of individual cases in the dataset would reveal that a variety of causal mechanisms was at work, not a single one, thereby refuting the assumption of causal homogeneity. If there is causal *heterogeneity* in the dataset, however, then it invalidates the inference that the single mechanism in the game theory model explains the observed outcomes.

Many statistical studies analyze the role of power in militarized disputes that escalate into war. These studies assume that all disputes have the same power properties and escalation occurs by the same causal mechanism in all cases. Yet the case studies to be presented later in this study show that these assumptions are not true. In fact disputes differ in their power properties and, as a result, the causal mechanism of escalation differs from one dispute to another. These conclusions suggest that the strong homogeneity assumptions made in statistical analyses of crisis escalation and conflict initiation are probably not true in the set of cases analyzed.

This naturally raises a question. When such regression studies produce statistically significant coefficients, how are these coefficients to be interpreted? Internal critics in the statistical community acknowledge that such "mongrel" regressions are entirely uninformative about causation – despite their significant coefficients. External critics note the problem of "equifinality". When regressions are performed on a data set that contains heterogeneous causal mechanisms, the significant coefficients say nothing about which causal mechanism operated in which cases, and therefore leave open the possibility that the single causal mechanism identified in the game theory model is not the one that operated in many cases – or even most of the cases – in the dataset.<sup>5</sup> Worse, omitted variable bias can generate a statistically significant coefficient on a variable that in fact played no role in *any* of the cases

---

<sup>5</sup>Goertz and Mahoney 2012.

in the dataset. Of course it is up to the critic to identify the omitted variable and show that it played a causal role. This is one of the main goals of the present study.

Methodologists propose a simple solution for the problems of causal heterogeneity, equifinality and mongrel regressions. The solution is to derive additional observable implications of the causal hypothesis, especially implications that can discriminate between that hypothesis and competing hypotheses. The more implications a researcher derives and tests on the evidence, the more opportunities there are to find either supporting or refuting evidence for the theory. Hence methodologists advise researchers to derive as many observable implications as possible from each theory, especially when testing rival theories.

To understand the logic behind this solution, consider an analogy to the statistical model called "seemingly unrelated regressions". Consider a set of  $n$  regression equations that share the *same* independent variable but have *different* dependent variables. Imagine that the  $n$  equations represent  $n$  coins manufactured by the same foundry.<sup>6</sup> Consider two competing hypotheses:

$H_0$ : The foundry only manufactures fair coins.

$H_1$ : The foundry manufactures biased coins that are weighted toward heads.

Suppose that each of the  $n$  coins is tossed one time, so the number of cases in each of the  $n$  equations is one. If the foundry manufactures only fair coins, then what is the probability of obtaining all  $n$  heads from the  $n$  tosses? The probability is  $p = .5^n$ . Thus if  $n = 5$  then  $p = .03$ .

This model provides a direct analogy to the method of deriving as many observable implications as possible from a causal hypothesis. Each observable implication represents a different dependent variable  $1, \dots, n$ . Each of these dependent variables represents an event that occurs with probability  $p_i$  in the *absence* of the hypothesized cause (where  $i = 1, \dots, n$ ). Thus, if the hypothesized cause had *not* in fact played any role, then the probability of observing all  $n$  events in the historical record is  $\prod_{i=1}^n p_i$ . Notice that as  $n$  increases,  $\prod_{i=1}^n p_i$  approaches zero asymptotically. That is, the larger  $n$  is, the *lower* is the likelihood of

---

<sup>6</sup>Here  $n$  refers to the equation number, not the case number within each equation.

observing all  $n$  events in the historical record if the hypothesized cause had *not* played any role. Hence if all  $n$  implications are indeed born out by the evidence, then it lends support to the theory in this probabilistic sense: Such a pattern of evidence would be very unlikely to be observed in the historical record if the theory were *wrong*. Thus, the more observable implications one derives and corroborates with empirical evidence, the more support it lends to the theory (for the lower is the likelihood that all  $n$  events would be observed in the historical record if the theory were wrong).

This analogy suggests a probability foundation for conducting a critical test between two rival theories. If one theory only generates a *few* implications that are corroborated by the evidence, while the other theory generates *many* implications that are corroborated by the evidence, then the second theory is less likely to be wrong than the first theory. One of the main goals of this study, therefore, is to derive as many observable implications as possible from the main causal hypotheses. A related goal is to show that previous theories generated fewer observable implications, and hence the theory presented here offers a more probable explanation of the patterns of evidence observed in the historical record.

### **2.3 "Proving the rational causes of observed behavior": Two views of the problem of rational over-determination**

By its nature, rational choice theory leads to an optimistic view of the explanatory power of models: If a model provides *an* explanation of some observed behavior, then it provides *the* explanation. The problem with this view is that a single observed behavior might be driven by any *one* of a number of different motivations – and not necessarily the one in the model. In other words, simply providing a model of one motivation does not guarantee that it was *the* motivation. The behavior might have been motivated by any of the other motivations instead. Consider two examples from the field of economics.

Before the rise of game theory, unusual behavior was explained as a rational consequence of price discrimination incentives. Yet however rigorous the models of price discrimination were, they have since been superseded by different explanations of the *same* behavior.

It used to be that an economist's generic answer to someone who brought up

peculiar behavior that seemed to contradict basic theory was "It must be some kind of price discrimination." Today, we have a new answer: it must be some kind of asymmetric information.<sup>7</sup>

Before the rise of transaction cost economics, vertical restraints on market exchange were explained as a rational consequence of incentives to engage in anti-competitive practices such as collusion, market foreclosure or exclusive dealing. Yet however rigorous those models were, they have since been superseded by different explanations of the *same* behavior (see see Chapter 7).

An economic methodologist stated the problem of rational over-determination in this way:

"[There is a] connection between functionalism and the optimization-and-equilibrium approach in economics. To say that a structure has a function is to say that it solves some kind of problem for the group in question – a problem usually cast in terms of the selection mechanism thought to be operating. The structure is functional because it solves a problem linked to the group's relative success or survival; the structure is efficient in some sense.... [T]his functionalist problem is easily recast in the form of an operations-research problem [in which each agent solves an optimization problem in order to maximize his utility or minimize his cost]. The substitution of an optimization problem for the functionalist problem...is an "as if" exercise: it is as if the agents were attempting to solve a certain (global) problem (whose optimal solution becomes a normative standard). The agents may in turn be represented as if they were solving their own little pieces of the global problem (with any inability to solve this superimposed global problem labeled a "market failure").... That is to say, we can easily transform the functionalist explanation into an intentional explanation: It is as if the agents possessed certain information and consciously brought about certain outcomes in view of that information.

It is, of course, a quick jump from saying that something is functional to saying

---

<sup>7</sup>Rasmusen 1989, p. 133.

that it is optimal, especially when we get to choose (within limits) the super-imposed operations-research problem that we think best captures the functional problem. This is not an incidental matter, since there is often enough leeway that we can find, for any given structure, a corresponding operations-research problem to which the structure is an optimal solution.... Social institutions may be susceptible to invisible-hand explanations employing any number of selection or filtering processes, and may also involve elements of intention. This make it a much trickier business to find *the* problem that the structure solves and, not incidentally, makes it easier to find *a* problem to which the structure is an optimal solution.”<sup>8</sup>

This problem is difficult to detect when evaluating a model that has a rational choice foundation and empirical data that support it. For the model provides a rational link between the cause and the effect and the empirical data show a correlation between the causal variable and the effect variable. Thus there seems to be no reason to look any further for any different explanation of the empirical correlation. Yet in fact, the model may *not* be the true explanation of the observed empirical correlation, due to the problem of rational overdetermination.

The solution to this problem is to conduct critical tests between rival models of the same phenomenon. One derives as many observable implications of each model as possible – until one finds implications of one model that are different from the implications of the other model. Then the empirical data on the rival implications discriminate between the models. The solution is similar to the method of differential diagnosis that a doctor uses to diagnose a patient. Suppose the doctor has two rival hypotheses for what is causing the patient’s symptoms. The doctor applies a non-invasive treatment that would produce one observable effect on the patient if the first hypothesis is correct but a different observable effect if the second hypothesis is correct. Then the patient’s actual response to the treatment will discriminate between the two hypotheses. Similarly in social science, if two rival models generate different implications for their observable effects in the real world, then real data can provide a critical test between the rival models.

---

<sup>8</sup>Parentheses in original. Italics added. Langlois 1986a, pp. 249–50.



### 3. The challenge of producing generalizations about the causes of conflict

In the existing literature there are two approaches to the goal of producing generalizations about the cause of war: the statistical approach and the case study approach. The statistical approach assumes that a generalized causal inference can only be valid if it is supported by a statistical result from a large- $n$  data set. While case studies might be able to *illustrate* a causal mechanism in operation in the cases that were studied, such analysis can never *prove* the operation of that causal mechanism beyond those particular cases without a large- $n$  statistical result to back it up. By contrast, the case study approach assumes that at least some of the driving forces of politics are timeless: They operate in more or less the same way in most places at most times. Therefore conclusions drawn from case studies can be presumed to support generalized statements about "how states behave" and "how politics works". This section examines the obstacles to generalization encountered by each approach and then suggests a new way to produce generalizations about the causes of conflict.

#### 3.1 The statistical method for generalizing about the causes of conflict

There are many statistical studies that claim to produce generalizations about the causes of conflict. These generalizations are based on a number of homogeneity and stationarity assumptions. The assumptions may go unquestioned for many years, leaving the impression that the generalization is valid in the meanwhile. For example, a number of statistical studies concluded that democratizing states are especially war-prone. Years later, another statistical study discovered that this generalization was unwarranted by the data. It was driven largely by a few cases in which a new democracy initiated war, while many *other* new democracies were *not* especially war prone (Narang and Nelson 2009). This example shows that even when a study produces a significant regression coefficient, this does not necessarily mean that a *single, general* causal mechanism drove the majority of the cases in the dataset. In other words, statistical studies may not yield the *generalizations* about the causes of conflict that they claim to yield. How many other statistical studies of conflict initiation suffer from this problem of unwarranted generalization – yet have not been discovered to date?

Once a dataset with a large number of cases has been created, there is a tendency to assume that it represents a kind of uniformity across the cases – that they are *all cases* of a particular phenomenon. This presumption is fostered by the "coding rules" used to determine what constitutes a "case". Once the coding rules have been followed and the dataset has been created, its cases are presumed to constitute a *uniform* representation of that type of case. Yet the coding rules may mask significant heterogeneities across cases that render them *non-uniform*. In this situation, the coding rules have functioned, in effect, as a kind of "shoe-horning" device – enabling the coders to "shoe-horn" heterogeneous cases into a single box. While this creates an impression of uniformity across cases, it is a misimpression, an illusion. And "generalizations" derived from such a dataset are equally an illusion. In fact, heterogeneities across the cases can invalidate causal inference. For example, based on an assumption of uniformity across cases, the studies cited above drew the inference that new democracies are especially war-prone. Yet in fact, there was heterogeneity across cases that invalidated this inference.

This problem calls into question the conventional coding rules for defining conflict for the purpose of drawing causal inferences about the causes of conflict. If conventional coding rules mask heterogeneity that invalidates causal inference, then how else can generalizations about the causes of conflict be achieved?

### **3.2 The logical method for generalizing about the causes of conflict**

This method proposes that the causes of conflict fall into a few main categories and most conflicts in human history were caused by one of these factors.

- Biological causes
  - the territorial imperative (Ardrey)
  - human nature (Thayer)
- Psychological causes
  - misperceptions (Jervis)
  - emotions (such as honor and dishonor) [Oneill]
- Domestic political causes

- diversionary motives (Morgan 1992)
- log-rolling (Snyder 1991)
- in-group out-group frictions
- Rational causes
  - indivisibility of the stakes in a dispute (Toft 2003, Hensel 2005)
  - commitment problems related to relative power (Powell 2002)
  - incentives to bluff about power (e.g. to bluff strength when weak) [Fearon 1995]
  - breakdown of common knowledge about relative power (Reiter et al 2000? 2004?)

Within the category of rational causes, the conventional wisdom is that any rationalist explanation of conflict must fall into one of the first three subcategories listed above (Fearon 1995). In addition, Fearon claims that the first two subcategories are relatively infrequent causes of war, and hence the third subcategory – bluffing about power – is the main rational cause of war.

According to the conventional wisdom, then, the incentive to bluff about power provides a *general* explanation of war. This explanation is general in two senses. First, the explanation has been characterized in game theoretic form in terms of the necessary and sufficient conditions for bluffing about power to cause war. Because these conditions are very general in nature, they are assumed to explain a wide range of conflicts and wars. Second, the explanation has been embedded in regression studies that produced statistically significant coefficients. Since each study included a large number of cases, the studies together are assumed to provide proof that this factor – the incentive to bluff about power – caused war in a wide range of cases. Given this factor’s assumed generality, it has become the default hypothesis for researchers conducting any regression analysis of conflict. For example, studies of local land disputes and violence between neighboring farmers have relied on this hypothesis by default.<sup>9</sup>

---

<sup>9</sup>Blattman et al. (2015b) cite Fearon’s (1995) model of conflict due to asymmetric information *about relative power*:

”Broadly, conflict is a breakdown of peaceful bargaining, and is more likely with information asymmetries and absence of commitment mechanisms. Strong, well-functioning social and political institutions at the town level reduce information asymmetries and improve coordination and enforcement (Fearon, 1995; Blattman et al., 2014).”

Within the field of international relations theory, the conventional wisdom is now that most conflicts are not caused by shifts in relative power, but by asymmetric information or commitment problems. Yet upon a closer look, it turns out that scholars are really saying: asymmetric information *about relative power* and commitment problems *related to relative power*.

”Realists argue that balancing occurs in response to changes to the balance of power. Recent informational approaches have focused primarily on informational asymmetries or commitment problems.”<sup>10</sup>

The authors then cite Fearon’s (1995) model of conflict due to asymmetric information *about relative power*, and Powell’s (2006) model of conflict due to commitment problems *related to relative power*.

Thus the conventional wisdom is that war is generally caused by relative power factors (e.g. bluffing about power or commitment problems related to relative power). If this claim were true, then it would indeed constitute a generalization about the rational causes of conflict. However, there are both logical and empirical problems with this claim. On the logical side, the rational causes of war listed above are not the only conceivable rational causes of war. In the present study I characterize a different rational cause of war (and conflict generally) and demonstrate the operation of this cause throughout European history both within and among states.

On the empirical side, the claim that war is generally caused by bluffing about power or commitment problems related to relative power suffers from four problems. First, the statistical studies intended to provide general, empirical support for this claim are flawed. They suffer from the same inference fallacies described earlier: causal heterogeneity, rational over-determination, and a skewed measurement of variation on the dependent variable. Second, these studies assume that bluffing about power or commitment problems related to relative power caused most of the conflicts in the dataset, yet provide no *within case*

---

Similarly, Blattman et al. (2014) explain conflict over land disputes between neighboring farmers by assuming that conflict results from a breakdown in bargaining due to asymmetric information *about relative power*. They cite Fearon’s (1998) other model which defines relative power as one side’s cost of bargaining delays relative to the other side’s cost of bargaining delays.

<sup>10</sup>Savic and Shirkey 2009.

evidence to demonstrate the operation of this causal mechanism in specific cases. Third, in some instances the regression analysis is accompanied by one or two case studies. Yet, as research methodologists have noted, the purpose of such case studies is simply to *illustrate* the main independent and dependent variables used in the regression analysis, not to *prove* the operation of a causal mechanism on the basis of *within case* evidence. In other words, the main proof of causation is still assumed to reside in the regression analysis (Goertz and Mahoney 2012). The purpose, design and execution of such an "illustrative" case study is completely different from the purpose, design and execution of a "dispositive" case study intended to prove the operation of a causal mechanism using *within case* evidence. This leads to the fourth and most fundamental problem with the claim that war is generally caused by bluffing about power (or one of the other rationalist hypotheses listed above).

These hypotheses generate many observable implications for the expected patterns of evidence in the most costly wars in European history. Although previous studies derived and tested a few of these implications, I derive and test many additional implications of these hypotheses and find that *the implications are refuted by the evidence*. Insofar as these wars provide the large majority of the variation from peace to costly war (as this variation was defined above), any hypothesis that fails to explain these wars cannot be regarded as a *general* explanation of costly conflict.

### 3.3 A new generalization about the causes of conflict

This study develops and tests a new rational explanation of conflict.<sup>11</sup> This explanation focuses on the problem of legal incompatibilities.<sup>12</sup> A legal incompatibility is a situation where one contract term gives one actor the legal right to an asset, but another contract term gives another actor the legal right to the *same* asset. The incompatibility causes a breakdown in common knowledge about the legal meaning of the contract terms. This is

---

<sup>11</sup>It is new in a few senses. It is not among the four rationalist causes listed above (which the conventional wisdom regards as the only rationalist causes of conflict). A literature search on the Social Sciences Citation Index produced only one published article whose theme is directly related to my new causal hypothesis – the review article by Farrell and Newman (2016) – and that review does not mention war as a dependent variable in any of the studies it surveys. See the next paragraph for a summary of the existing literature that is tangentially related to my new causal hypothesis.

<sup>12</sup>I first analyzed legal incompatibilities as a cause of war in my earliest studies of this dissertation topic between 1997 and 2006. I provide the relevant quotations from these studies below in Section 4.3.

not a breakdown in common knowledge about relative power. It is a breakdown in common knowledge about a non-power factor: the legal meaning of contract terms. This breakdown can occur even when none of the four rationalist causes listed above is present (e.g. even when there are no indivisibilities, no commitment problems relating to relative power, no bluffing about power, and no lapse in common knowledge about relative power).

This cause of conflict is tangentially related to other factors that scholars have already identified as causes of inefficiency, dispute or conflict. In the field of mathematical economics, scholars have identified "imperfect common knowledge" as a cause of inefficiency in the macro-economy.<sup>13</sup> In the field of law and economics, scholars have identified the problem of "overlapping contract terms" as a cause of contract disputes.<sup>14</sup> In the field of political science, scholars have identified "disputes about the meaning of words" as a cause of conflict.<sup>15</sup> In the field of international relations theory, scholars have identified "contradictions in society",<sup>16</sup> "contentious issues"<sup>17</sup> and "indivisible stakes"<sup>18</sup> as causes of conflict. In the field of international political economy, scholars have identified "rule overlap between different national jurisdictions" as a cause of international frictions.<sup>19</sup>

I propose a model that locates the source of conflict in a particular type of contradiction in society that causes a particular type of dispute over words: incompatibilities between the existing laws that create problems of overlapping legal rights which, in turn, generate disputes over the meaning of contract terms. According to the existing laws, one term gives one actor the legal right to an asset, while another term gives another actor the legal right to the *same* asset.

In the following section I explain the methods that this study uses to test the hypothesis that legal incompatibility is a *general* cause of conflict.

---

<sup>13</sup>Amato and Shin 2006.

<sup>14</sup>Hermalin, A. Katz, and Craswell 2007.

<sup>15</sup>CITE APSR article on Hobbes' Leviathan.

<sup>16</sup>Wendt 1999.

<sup>17</sup>Aminzade *et al* 2001

<sup>18</sup>Goddard 2006; Hassner 2003; Hensel and Mitchell 2005.

<sup>19</sup>Farrell and Newman (2016). Their review of the literature on this topic does not mention any studies in which the dependent variable is war. (Their review does not even contain the word "war".)

## 4. Methodological goals of this study

### 4.1 Distinguishing between the theory's hard-core assumptions and its protective belt assumptions and implications

I identify the theory's hard-core and protective belt in chapters 10-12. This distinction is necessary for several reasons. First, the theory's hard-core assumptions must be identified to establish that the theory is a *rationalist* explanation of conflict and settlement. This distinguishes the theory from alternative rival theories based on non-rational behavior. Such theories are not compelling because they fail to explain why the non-rational behavior wasn't weeded out by competitive pressures from (more) rational actors. Second, the theory's hard-core assumptions must be identified and characterized as rationalist in order to derive as many observable implications from them as possible. The assumption of rationality is what enables the derivation of so many implications. Third, it is necessary to distinguish the hard-core assumptions from the protective belt assumptions and implications in order to perform critical tests without the risk of false negatives. This distinction clarifies which anomalies are fatal to the theory and which anomalies are not. If the observable implications of the *hard-core assumptions* are refuted by the empirical evidence, that would be fatal to the theory. But if the observable implications of certain *protective belt assumptions* are refuted by the evidence, that would only be fatal to that sub-model built on those protective belt assumptions.<sup>20</sup>

### 4.2 Expanding the number of observable implications of the theory

The primary method for testing this hypothesis is to derive as many observable implications as possible from the hypothesis. Research methodologists recommend this approach for several reasons. First, it enables valid causal inferences in situations where statistical inferences are not possible. Second, it maximizes the probability of producing a valid causal inference because it maximizes the opportunities for validating an inference (e.g. when an implication is corroborated by the evidence) and for invalidating an inference (when an implication is refuted by the evidence). Third, it enables critical tests between alternative rival explana-

---

<sup>20</sup>Lakatos 1970. See chapters 3-6 for further analysis of this issue.

tions. The explanation with the highest number of corroborated observable implications is the most likely explanation, other things being equal.

In using this method I follow in the tradition established by one of the first modern economists to study coercion, conflict and settlement. Hirschman (1945) developed a simple economic model to explain German foreign economic policy toward Eastern Europe in the 1930s and 40s. His methodology was to derive as many observable implications as possible from the model and test the implications through an austere case study that provided only one or two sentences of historical evidence to corroborate each implication. I share his methodology of deriving as many observable implications as possible from a single economic model. I expand on his methodology by deriving implications that explain variation across many cases as well as variation within each case.<sup>21</sup> In using economic assumptions to explain conflict and settlement, I follow in the tradition established by several economists of Hirschman's period (Lane 1942; Lane 1950; Lane 1958; Knorr 1947; Knorr 1957; Demsetz 1967). More recently, political economists have coined the term "analytic narrative" to describe the method of writing an economic or game-theoretic model, deriving its observable implications, and testing the implications through narrative case studies.<sup>22</sup>

---

<sup>21</sup>See King, Keohane, and Verba (1994) for the importance of explaining variation both across cases and within cases.

<sup>22</sup>See Bates (1998). I first wrote that I would use the methodology of game-theoretic analytic narrative for this study in Koppel (2002b). There I wrote:

"The methodology for discriminating between the power-shift theory and the loss-avoidance theory is to construct a theoretical framework in which the two theories are derived from a single set of primitive assumptions. The theories then generate contrasting empirical implications. These implications will be evaluated against the historical evidence; and the theory that explains more of the facts more accurately is accepted. This method of testing is recommended by King, Keohane, and Verba (1994) for inference problems involving a small number of cases without hard data. By deriving as many empirical implications as possible for each case, one can expand the opportunities for a conclusive test. To further increase explanatory leverage, the methodology of analytic narrative is employed (Bates 1998). This involves grounding rich case studies of complex events in the rigorous logic of cause and effect provided by a game-theoretic model. The function of the model is to distill the root cause of each event from the wealth of factors that played intervening or subsidiary roles between that cause and its many effects. The model's assumptions are few and weak (e.g. empirically accurate), while its implications are many and broad. Hence a close match between its implications and the evidence strongly supports the theory in the absence of another theory with equally weak assumptions and equally many corroborated implications." (Koppel 2002b, p. 6)



### 4.3 Widening the range of conflicts explained by the theory

A complementary method is to test the theory on a variety of different kinds of conflict and settlement. The more varieties are explained by the theory, the more valid is the claim that the theory provides a *general* explanation of conflict and settlement.

The study examines both international and domestic conflicts. Among international conflicts, the study examines two kinds. First, the study examines seven cases of symmetric military conflict, e.g. conflict between peer competitors (powerful states of roughly equal strength). These "hegemonic" wars were seven of the nine costliest wars in European history. Hence they constitute the vast majority of the variation between peace and costly war among peer competitors. Second, the study examines five cases of asymmetric military conflict, e.g. conflict between a powerful state and a much weaker state. These "imperial" wars were five of the seven costliest wars in British imperial history. Hence they constitute the vast majority of the variation between peace and costly asymmetric war in that state's imperial history. By explaining both symmetric and asymmetric conflicts, the model identifies a deeper common factor that causes conflict independent of the relative power of the belligerents.

The study also examines a wide range of domestic conflicts in the main states of western and central Europe between 1000 and 1900. These include:

- political, economic and social conflicts;
- militarized and non-militarized conflicts;
- symmetric conflicts across groups of roughly equal power in society;
- asymmetric conflicts between strong and weak groups in society; and
- asymmetric conflicts between a strong government and a weak society.

In explaining such a wide range of domestic conflicts, the model identifies a deeper common factor that causes conflict independent of the domestic issue in contention, the relative power of the contenders, or their means of imposing costs on each other.

The method of wide testing is intended to bolster the claim that the model provides a *general* explanation of conflict. To make this clear and convincing, the wide testing process must meet two standards. First, it is necessary to derive enough observable implications for each kind of conflict to establish the causal mechanism that drove that kind of conflict.

Second, it is necessary to show that the causal mechanism driving each kind of conflict bears enough similarities to the causal mechanism driving the other kinds to warrant calling it the same mechanism.

Part III

**Contractual Realism: Causal links  
between power and law**

## CHAPTER 10

### The world the power, conflict and high-cost deterrence

In the world of power, actors divide goods among themselves according to relative power. The clearest example is a conflict. Each actor imposes costs on the opposing actor to force him to relinquish a share of the good at stake. In the end each actor's share is proportional to his *power to impose costs* on the other actor. Their *relative* shares are determined by their relative power to impose costs on each other. The front line in a war provides the simplest illustration. If one army grows in size relative to the other, it gains territory at the other's expense.

When actors are in the world of power, they divide goods according to relative power not only in war but in peacetime as well. A good example is when two armies stop fighting – stop imposing costs on each other – and strike a peace agreement. The agreement specifies each side's share of the good at stake. Each side will be deterred from seizing some of the other side's share unilaterally – violating the agreement – if it expects to incur a punishment greater than its gain from such opportunism. Opportunism will be deterred as long as each side maintains the capacity to impose such punishment in response to violations. If one side reduces its capacity to impose punishment, however, then the other side will be able to commit violations with impunity – effectively shifting the terms of the agreement in its favor *without conflict*. Thus, even in peacetime, the terms are determined by the actors' relative power: their relative capacities to impose punishment costs on each other.

It seems obvious that each side needs to maintain the capacity to impose punishment on the other side for committing violations. Yet through European history, there was variation in the amount of enforcement capacity that rulers and factions maintained to deter rivals from violating their rights. And counter-intuitively, it was sometimes the actors who maintained *less* enforcement capacity that experienced *more* security for their goods. This is a puzzle worth exploring.

## 1. Variations in enforcement capacity

In medieval Germany, the nobility maintained high levels of enforcement capacity. They retained large numbers of knights who invested heavily in military training and hardware. And they maintained these high capacity levels after active conflicts ended. In medieval England and France, by contrast, the nobility gradually reduced their enforcement capacity. They demanded less knight service from their noble tenants, who in turn invested less money in military training and hardware and more money in agricultural production. And they returned to these low capacity levels after active conflicts ended.<sup>1</sup>

In the early modern period, this pattern was reversed. In France the wealthiest nobles maintained large patronage networks to ensure they would have armies of clients to wage their political and administrative turf wars. And they maintained these high capacity levels after active turf wars ended. In Prussia, by contrast, the magnates gradually reduced the size of their patronage networks. And they maintained these low capacity levels during and after their turf wars.

Variations in enforcement capacity were also evident in international politics. The largest increases in capacity were caused by hegemonic wars. Once each war erupted, states made heavy investments in the capacity to wage war to punish their opponents for violating their rights and deter further violations. The extent of the reduction in enforcement capacity after each war varied. After the Peace of Augsburg in 1555, Germany's territorial rulers continued to make large investments in enforcement capacity in order to defend the boundary lines between their rival religions. After the Peace of Westphalia in 1648, by contrast, rulers reduced their investments in enforcement capacity to defend religious boundaries. After the Peace of Utrecht in 1713, France and Britain continued to invest in the capacity to defend themselves from the threat of military invasions aimed at imposing regime change. After the Peace of Paris in 1815, by contrast, they reduced their investments in the deterrence of such invasions. After the Treaty of Versailles in 1918, the great powers continued to make large investments in economic spheres of influence to maintain their capacities to raise tariffs on each other's exports (in punishment for previous tariff hikes). After the peace treaties

---

<sup>1</sup>Citations for all of the evidence in this section are provided in chapter 30.

of 1945, by contrast, they reduced their investment in this means of deterring and waging economic conflict.

In sum, actors were able to reduce their levels of enforcement capacity after active conflicts in *some* regions and periods of European history, but not others.

## **2. The benefits and costs of varying enforcement capacity**

The main benefit from maintaining enforcement assets is that they provide the capacity to impose punishment on opportunists who commit holdups and thereby deter further holdups. The main cost of maintaining enforcement assets is the standard opportunity cost of withdrawing resources from the productive economy and maintaining them in the enforcement sector for the duration of the enforcement action. This is the income forfeited by taking the resources out of economic production for that length of time. Given this opportunity cost, it would be preferable to keep the resources in the productive economy until an opportunist commits a holdup, and then transfer the resources to the enforcement sector for the time it takes to impose punishment on the opportunist. After the enforcement conflict is over and deterrence has been restored, the resources would be returned to the productive economy and kept there until the next opportunist commits a holdup. This strategy would increase income by keeping resources in productive use in the economy when they are not needed for enforcement purposes. Many actors who value income do not employ this strategy, however. This is a genuine puzzle.

One might try to resolve the puzzle by assuming that when actors are competing for relative power, they need to maintain resources in the enforcement sector at all times, even in peacetime. Even when actors are forced to compete for power, however, they still need to maximize income under this competition constraint to sustain their power in the long run. The strategy of returning resources to the productive economy when not needed for enforcement purposes would help to maximize income. So why don't actors use this strategy? The answer lies in the nature of enforcement technology.

## 2.1 Generic *versus* specialized enforcement capacity

If enforcement actions do not require any specialized technology, then the strategy of returning resources to the productive economy when not needed for enforcement purposes is perfectly feasible. Generic economic assets can be used for enforcement purposes and then returned to economic production without having to undergo any transformation. (In technical terms, it is *costless to transform* resources from their productive use in the economy to their power-based use in the enforcement sector.) Under these conditions, the only cost of waging an enforcement conflict is the standard opportunity cost of withdrawing resources from the productive economy and maintaining them in the enforcement sector for the duration of the conflict. This opportunity cost is *always* less than the potential loss from allowing the opportunist to continue to commit holdups. So the victim of a holdup *always* has an incentive to wage an enforcement conflict. More precisely, the victim always has an incentive to transfer resources from the productive sector to the enforcement sector to punish each opportunist for each holdup. So the potential victim can maintain *all* of his resources in the productive sector at *other* times, knowing that he will always have an incentive to transfer them into the enforcement sector in response to a holdup. This incentive alone is sufficient to deter *all* holdups. That is, the *threat* of punishment alone is sufficient to deter all holdups. None occur. No active conflicts are needed to deter holdups. So the potential victims can maintain all of their resources in the productive economy at *all* times.<sup>2</sup>

This deterrence strategy depends critically on the assumption that the enforcement technology is generic, so it is costless to transform resources from their productive use in the economy to their power-based use in the enforcement sector. By contrast, if the enforcement technology is specialized, then it is *costly* to transform resources from productive to power-based form. Then this strategy is not feasible. The cost of transforming the resources to wage an enforcement conflict would be larger than the losses from simply letting the opportunist continue to commit holdups. It would be cheaper for the victim to permit the holdups than to pay the resource transformation cost necessary to wage the enforcement conflict.<sup>3</sup>

---

<sup>2</sup>See appendix chapter 50 for a formal proof.

<sup>3</sup>See appendix chapter 50 for a formal proof.

## 2.2 The benefits and costs of varying *specialized* enforcement capacity

When the enforcement technology is specialized, the only way to deter opportunism is to maintain some resources in power-based form in the enforcement sector in peacetime, e.g. even when no one is committing holdups. This reserve strategy reduces the amount of resources the victim of opportunism needs to transform into enforcement assets in the event of a holdup, thereby reducing his transformation cost in response to each holdup. Then his transformation cost will be *less* than his loss from simply letting the opportunist continue to commit holdups. So the victim will find it cost-effective to wage enforcement conflicts in response to holdups.

It pays to increase the size of the peacetime reserve force right up to the point where the marginal benefit from deterring holdups is equal to the marginal opportunity cost of maintaining resources in the enforcement sector in peacetime. This deterrence strategy guarantees that actors will have the capacity to deter larger holdups, while not wasting productive resources on deterring smaller holdups. Moreover, the threat of punishment alone is sufficient to deter the larger holdups. No active conflicts are needed to deter them. And no active conflicts erupt over the smaller ones, because the victims simply permit them without retaliating. So this deterrence strategy, like the previous one, avoids active conflicts completely.<sup>4</sup>

In sum, the main benefit of maintaining specialized enforcement capacity at all times is that it deters some holdups that would not be deterred otherwise. The main cost of maintaining specialized enforcement capacity at all times is that these resources are lost from the productive economy permanently, thereby forfeiting the income they would have generated in the productive economy permanently. Thus, while this deterrence strategy is necessary to secure goods from predation, it creates permanent income losses. The key question is whether the income loss rate from this strategy is higher or lower than the output rate of the productive economy. I will return to this point in a moment.

---

<sup>4</sup>See appendix chapter 50 for a formal proof.



### 2.3 Specialized enforcement capacity takes many forms

The logic described above applies not only to military enforcement capacity but to any form of power-based enforcement capacity that requires specialized assets. Consider the example of a coalition that has formed to impose an economic boycott on an opportunist who committed a holdup. As long as the boycott lasts, each member of the coalition incurs the standard opportunity cost of not trading with the opportunist. If this were the only cost of participating in the boycott, then a coalition could be assembled after each holdup occurs and disbanded after each boycott has imposed a punishment sufficient to deter further holdups. Each coalition member would always have an incentive to join the boycott because his opportunity cost from losing trade during the boycott would always be less than his potential loss from allowing the opportunist to continue to commit holdups.<sup>5</sup>

But this is not the only cost of participating in the boycott. Each member also bears a transformation cost from having to transform his productive assets that are specialized for trade with the opportunist into another form that is specialized for trade with others. Because of this transformation cost, the victim of a holdup may find it more costly to join the boycott than simply to allow the opportunist to continue to commit holdups. So the potential victims of holdups face a choice. They can allow holdups to go unpunished or they can maintain some degree of boycott at all times (e.g. lower levels of specialized investment in trade with potential opportunists at all times). This example shows that the problem of resource transformation costs extends beyond the sphere of military enforcement capacity to other spheres of enforcement that rely on specialized capacity to impose punishments on opportunists.

## 3. The factor that makes power based enforcement unsustainable

When enforcement requires specialized technology, each actor employs a strategy of maintaining some specialized enforcement capacity in reserve at all times to deter holdups (as

---

<sup>5</sup>I abstract away from the free rider problem by assuming: (1) it is a two-member coalition, (2) both members are required to impose the sufficient punishment, and (3) each member is better off joining the boycott than continuing to be the victim of holdups. See Weingast and Hadfield (CITE) for the canonical form of this model.

explained above). Since these resources are lost from the productive economy permanently, the reserve strategy creates permanent income losses. The strategy is sustainable as long as the income loss rate from maintaining a reserve force all the time is *lower* than the output rate of the productive economy. If the income loss rate becomes *higher* than the economic output rate, however, then the strategy is not sustainable. It consumes resources at a higher rate than the economy is producing them, so the economic base shrinks over time. A shrinking economic base cannot sustain power assets for long. Reductions in power assets are inevitable one way or another unless something is done to raise the output rate of the economy or lower the loss rate from the strategy.

Through European history there was one factor that drove statesmen into this dilemma: legal incompatibility disputes. When such disputes arose, it forced statesmen to increase the size of their reserve forces significantly and use the forces frequently enough that the income loss rate became higher than the economic output rate. Their strategies eventually became unsustainable.

### **3.1 Why legal incompatibility disputes are so costly**

A *legal incompatibility* is a situation where one term of a contract gives one actor the right to an asset, but another term of the contract gives another actor the right to the same asset. When a legal incompatibility arises, each actor claims that the other actor is in violation of the contract. It becomes unclear what constitutes a contract violation. The standards for deciding the issue are contested. So each side in the dispute views the other side as the opportunist.

In such a dispute, each actor expects that if he concedes any of the asset to the other one, then the other one will receive a net gain from the dispute and have an incentive to create more legal incompatibilities intentionally to gain more concessions – potentially until he owns the entire asset. That is, each actor fears that if he makes any concession in the current dispute, it will set a "runaway precedent" that eventually leads to his loss of the entire asset. So each actor refuses to agree to any compromise that would allow the other actor to reap a net gain from the current dispute (such as a compromise settlement of "splitting the difference"). When both actors stand firm in this way, the dispute becomes irreconcilable.

In this environment, anchor-based enforcement no longer works. Each actor reverts to power-based enforcement – imposing costs on the other actor by power based methods to punish his perceived opportunism and deter further opportunism. Each actor resists punishment, however, in the belief that he has been nothing wrong and the other actor’s opportunism must be punished or it will continue ad infinitum. At this point, each actor perceives that he must impose costs on the other actor until the other actor has incurred more losses than whatever he would gain from a final compromise settlement of ”splitting the difference”. Only then will he suffer a net loss from the dispute and be deterred from creating legal incompatibilities intentionally in the future. The larger is the asset at stake in the dispute, the longer it takes for each actor to impose enough costs on the other actor to meet this deterrence condition. The result is instability and conflict that lasts as long as the underlying legal incompatibility remains unresolved. Instability and conflict erupt even if there are no uncertainties about relative power and no large shifts in relative power between the two actors. Hence the problem cannot be solved simply by stabilizing the relative power balance and providing full information about relative power to both actors. The problem continues as long as the anchor based governance system remains hobbled by the legal incompatibility problem. This is why legal incompatibility disputes are so costly.

## 3.2 Historical examples

### 3.2.1 The Thirty Years War

In the 1500s and early 1600s, the Protestants in southern Germany were practicing their religion in a way that inspired some Catholics across the border in Austria to convert to Protestantism, thereby threatening the Catholic led government there with popular disaffection and rebellion. The German Protestants denied they were responsible for the problem on the grounds that the conversions in Austria were unintentional on their part. The key question for the Austrian government, therefore, was whether the German Protestants were *knowingly* practicing their religion in a way that was intended to inspire conversions in Austria. If they were doing it knowingly, and the Austrian government permitted them to continue doing it, this would give them an incentive to look for *other* ways to practice their religion so as to inspire *more* conversions in Austria.

To deter this kind of opportunism, the Austrian government would have to demand that the German Protestants change the way they were practicing their religion, and threaten to go to war over the issue if the demand were not met. If the Germans rejected the demand, Austria would have to fight over the issue for long enough to deter the Germans from continuing to practice their religion in the way they were. That is, Austria would have to impose such high war costs on the Germans that they would be better off changing their religious practice than continuing to fight on in defense of it.

This was a new type of spillover problem. There was no adjudication system to distinguish whether the German Protestants were causing it knowingly or unknowingly. For all Austria knew, therefore, the Germans were doing it knowingly and only bluffing that they were doing it unknowingly. Austria could not permit them to continue doing this (for the reason explained above). Hence the Austrian government demanded that the German Protestants change the way they were practicing their religion, and threatened to go to war over the issue if the demand was not met.

In turn, the key question for the German Protestants was whether the Austrian government truly could not tell they were causing the problem unknowingly, or it was only bluffing it could not tell when it really could. Since there was no adjudication system, there was no way for the German Protestants to know what evidence the Austrian government had about the Germans' intention or lack of intention in causing the problem. So there was no way to know whether the Austrian government was ignoring any evidence it had. For all the German Protestants knew, therefore,

1. The Austrian government had evidence that the German Protestants were simply minding their own religious business and not intentionally conducting it in a way so as to inspire political agitation in Austria; but
2. the Austrian government was ignoring the evidence and bluffing that for all it knew, the German Protestants were intentionally inspiring conversions in Austria; so that
3. the Austrian government could demand moderation in the German Reformation for its own gain, even though the demand was not necessary to deter opportunism by the Germans.

So if the German Protestants were to acquiesce to the demand, it would give Austria an incentive to do the same thing on other issues in order to make more demands opportunistically (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew

the opposite). So the German Protestants rejected the Austrian demand and went to war over the issue.

Once war erupted, Austria had to fight over the issue for long enough to deter the German Protestants from continuing to practice their religion in a way that was intended to inspire religious conversions in Austria. In turn, the German Protestants had to fight over the issue for long enough to deter the Austrian government from claiming that, for all it knew, the Germans were inspiring conversions in Austria intentionally.

### **3.2.2 The French Revolutionary Wars**

In the early 1790s, the revolutionaries in France were restructuring their domestic government in a way that encouraged activists in Austria to advocate for a similar restructuring there, thus threatening the noble led government in Austria with being overthrown. The French revolutionaries denied they were responsible for the problem on the grounds that the political agitation in Austria was unintentional on their part. The key question for the Austrian government, therefore, was whether the French revolutionaries were knowingly conducting their domestic revolution in a way that was intended to inspire political agitation in Austria. If they were doing it knowingly, and the Austrian government permitted them to continue doing it, this would give them an incentive to look for more ways to conduct their domestic revolution so as to inspire more political agitation in Austria. Again this was a new type of spillover problem. There was no adjudication system to distinguish whether the French revolutionaries were causing it knowingly or unknowingly. For all Austria knew, therefore, the French were doing it knowingly and only bluffing that they were doing it unknowingly. Austria could not permit them to continue doing that. Hence the Austrian government demanded that the French revolutionaries moderate their domestic revolution – and threatened to go to war over the issue if the demand was not met.

In turn, the key question for the French revolutionaries was whether the Austrian government truly could not tell they were causing the problem unknowingly, or it was only bluffing it could not tell when it really could tell. Since there was no adjudication system, there was no way for the French to know what evidence the Austrian government had about French intention or lack of intention in causing the problem. So there was no way to know

whether the Austrian government was ignoring any evidence it had. For all the French knew, therefore,

1. The Austrian government had evidence that the French revolutionaries were simply minding their own domestic political business, and not intentionally conducting it in a way to inspire political agitation in Austria; but
2. the Austrian government was ignoring the evidence and bluffing that, for all it knew, the French revolutionaries were inspiring the political agitation in Austria intentionally and knowingly; so that
3. the Austrian government could demand moderation in the French Revolution for its own gain, even though the demand was not necessary to deter opportunism by the French.

So if the French were to acquiesce to the demand, it would give Austria an incentive to do the same thing on other issues in order to make more demands opportunistically (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew the opposite). So the French rejected the Austrian demand and went to war over the issue.

Once war erupted, Austria had to fight over the issue for long enough to deter the French from continuing to conduct their domestic revolution in a way that would inspire political agitation in Austria. In turn, the French had to fight over the issue for long enough to deter the Austrian government from claiming that, for all it knew, the French were inspiring political agitation in Austria intentionally.

### **3.2.3 World War II**

In the 1930s, France, England and the United States raised the prices that their domestic consumers were paying for imported goods, thereby lowering the export earnings of German firms.<sup>6</sup> France, England and the United States denied they were responsible for the income loss in Germany on the grounds that it was unintentional on their part. The key question for the German government, therefore, was whether the Western powers were knowingly setting their economic policies in a way that was intended to increase their incomes at Germany's expense. If they were doing it knowingly, and the German government permitted them to

---

<sup>6</sup>The financial markets in France, England and the United States also increased the interest rates that investors were paid for providing capital, thereby raising the capital costs of German firms and lowering the wages of German workers. The same kind of analysis applies to this part of the spillover problem.

continue doing it, this would give them an incentive to look for other ways to set their economic policies so as to increase their incomes at Germany's expense again. Once again this was a new type of spillover problem. There was no adjudication system to distinguish whether the Western powers were causing it knowingly or unknowingly. For all Germany knew, therefore, the Western powers were doing it knowingly and only bluffing that they were doing it unknowingly. Germany could not permit them to continue doing this. Hence the German government demanded that the Western powers moderate their economic policies, or failing that, accept Germany's right to conquer other states and establish direct control over their economic policies.

In turn, the key question for the Western powers was whether the German government truly could not tell they were causing the spillover problem unknowingly, or it was only bluffing it could not tell when it really could. Since there was no adjudication system, there was no way for the Western powers to know what evidence the German government had about Western intention or lack of intention in causing the problem. So there was no way to know whether the German government was ignoring any evidence it had. For all the Western powers knew, therefore,

1. The German government had evidence that the Western powers were simply minding their own economic business, and not intentionally setting their economic policies so as to impose income losses on Germany; but
2. the German government was ignoring the evidence and bluffing that, for all it knew, the Western powers were setting their economic policies at Germany's expense intentionally and knowingly; so that
3. the German government could demand moderation in their economic policies for its own gain – or acceptance of its right to conquest – even though these demands were not necessary to deter opportunism by the Western powers.

So if the Western powers were to acquiesce to these demands, it would give Germany an incentive to repeat the same opportunistic behavior on other issues (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew the opposite). So the Western powers rejected the demands and went to war over the issue.

Once war erupted, Germany had to fight over the issue for long enough to conquer other states and establish direct control over their economic policies – and to deter the Western powers from adopting economic policies designed to increase their incomes at Germany's

expense again. In turn, the Western powers had to continue fighting for long enough to deter Germany from engaging in such conquests – and to deter Germany from claiming that, for all it knew, they were setting their economic policies to gain income at its expense intentionally.

### **3.3 Costly competitions in raising relative power**

Once a conflict like this begins, each actor increases his power based enforcement capacity simply to be able to defend its perceived rights in the dispute. Each actor's imperative is to insure that his power to impose costs on the other does not decline too far. For if either actor's power to impose costs on the other declines too far, then he will have to capitulate in the dispute, leaving the other actor with a net gain and an incentive to create more disputes in the future for similar gains – potentially until he owns the whole asset. To avoid such a defeat, each actor must invest in power-based enforcement tools and match the other's investments one-for-one. This imperative generates a competition in raising relative power.

This competition goes on continuously – day and night – until the deterrence condition explained above is met. Although shifts in relative power between the opponents may cause one or the other to capitulate temporarily, that will not end the competition if the deterrence condition is not met yet (as explained above). The side that capitulated will look for ways to re-boost his relative power so he can get back in the fight and continue imposing costs on the opponent until the deterrence condition is met. Each side perceives that he must continue imposing costs on the other side to avoid the establishment of a runaway precedent that might eventually lead to his loss of the entire asset (as explained above). Hence each side is willing to pursue his competitive strategy into the unsustainable range: where his income loss rate is higher than his economic output rate.

## **4. Power-based enforcement strategies: Effects, costs and trade-offs**

Once a legal incompatibility dispute becomes irreconcilable, each actor has four strategies for imposing costs on the other actor.



1. Seize shares of the asset at stake in the dispute from the opponent unilaterally.
2. Produce and stockpile power-based enforcement tools.
3. Maneuver the enforcement tools to deter the opponent from committing asset seizures and defending his assets from seizure.
4. Use the enforcement tools to destroy some of the opponent's enforcement tools or assets.

The benefit of each strategy is that it imposes costs on the opponent and thereby hastens the day when the deterrence condition will be met and the dispute can end. Strategies 1 and 4 impose costs on the opponent *directly* by seizing or destroying his assets and enforcement tools. Strategies 2 and 3 impose costs *indirectly* by increasing his costs of employing all four strategies. The four strategies differ in their effects and costs, however. These differences create trade-offs that drive actors' decisions about which strategies to use and for how long. By elaborating each strategy's effects and costs, it is possible to derive observable implications that can be tested on the historical evidence.

**Strategy 1.** Seize shares of the asset at stake in the dispute from the opponent unilaterally.

**Effects** When both actors employ this strategy alone, the effect is instability in the distribution of assets between actors. When the dispute concerns a contract that divides political assets, the effect is political instability. For example, if the contract divides territory between two states, the effect is *geopolitical* instability. If the contract divides political authority between two branches of a government, the effect is *constitutional* instability. If the contract divides administrative authority between two agencies of the executive state, the effect is *bureaucratic* instability.

**Costs** Any type of instability creates insecurity of property rights that produces adverse incentive effects. Actors experience reduced incentives to invest in the economy and pay taxes for the production of public goods.

**Observable implications** When a legal incompatibility dispute becomes irreconcilable, the amount of instability will increase. Property rights will become less secure. Eco-

conomic investment and tax payment rates will decrease. Economic growth will slow.

**Strategy 2.** Produce and stockpile power-based enforcement tools.

**Effects** When both actors employ this strategy as well, the effect is a competition in raising relative power. Each actor is driven to invest in power-based enforcement tools and match the opponent's investments (for the reasons explained above). A major tactic for increasing power is to commit holdups against weaker actors not a party to the dispute. In the domestic arena, this takes the form of increasing the tax rates on non-elites without their consent. In the international arena, it takes the form of violating the territorial boundaries and political independence of weak states and then increasing the tax rates on their non-elites without their consent.

**Costs** The cost of these tactics is economic misgovernance. Authority holders expand their authority to tax and regulate the economy opportunistically. When they use that authority to increase expenditure on protection and predation, the cost is a misallocation of resources. When they use that authority to redistribute the tax burden to increase their power, the cost is an inefficient distribution of the tax burden – undertaxing the more influential elements of society and overtaxing the less influential elements. When they use that authority to impose arbitrary exactions on the less influential segments of society, the cost is corruption.

**Observable implications** When a legal incompatibility dispute becomes irreconcilable, investment in power based enforcement capacity will increase: institutions will become stronger in power based enforcement capacity. At the international level, strong states will violate the territorial boundaries and political independence of weak states. At the domestic level, corruption rates and tax rates on non-elites will be relatively high.

**Strategy 3.** Maneuver enforcement tools to deter the opponent from committing asset seizures and defending his assets from seizure.

**Effects** When both actors employ this strategy as well, the first effect is to intensify the competition to raise relative power. Since each side perceives the opponent is jockeying for a relative power advantage at all times, each side is driven to jockey for a countervailing power advantage simply to avoid declining in relative power to the point of having to capitulate and lose the entire asset at stake in the dispute (as explained above). The drive to maintain power, in turn, intensifies the pressure to commit holdups against weaker actors. It also creates pressure to find allies who share the fear that they will be harmed by the other side's opportunism if it is not punished and deterred. The net effect of all this jockeying and competitive coalition building is a decline in social trust due to the elites' fear of opportunism by peer competitors and the non-elites' fear of opportunism by stronger actors.

**Costs** The costs of this competition are further misallocation of resources and social capital depletion.

**Observable implications** When a legal incompatibility dispute becomes irreconcilable, there will be social capital depletion due to fear of opportunism by others. Actors will avoid direct dialogue over disputes and employ indirect subversion instead. To defend against indirect subversion, actors will pursue jurisdictional separation from opponents. To attract allies in the competition for relative coalition power, actors will express social disapproval of opponents.

**Strategy 4.** Use the enforcement tools to destroy some of the opponent's enforcement tools or assets.

**Effects** When both actors employ this strategy is well, the effect is active conflict. Here conflict is defined as an equilibrium in which *each* side imposes costs on the other side against his will by destroying the value of his assets or defenses (e.g. the *mutual* use of power based tools to impose costs). This definition refers not only to military conflict but to any form of conflict in which one side imposes costs on the other side backed by a demand for concessions, but the other side does not back down, and *both* sides incur costs as a result. To take an example from the case studies, a political strike initiated by public-sector workers is

costly to both sides; and a lockout of the workers initiated by the chief executive is equally costly to both sides.<sup>7</sup>

**Costs** Active conflict carries the cost of destruction of resources.

**Observable implication** When a legal incompatibility dispute becomes irreconcilable, the amount of active conflict will increase. Disputes will be settled by relative power rather than facts and law: stronger actors will win, weaker actors will lose.

## 5. Conflicts *versus* competitions in raising relative power

Actors cannot engage in active conflict continuously – day and night – for years. The rate of resource destruction would be too high compared to the growth rate of the underlying economic base. It would drive the economy into the ground. This constraint applies to all forms of conflict. No army can fire its weapons continuously for years without breaks. No state can sustain an economic boycott coalition against a major trading partner for years without breaks. No branch of government can sustain a political strike or lock out continuously for years without breaks. No interest group can sustain a demonstration continuously for years without breaks. Each actor needs to take breaks for several reasons: to replenish his economic base, secure financial loans, find allies, and maneuver in search of a power advantage in preparation for the next bout of active conflict. Because actors need to take breaks, conflict is never continuous. It is always intermittent – punctuated by breaks that produce periods of apparent calm.

However, actors *can* engage in a competition in raising relative power continuously for years. This is what happens when a legal incompatibility dispute becomes irreconcilable. Actors bear the costs of stockpiling enforcement tools and jockeying for power continuously – without breaks – as long as the legal incompatibility problem remains unsolved. If the

---

<sup>7</sup>In early modern France, the king and his ministers locked out the parliamentary magistrates, thereby destroying their official revenue streams from dispensing justice to the public for the duration of the lockout. At other times, the magistrates staged political strikes that halted the provision of justice to the public, thereby destroying the legitimacy of the king and his ministers in the eyes of the public for the duration of the strike.

governance system itself is incapable of managing legal incompatibility disputes and *most or all* of them become irreconcilable, then the entire polity descends into a competition in raising relative power that lasts continuously for years or decades until the governance system is repaired. This problem applies equally to domestic polities and the international polity.

Thus, once a legal incompatibility dispute becomes irreconcilable, the opponents go back and forth between two equilibria:

- Active conflict: an equilibrium in which each of the opposing sides employs strategies 1-4.
- Competition in raising relative power: an equilibrium in which each of the opposing sides employs strategies 1, 2 and 3, but not strategy 4.

This comparison makes clear that an active conflict is a *special form* of competition in relative power – a form in which the opponents resort to strategy 4 *as well as* the first three strategies. Their resort to strategy 4 makes the rate of income loss for each side much higher. That is why they need to take breaks: to manage the higher rates of income loss economically to ensure they can sustain their long-term strategies in the underlying competition to raise power and meet the deterrence condition eventually.

Breaks vary in length depending on the loss management plan they are intended to support. A *tactical retreat* supports the plan of withdrawing from battle to regroup and choose a more favorable time and place to reengage the opponent. A *strategic pivot* supports the plan of accepting a stalemate in one theater in order to conserve resources for pursuing a victory in another theater. A *truce* supports the plan of withdrawing from an active conflict temporarily to restore the economic base, secure loans or find allies. An *entente* supports the plan of pausing an active conflict with one peer competitor in order to defeat another peer competitor that poses a more immediate threat. In each case the competition to raise relative power continues during the break.

In analyzing historical cases of legal incompatibility disputes that became irreconcilable, it is essential to distinguish between the factors that caused breaks in the active conflicts, on one hand, and the factor that enabled an end to the underlying competitions in raising relative power. Breaks in the conflicts were caused by the need to manage income loss rates in

order to sustain the longer-term strategies in the underlying competitions. By contrast, the underlying competitions were only ended by the development of institutions for managing legal incompatibility disputes. This distinction explains why, in some cases, the underlying competition continued after the apparent settlement of the active conflict. The settlement failed to include institutions for managing legal incompatibility disputes effectively.

### 5.1 The deeper cause of conflicts and competitions in raising relative power

The conventional academic view is that there are two main causes of costly conflict: (1) uncertainty about relative power and (2) commitment problems related to large shifts in relative power. CITES. According to this view, actors *always* divide goods according to relative power in peacetime. They *never* escape from this world. That is why uncertainty about relative power *in peacetime* is seen as so important: It enables actors to bluff that they are more powerful than they really are *in peacetime* to gain a greater share of the world's goods. And that is why large shifts in relative power *in peacetime* are seen as so important: They enable rising actors to exploit their relative position *in peacetime* to gain a greater share of the world's goods.

From this perspective the function of conflict is to impose costs on the opponent to solve these problems: to prevent large shifts in relative power *in peacetime* and to deter bluffing about power *in peacetime* in the future. This function is served only by the actual fighting – when each side is firing its weapons at the opponent to impose costs on the opponent. When the fighting stops, the function of conflict is no longer served. From this perspective, the only source of conflict costs is the actual fighting. The conflict costs stop mounting when the fighting stops.

This perspective assumes that actors always divide goods according to relative power in peacetime. On this view, the farthest that actors can fall is: Into an active conflict. The metric of whether they have fallen this far is whether they are actively imposing costs on each other by firing their weapons at each other. By contrast, the theory presented in this study assumes that actors can fall a much greater distance. For they do *not* always divide goods according to relative power in peacetime. Sometimes they *leave* the world of power and enter the world of law in peacetime – a world in which they divide goods on terms that are

*independent* of relative power. In this world they can reduce their power-based enforcement capacity to a much lower level, thereby saving significant costs (see next chapter). When an active conflict erupts, therefore, they are not merely falling into the act of firing their weapons at each other. They are falling from the world of law back into the world of power, where they need to increase their power-based enforcement capacity to a high level again *and keep it there all the time – whether they are firing their weapons or not*. In this world of power, actors pay not only the cost of the active conflict – the cost of firing their weapons at each other – but also the cost of maintaining their power-based enforcement capacity at a high level all the time.

When a legal incompatibility dispute becomes irreconcilable, therefore, it is incomplete simply to say that actors have fallen into an active conflict. The complete truth is to say that they have fallen into a competition in raising relative power. This competition involves phases of active conflict and other phases without active conflict. *Both* phases are costly compared to the period before the conflict erupted (when the actors lived in the world of law). For actors must maintain high levels of power-based enforcement capacity during both phases.

The theory presented in this study identifies a deeper cause of these outcomes. As long as actors have an effective method of dividing goods among themselves on terms that are *independent* of relative power, they can remain in the world of law and avoid *both* active conflicts and competitions for power. If that method of dividing goods fails, however, then actors descend from the world of law into the world of power and suffer *both* active conflicts and competitions for power. Thus the deeper cause of both adverse outcomes is the fact that actors *had* the capacity to divide goods independently of relative power, but *lost* it.

## **5.2 Examples from European history**

In medieval Germany actors divided territory and political authority according to relative power. In medieval England, by contrast, they divided these assets according to anchor contracts – agreements whose terms were independent of relative power. Medieval Germany experienced significantly more conflicts. The conflicts were not continuous, but the competition in raising relative power was continuous. After the Peace of Augsburg, actors

divided territory and religious rights according to relative power (1555-1618). After the Peace of Westphalia, by contrast, they divided these assets according to anchor contracts (1648-1689). The first period experienced significantly more conflicts over religious differences. The conflicts were not continuous, but the competition in raising relative power was continuous. Finally, in each of the seven hegemonic wars among the great powers between 1550 and 1950, the periods of active conflict – when the opposing armies were actually firing their weapons at each other – were intermittent. Yet the competition in raising relative power was continuous from the beginning of each war to its termination and settlement – and in some cases after that. Three of the seven settlements solved the legal incompatibility problem that had caused the war, thereby ending the competition in raising relative power. The other four settlements failed to solve it, so the competition in raising relative power continued for years after the war ended.

## **6. Observable implications of the descent into the world of power**

In the world of law actors divide goods among themselves according to anchor contracts – agreements whose terms are *independent* of relative power. After actors get into an irreconcilable dispute and switch to the world of power, they divide the goods according to relative power. Each of the opposing teams is driven to increase its power simply to avoid falling behind in relative power and losing the entire asset at stake in the dispute to the opposing team. At the same time, though, each team must safeguard the long-term viability of its competitive strategy by sustaining the economic base used to pay team members for their labor. This problem of income sustainment imposes a constraint on the team's level of investment in its power. If it invests too much resources in its power, then its economic base will shrink to the point where the team members' incomes are so low that they replace the team leader with another who will raise their incomes by reducing investment in power. The statesman's decision problem, then, is to maximize his team's power subject to this income sustainment constraint.

To solve this decision problem, each statesman increases the power of his team up to the point that the marginal benefit from the last unit of investment in power is just equal to the marginal cost of that investment. This is the point at which the team has the maximum



amount of power that is consistent with continuing to pay the team members in the long run. While no statesman ever hits this optimal point exactly, history shows that when a statesman missed this point by too much in either direction, his team members replaced him with another statesman who would get the team closer to the optimal point. In some cases, the statesman missed the optimal point by investing too *little* in power. Then the team's share of the good at stake in the dispute declined significantly and the team members were driven to replace the statesman with another who would increase the team's power to gain more of the good. In other cases, the statesman missed the optimal point by investing too *much* in power. Then the economic base used to pay the team members for their labor shrank significantly and the team members were driven to replace the statesman with another who would increase the team's income by reducing its level of investment in power.

### 6.1 The economic structure of security competition varies with dispute type

Each statesman's decision problem can be represented by the standard microeconomic model of a declining marginal benefit curve and a rising marginal cost curve. This model is shown in figure 30. The optimal point is at the intersection of the two curves. This is the point at which the marginal benefit from the last unit of investment in power is just equal to the marginal cost of that investment.

Through European history these curves were shifted in different ways by different types of dispute over legal incompatibility problems. There were two main types of dispute. In disputes over *economic* exclusionism, one team excluded the other team's members from entry into economic markets. In disputes over *political* exclusionism, one team excluded the other team's members from participating in political decisions on tax rates and spending priorities. The two types of exclusion differed in their impacts on each of the opposing team's marginal benefit and cost schedules for investing in power. These differences are represented in figure 30.

In disputes over *economic* exclusionism, the excluders imposed economic market restrictions on the victims. These restrictions gave the excluders monopoly profits, thereby increasing the income-productivity of their civilian economic labor. This is represented in the top left panel of figure 30 as an *increase* in that team's marginal cost of investment in

power. The entry restrictions denied the victims market opportunities, thereby decreasing the income-productivity of their civilian economic labor. This is represented in the top right panel of figure 30 as a *decrease* in that team's marginal cost of investment in power.

In disputes over *political* exclusionism, the excluders denied the victims' right to participate in political decisions on tax rates and spending priorities. These restrictions gave the excluders the capacity to raise taxes on the victims and reduce spending on their policy priorities, while at the same time lowering taxes on excluder team members and raising expenditure on their policy priorities. The latter is represented in the middle left panel of figure 30 as an increase in the excluder team's marginal benefit from paying taxes to its own political authority to invest in its power. The former is represented in the middle right panel of figure 30 as a decrease in the victim team's marginal benefit from paying taxes to that authority.

These shifts in the marginal benefit and cost of paying taxes to the excluders' political authority are representative as long as the victim team continues to respect that political authority and its fiscal agency. Once the victim team rejects that authority and launches a separatist movement, however, the benefit and cost schedules shift in the opposite direction. The victim team experiences an increase in the marginal benefit of paying taxes to invest in the power of its own (separate) political authority to resist the excluder team, since its taxes pay exclusively to fund its own policy priorities. Conversely, the excluder team experiences a decrease in the marginal benefit of paying taxes to its own leaders, since it is a long and costly process to quash a secessionist movement. That is, the marginal income-productivity of military labor devoted to quashing a secessionist movement is low.

These differences between the two types of dispute generate a range of observable implications for patterns of conflict, settlement and long-run development in European history. The main implications are derived in Chapter 12. Further implications are derived in chapters 22, 27 and 31.

## CHAPTER 11

### The world of law, justice and low-cost deterrence

#### 1. The institutions of anchor-based governance

##### 1.1 Anchor contracts

When actors divide assets according to relative power, it provokes a costly competition in raising relative power. Competitive investments in power leave relative power unchanged but all of the competitors poorer. To avoid this outcome, actors divide assets according to terms that are independent of relative power by forming *anchor contracts*. These are agreements whose terms are tied to anchors in the world that do not move even when relative power shifts. With such a contract in place, neither of the two contractors has an incentive to invest in his power. For even if one invests and the other does not – shifting the power balance – the contract terms would not move in his favor. They are fixed: tied to anchors in the world that do not move. Since neither contractor has an incentive to invest in raising his power at the other's expense, a competition in raising relative power is avoided.

Because the world is always changing, the anchors in the real world are always moving too. When the anchors move the terms of the contract move with them, because the terms are tied to the anchors. For example, in a long-term wage contract with an inflation adjustment clause, an increase in the inflation rate (the anchor) results in an increase in the nominal wage rate (the contract term). All kinds of anchor contracts have adjustment clauses designed to ensure that the contract terms keep pace with a changing world. The purpose of an adjustment clause is to specify how the actors' relative shares of the asset in question will be allowed to shift when an anchor in the world moves.

As long as each actor knows that the other actor did not move the anchor in his favor intentionally to gain a greater share of the asset, then each actor can allow the other to receive

a greater share of the asset when an anchor moves in his favor. But if one actor knows that the other actor moved an anchor in his favor intentionally, then the first actor cannot allow the second actor to receive a greater share of the asset. Actors therefore need some kind of adjudication system that can tell the difference between intentional and unintentional movements of the anchors.<sup>1</sup>

## 1.2 anchor-based adjudication with a high capacity for Legal Incompatibility Management (LIM)

If anchor term contracts are to function well, there must be an adjudication system that can distinguish between cases in which one party moved the anchor in his own favor intentionally and cases in which an anchor moved in his favor due to forces beyond his control. The effectiveness of the system in drawing this distinction depends on four capacities.

- The capacity to manage contractual and legal ambiguities by distinguishing between questions of fact and questions of law (or more generally, between questions of fact and questions of the interpretation of contract terms, rules and laws).
- The capacity to manage legal incompatibilities by fitting new cases of legal incompatibility into existing categories of legal doctrine that have the properties of an anchor contract.<sup>2</sup>
- The capacity to preserve adjudication decisions for future use as authoritative precedents.

If the adjudication system lacks any of these capacities, then cases of legal incompatibility may arise in which each actor suspects, but cannot prove, that the other actor an anchor intentionally to create the legal incompatibility for his own gain. Then the only way to deter repeats of such opportunism is for the contracting parties to engage in costly conflict whenever an anchor moves in either party's favor for any reason – *even reasons beyond his control*. In this kind of conflict, each actor must impose enough costs on the other actor

---

<sup>1</sup>The ideas in the first three paragraphs of this section are due to Rowe (1989, 1990). The ideas in the following paragraphs are mine.

<sup>2</sup>A legal incompatibility is a situation where the real world anchors have moved in a new way that renders the terms of the existing contracts *incompatible* with each other. One term gives one actor the right to an asset, but another term gives another actor the right to the *same* asset. See Chapter 10 for further explanation of the problem of legal incompatibilities.

to deter him from moving the anchor opportunistically in the future – and from lodging possibly fraudulent claims of opportunism again in the future.<sup>3</sup>

**Historical examples from domestic politics** In medieval England, the legal system had a set of “forms of action” whose function was to fit new cases of legal incompatibility into existing categories of legal doctrine with fixed terms (see Chapter 13). In medieval Germany, by contrast, the legal system had no capacity to fit new cases of legal incompatibility into existing categories of legal doctrine. The result was a proliferation of new legal categories that hobbled the adjudication process (see Chapter 14).

**Historical examples from international politics** In the decades after the Peace of Westphalia, the legal system of the Holy Roman Empire developed new legal doctrines with fixed terms for use in solving legal incompatibility problems (see Chapter 35). In the decades after the Peace of Augsburg, by contrast, the legal system lacked legal doctrines with fixed terms. Quite the opposite, the most important legal doctrine – the ruler has the right to determine the religion of his subjects – varied with the relative power of the rulers and territories. (see Chapter 35).

### **1.3 Anchor-based enforcement: Harnessing the natural variability of the world for enforcement purposes**

Whenever an anchor moves in one actor’s favor, he receives a random variable benefit (as explained above). What prevents or deters an actor from intentionally moving an anchor in his favor to increase his share of the asset? What form of punishment can the other actor threaten to impose to deter him – or impose after the fact to restore deterrence? If the imposition of punishment requires the use of power-based enforcement tools, then the outcome is a costly competition in raising relative power. In this case the honest actors must invest in power tools to maintain deterrence. In response the dishonest actors invest in countervailing tools that increase their power to evade detection and withstand punishment. The competitive investments do not shift relative power, but they leave both the honest and

---

<sup>3</sup>See the previous chapter for a full explanation of how legal incompatibility disputes because conflict.

the dishonest actors poorer.<sup>4</sup> How can the parties to an anchor contract threaten to impose punishments to deter violations – and impose them after the fact to restore deterrence – without relying on power-based enforcement tools to make their deterrent threats credible?

The natural variability of the world provides an alternative enforcement mechanism that does not rely on power-based enforcement tools. Because the world is always changing, the anchors in the world move periodically too. An anchor contract specifies how the distribution of contractual rights will shift between the contracting parties when the anchors move due to random forces beyond either party's control. As long as each party knows that the anchor in the world moved due to forces beyond the other party's control, he can allow the distribution of contractual rights to shift in the other party's favor when that is what the contract specifies. The party that is favored by the anchor's movement thus receives a *variable benefit*. His benefit is variable because he receives it only periodically – at those times when the anchor moves in his favor due to forces beyond his control.<sup>5</sup>

This feature of the contractual world enables the two parties to the contract to strike a special kind of bargain that maintains deterrence without reliance on power-based enforcement tools.

- The first party commits that whenever he is found guilty of moving the anchor in his favor intentionally for his own gain, he will admit guilt and concede defeat voluntarily – without resistance.
- In return, the second party commits that whenever the anchor moves in the first party's favor due to forces beyond his control, the second party will grant the first party his variable benefit voluntarily – without resistance.

---

<sup>4</sup>See the appendix in chapter 45 for a formal model.

<sup>5</sup>The ideas in this paragraph and the previous paragraph are due to Rowe (1989b). The ideas that follow in the remainder of this chapter and this study are mine. Although Rowe identified the potential to avoid competitions in raising relative power by using contracts whose terms do not shift with relative power, he did not identify the key condition that actors need to meet to achieve this solution (justice bargains). He also did not derive the observable implications of his ideas or test them methodically on empirical evidence. (For example, he does not discuss European history at all.) I undertake all of these tasks of theory building and theory testing in this study. Nonetheless, Rowe's ideas are seminal and without them this study would never have been conceived or conducted. I first cited and applied Rowe's ideas to the analysis of conflicts and settlements in European history in my earliest studies of this dissertation topic between 1997 and 2002. In the appendix, I provide citations to my earlier studies and quotations of the passages that link my analysis to Rowe's prior work.

As long as the first party gains more from receiving his variable benefit periodically than he loses from his concessions in the cases where he moved the anchor intentionally, he is better off adhering to his half of the justice bargain (e.g. conceding defeat voluntarily when found guilty of such opportunism). In turn, as long as the second party loses less from granting away the variable benefit periodically than he would lose from having to switch to a costly power-based method of enforcement, he is better off adhering to his half of the justice bargain as well (e.g. granting away the variable benefit periodically).

As long as the random movements of the anchors are large or frequent enough, the threat of losing the random variable benefits will be sufficient to deter the first party from moving an anchor in his favor intentionally. I call this enforcement method a *justice bargain* for two reasons. It requires an adjudication system to distinguish between the intentional and unintentional movements of the anchors. And it relies on anchor contracts because these contracts provide random variable benefits in return for compliance with the contract – a provision that can be withheld in punishment for violating the contract. I also call it *anchor-based enforcement*.

This kind of bargain can be enforced without either party having to raise his relative power to make his enforcement threats credible (or any other party having to do so for that matter). Each party can be deterred from renegeing on the bargain simply by the threat that the other party will renege on his half of the bargain in retaliation.

- The first party can be deterred from moving the anchor in his own favor intentionally by the threat that if he is found guilty of doing so, then he will be denied his variable benefit *at the other times* when the anchor moves in his favor due to factors beyond his control (e.g. the threat that the second party will renege on his half of the justice bargain in retaliation). This form of retaliation is cost-free for the second party. In fact the second party actually *saves* resources by retaliating in this way. He does not have to grant the first party the variable benefit any more, so he can keep the variable benefit for himself. In this sense the imposition of the punishment is beneficial to the imposer rather than being costly to the imposer (as is usually the case).
- The second party can be deterred from withholding the variable benefit opportunistically (when the first party actually has a right to it under the contract terms) by the

threat that if he is found guilty of doing so, then the first party will not admit guilt and concede defeat voluntarily *at the other times* when is found guilty of having moved the anchor in his own favor intentionally.

Credible threats to retaliate in these ways are sufficient to enforce the justice bargain. So neither party needs to maintain or deploy power assets for enforcement purposes. Hence neither party needs to invest in raising his relative power to render his retaliation threat credible. In this way the natural variability of the world enables anchor contracts to be enforced without recourse to costly power-based methods.

### **1.3.1 The complementarity between anchor contracts and justice bargains**

Natural variability can be leveraged in this way only because of the first feature of anchor contracts: the terms are tied to anchors in the real world that remain stationary at least some of the time. Thus the two main features of anchor contracts are mutually reinforcing. The more effective actors become at choosing anchors that do not shift with relative power, the more low-cost enforcement leverage they gain. In turn, the more enforcement leverage they gain and the lower enforcement costs become, the higher are the expected returns from forming more anchor contracts on top of the first set of contracts. By implication, political systems whose members become skilled at this iterative process are predicted to perform well in the process of long run development, while systems whose members lack incentives to develop these skills are predicted to perform poorly.

### **1.3.2 The relation between anchor-based enforcement and power-based enforcement**

When anchor contracts and justice bargains function well, actors can secure a variety of rights and resolve a variety of disputes without recourse to costly power-based methods (e.g. without costly conflicts and competitions in raising relative power). Hence actors in society do not need to amass or deploy power assets to secure themselves from each other's predations – or the government's predations. In turn, the government does not need to amass or deploy power assets to secure actors in society from each other – or to secure its own agencies from predation on each other. Large amounts of resources are thereby released for investment in



the productive economy without compromising deterrence or risking a descent into conflict.

Not all contracts are anchor contracts. There are always some contracts that require power-based enforcement methods. Nonetheless, the enforcement cost savings from anchor contracts makes the difference for long run development, because it reduces the equilibrium tax rate on private economic actors. In turn this lowers the private opportunity cost of public spending on the power-based enforcement of other contracts, thereby enabling enough spending on it to turn the tide from disorder to order.

In times and places where this road to development is successful, later observers may assume that *all* of the contracts were enforced by power-based methods and conclude *a fortiori* that the developmental success was produced by a "strong state" or a "hegemonic state". Yet the inference of power-based strength is unwarranted.<sup>6</sup>

## **2. Anchor-based enforcement has comparative advantages over enforcement based on collective boycotting**

In the standard model of collective boycotting, the single victim of opportunism cannot impose sufficient costs on the opportunist through boycotting to deter further opportunism. The victim must find enough third parties to join his boycott that the *coalition* of boycotters can impose sufficient costs on the opportunist to deter further opportunism. This enforcement method has three disadvantages, however. And anchor-based enforcement has

---

<sup>6</sup>In technical terms, anchor-based enforcement has a comparative advantage over power-based enforcement in *some* domains of contracting. Within these domains, the two enforcement methods are strategic substitutes. That is, the more effective a state becomes at anchor-based enforcement in these domains, the less profitable are its investments in power-based enforcement in these domains. Hence the less incentive it has to make such investments and the less it makes in equilibrium. However, between these domains and the other domains where power-based enforcement has the comparative advantage, the two enforcement methods are strategic complements. That is, the more effective a state becomes at anchor-based enforcement in the former domains, the more profitable are its investments in power-based enforcement in the latter domains. Hence the more incentive it has to make such investments and the more it makes in equilibrium. Overall, across all domains, the state makes less investment in power-based enforcement as it becomes more effective at anchor-based enforcement. But it makes more investment in power-based enforcement in those domains where that method has a comparative advantage, thereby turning the tide from disorder to order. See Bulow, Geanakoplos, and P. D. Klemperer (1985) for an explanation of strategic complements and substitutes. See Tirole (1988) and Milgrom and Roberts (1990, 1995) for applications. See Garfinkel and Skaperdas (2012) for a different approach to the relation between power-based enforcement and other methods of enforcement.

a comparative advantage on all three points.<sup>7</sup>

## 2.1 Costly joining *versus* beneficial joining

In the standard model of collective boycott, each potential boycotter would find it costly to boycott the opportunist, so the beneficiaries of boycotting face a collective action problem. They must find a way to ensure that each potential beneficiary joins the boycotting coalition, or else he is punished himself for not joining. This introduces a second-order punishment problem on top of the first one – and therefore a second collective action problem on top of the first one. In fact there is an infinite regress of such problems in this class of models. The only ways to resolve the infinite regress are: (1) to assume that some actor with a relative power advantage over the others solves the last problem in the chain or (2) to assume away the last problem in the chain by making some other assumption.

Anchor-based enforcement avoids such problems. Each punisher actually *gains* something by imposing the punishment on the opportunist (as explained above). So third-party punishers have an incentive to *join* the boycott rather than an incentive to *free ride* on it. Correspondingly, they have an incentive to *inform themselves* of each actor's reputation as an honest contractor or an opportunist – as long as their costs of gathering such information are less than their gains from imposing the punishments on the opportunists. Thus, this model does not assume away any collective action problems or rely on any actor having a relative power advantage over others.

## 2.2 The centrality of anchor-based enforcement in two-actor systems (and any other systems that rely on them)

In some contexts there are only two parties to a contract and neither party has access to any third party support that would be sufficient for enforcement purposes. In such contexts collective boycotting is not a feasible enforcement method. Anchor-based enforcement is feasible, however, because it does not require third parties. If either one of the two parties reneges on the contract, then the other party has both the incentive and the capacity to

---

<sup>7</sup>For standard models of collective boycotting, see among others, Randall Calvert (date). For the latest application of such models to the rule of law, see Weingast and Hadfield 2011.

retaliate sufficiently to restore deterrence without help from a third party. He has the incentive because he actually *saves* resources by retaliating (as mentioned above: he does not have to grant the first party the variable benefit any more). And he has the capacity to retaliate sufficiently whenever the first party's variable benefit is large enough (e.g. whenever the anchor moves in the first party's favor frequently and widely enough to enable formation of this kind of contract in the first place).

Are there any situations in which there are only two parties to a contract and neither party has access to third party support for enforcement? This situation arises in several contexts in the process of long-run development. For example, in a divided society where each ethnicity includes 40% or more of the population, neither ethnicity can turn to any third party for enough support to impose enough costs on the other ethnicity through boycotting. To take another example, in authoritarian regimes a succession crisis may divide the state elite into two factions, each of which includes 40% or more of the elite. In a contest where victory requires an overwhelming power advantage, neither faction can turn to any third party for enough support to impose enough costs on the other party to win outright. Finally, in a majoritarian system where each major party is supported by 40% or more of the population, each party could win a majority of the vote by convincing enough of the swing voters to join its side. But in a contest where victory requires a more overwhelming power advantage (such as a constitutional crisis, a domestic revolution or a civil war), neither party can turn to any third party for enough support to impose enough costs on the other party to win outright.

In two-actor systems like these ones, anchor-based enforcement has a comparative advantage over other methods of enforcement that rely on collective boycotting. Moreover, the two-actor contract in such systems is foundational for political order and economic governance, and therefore it is foundational for any other contracts that depend on these capacities (e.g. even contracts that can be enforced by collective boycotting by third parties). In this sense anchor-based enforcement is central not only for two-actor systems but also for other systems that rely on them.

### **2.3 Collective boycotting requires symmetric common knowledge of the meaning of contract terms, laws, constitutions, treaties etc.**

In the standard model of collective boycotting, the victim of opportunism and the potential joiners of his boycott are in complete agreement about what constitutes a violation. So they are in complete agreement that a violation was committed and the opportunist who committed it deserves punishment. When this "complete agreement" condition is met, the victim has a relatively easy time finding third parties to join his boycott while the opportunist has a difficult time finding confederates to defend himself from punishment. Since most honest contractors are not interested in joining the opportunist's coalition of self-defense, his main option is to go into outlawry and search for other outlaws with whom he can form a defensive coalition.<sup>8</sup> Since the average actor has no interest in going to that length to be able to commit opportunism, he can be deterred from committing opportunism by the threat of a collective boycott.

The story is different when there is disagreement about what constitutes a violation. The (putative) victim of (putative) opportunism may have an easy time finding third parties to join his boycott when he searches among those who agree with his interpretation of what constitutes a violation. But those who take the opposite view are more likely to join the putative opportunist in defending himself from what is seen as an undeserved punishment. If the disagreement over what constitutes a violation is deep enough, it may split the group of potential boycotters down the middle – with 40-60% taking the putative victim's side and the other 40-60% taking the putative opportunist's side. Then the whole logic of the collective boycott falls apart: the victim cannot amass a coalition of boycotters large enough to impose enough costs on the opportunist to deter him from repeating the action in question. Thus, in cases where the conflict of interest between the victim and the opportunist is due to a dispute over what constitutes a violation in the first place, collective boycotting does not work as a method of enforcing contracts, law and order.

This problem arose repeatedly throughout European history. Each time it began with a dispute over what constituted a violation in the first place, followed by a breakdown in the capacity of domestic institutions to resolve such disputes. Domestic politics then descended

---

<sup>8</sup>Or an offensive coalition.

from a world of law into a world of power where the distribution of valuable assets among actors was determined by costly conflicts and competitions in raising relative power. Each of the opposing sides sought to increase the size of its coalition by appealing to third parties to join its side. Yet neither side obtained enough of a power advantage to defeat the other side. In each case the main solution was not to restore the effectiveness of third party boycotts. The solution was to reform domestic institutions to increase their capacity to form anchor contracts and justice bargains and sustain them over time in the face of disputes over what constitutes a violation in the first place. This body of evidence suggests that Contractual Realism offers a better explanation of the foundations of contract enforcement, law and order than the collective boycott model.

Weingast and Hadfield (2011) provide a model of the rule of law in which the primary function of the stability of law is to enable complete agreement between victims of wrongdoing and third parties as to what constitutes a violation in the first place, so that the victims will have no problem convincing the third parties to join their collective boycotts of the opportunists.<sup>9</sup> This explanation of the stability of law differs fundamentally from the explanation provided by the theory of Contractual Realism developed in the present study. According to Contractual Realism, the primary function of the stability of law is to avoid costly competitions in raising relative power by enabling actors to enforce contracts, law and order *without* descending into rivalries over which side can amass a larger coalition to collectively boycott or embattle the other side. In so far as the Contractual Realist theory produces a range of observable implications that are corroborated by the historical evidence on the emergence of the rule of law in European history, this theory of the function of law seems to offer more explanatory power. This observation suggests a need to reconceptualize the basic function of law – and politics generally for that matter.

---

<sup>9</sup>In this approach Weingast and Hadfield follow North and Weingast (1989) and Weingast (1997).

### **3. The cost of deterring holdups: A cost comparison between power-based enforcement and anchor-based enforcement**

#### **3.1 Power-based punishment is costly to impose, while anchor-based punishment is beneficial to impose**

The two methods differ in the costliness of imposing punishments on violators. With power-based enforcement, the enforcer finds it *costly* to impose punishment on the violator. The costly element is usually armed force, which is costly to make and use. But non-kinetic tools of power-based enforcement are also costly to use.<sup>10</sup> By contrast, with anchor-based enforcement, the punishment takes the form of withholding the violator's random variable benefit. The enforcer gets to keep the random variable benefit for himself. So he actually *benefits* from imposing this punishment.

Since the enforcer finds it beneficial to impose anchor-based punishment, he can impose a more severe punishment (compared to the case with power-based enforcement). The more severe is the punishment, in turn, the lower the probability of detection needs to be to maintain deterrence. Thus, when anchor-based enforcement is used, monitoring efforts can be lowered and yet holdups can still be deterred by imposing severe punishments on the holdups that are detected. Anticipating such punishments, actors will comply strictly with the anchor contracts even when they are not monitored so closely. Thus, in the world of anchor contracts, compliance rates can be high even while monitoring efforts are low.

#### **3.2 Power-based punishment is resisted, while anchor-based punishment is accepted willingly**

The two enforcement methods differ in the target's response to being punished. With anchor-based enforcement, the target acquiesces willingly to the punishment (as explained above in the section on justice bargains). So the enforcer does not need to employ power-based enforcement tools to impose punishment against the target's will. By contrast, with power-

---

<sup>10</sup>An economic boycott is costly to impose because the boycotters must forego lucrative trading opportunities with the target. A labor strike is costly to impose on management because the workers lose wages during the strike. A lockout is costly to impose on workers because management loses output and profits during the lockout.

based enforcement, the target may resist if his relative power is high enough. So the enforcer must maintain and deploy power-based tools to be able to impose punishment against the target's resistance if necessary.

### **3.3 Power-based punishment requires investment in specialized enforcement assets, while anchor-based punishment does not**

In the case of power-based enforcement through kinetic means, it requires specialized investment in military and paramilitary technology. In the case of power-based enforcement through boycotting, it requires the withdrawal of productive investments specialized for trade with the target of the boycott and their redeployment to forms specialized for trade with others. In either case the transformation of the investments back and forth from the enforcement use to the other use is costly. And this creates a vulnerability to holdups (as explained in the previous chapter).

## **4. The incentive to invest in anchor-based governance**

### **4.1 The fundamental investment transformation**

Anchor-based enforcement does not rely on power-based enforcement tools to punish violators or make the threat to punish credible. Hence actors do not need power-based enforcement tools to protect the assets that are governed by anchor contracts. For such assets, actors can divest from power-based enforcement tools and reinvest in productive economic capital. This *fundamental investment transformation* amounts to "taking one's guard down". The investment transformation is never total, because there are always some assets that cannot be divided by anchor contracts. To protect these assets some power-based enforcement tools must be maintained. But many assets can be governed by anchor contracts – both economic and political assets. Hence there can be a significant investment transformation. The resulting increase in economic growth is the main value of reliance on anchor contracts and anchor-based enforcement. When this growth potential is high, actors have a strong incentive to switch from power-based enforcement to anchor-based enforcement and invest in the institutions of anchor-based governance. Similarly, after a long and costly competi-

tion in raising relative power, the value of switching to anchor-based enforcement is simply to get the economic growth rate out of the negative range – that is, to reduce the income loss rate from the competition. This too gives actors an incentive to invest in anchor-based governance institutions.

**Historical examples** Some of the most important governance reforms in European history – at both the domestic and international levels – followed long periods of costly competition in raising relative power. In medieval England, the governance reforms of Henry II followed the long civil war of King Stephen’s reign. The Westphalian governance reforms in 17th-century Germany followed the Thirty Years War. The domestic governance reforms in 17th and 18th century Prussia also followed the Thirty Years War. The diplomatic governance reforms in 19th-century Europe followed the French Revolutionary Wars. The international trade governance reforms in the Bretton Woods era followed World War II.

## 4.2 The new holdup problem

Once actors disinvest from power-based enforcement tools even partly and enter the world of law, they become more vulnerable to conventional ”holdups”.<sup>11</sup> For the vulnerability to holdups is a function of the cost of investing in enforcement tools to impose punishment. After disinvesting from power-based enforcement tools, the cost of reinvesting to the optimal level would be high – in part because it would unleash a competition in raising relative power again. This means that opportunists can commit relatively large holdups (e.g. ones with relatively high value). For they know that although the victim would lose much by doing nothing and simply accepting the loss from a large holdup, he would find it even costlier to reinvest in power-based enforcement tools to impose punishment.

---

<sup>11</sup>Recall from earlier that a holdup is when one party unilaterally seizes a share of the contracted asset from the other party, because the latter would find it costlier to impose punishment than to accept the loss of that share.



### 4.3 The incentive to invest in anchor-based governance

Anchor-based enforcement is less costly than power-based enforcement (as explained above). So it can deter holdups that are undeterrable by threats to revert to power-based enforcement. Since the cost of imposing punishment by anchor-based means is so much lower than the cost of imposing punishment by power-based means, opportunists know they can only get away with relatively small holdups. If they commit a holdup whose value is greater than that low threshold, then the victim will find it cheaper to impose punishment by anchor-based means than to do nothing and simply accept the loss from the holdup. This calculus gives actors vulnerable to holdups an incentive to make fixed investments in anchor-based governance institutions. The purpose of fixed investments in anchor-based governance is to keep the *marginal* cost of imposing punishment by anchor-based means – the cost per individual case – low. As long as that marginal cost remains low, actors can deter even the medium and small size holdups that were undeterrable through power-based enforcement.

This assumes that fixed investment in adjudicative institutions, procedures and doctrines is effective in lowering the marginal cost of adjudication per individual case. The effectiveness of such investments is determined above all by their capacity for Legal Incompatibility Management (LIM). When the LIM capacity is high, such investments achieve a low marginal cost of adjudication per individual case. When the LIM capacity is low, by contrast, such investments fail to lower the marginal cost of adjudication per case. That cost remains higher than the cost of simply accepting the losses from holdups. Then actors have no choice but to revert to power-based enforcement.

**Observable implication:** When LIM capacity is high, investment in anchor-based adjudication will be high and actors will remain in the world of law. When LIM capacity is low, investment in anchor-based adjudication will be low and actors will revert to the world of power.

**Historical examples from domestic politics** In medieval England, the legal system had high LIM capacity. Investment in anchor-based adjudication was high and actors

remained in the world of law (see Chapter 13). In medieval Germany, by contrast, the legal system had low LIM capacity. Investment in anchor-based adjudication was low and actors remained in the world of power (see Chapter 14).

**Historical examples from international politics** In the decades after the Peace of Westphalia, the legal system of the Holy Roman Empire had high LIM capacity. Investment in anchor-based adjudication was high and actors remained in the world of law (see Chapter 35). In the decades after the Peace of Augsburg, by contrast, the legal system of the Holy Roman Empire had low LIM capacity. Investment in anchor-based adjudication was low and actors remained in the world of power (see Chapter 35).

## 5. The sources of stability and rigidity in anchor-based governance

In the world of law there are two types of holdups. One type of holdup is driven by shifts in relative power between the contracting parties. The other type is driven by ambiguities in the contract terms. Both types of holdup can be deterred by anchor-based enforcement.<sup>12</sup> The deterrence of the first type generates *stability* in the distribution of assets across actors. The deterrence of the second type generates *rigidity* in adherence to contract terms as originally written.

### 5.1 The source of stability

The world is always changing in ways that produce marginal fluctuations in relative power between the two contracting parties. Depending on the type of contracted asset, this may include fluctuations in market power, political power, military power, or some other source of bargaining leverage. An opportunist might seize upon momentary power advantages to commit holdups – unilaterally seizing a share of the contracted asset from the other party. Such holdups can be deterred by anchor-based enforcement: the threat of withholding the opportunist's random variable benefits. This kind of deterrence removes the incentive for such opportunism. It removes the incentive to search for potential power advantages and

---

<sup>12</sup>For a formal model that reaches this conclusion see the appendix in chapter 36.

maneuver into position to seize them. When actors stop jockeying for power in these ways and simply adhere to the contract terms, it makes the distribution of asset shares between them *stable*. The distribution does not change even when relative power shifts between the contracting parties. In this sense the parties have left the world of power – where assets are divided according to relative power even in peacetime – and entered the world of law – where assets are divided on terms that are independent of relative power.

**Observable implication** When actors rely on anchor-based governance, there will be stability in the distribution of assets across actors. Stronger actors will not violate the contractual rights of other actors even when they have enough power on the margin to do so.

## 5.2 The source of rigidity

Even if relative power never changes, the world is always changing in other ways that cause the anchors to move. Sometimes the anchors move in a way that is not specified in the contract terms, exposing a gap or ambiguity in those terms. An opportunist might intentionally move an anchor into one of these gaps in order to create a dispute over the meaning of the terms and demand a renegotiation based on the principle of "splitting the difference" in disputes. If the victim gives in to the demand every time, then the opportunist can continue the ploy *ad infinitum*, whittling away the victim's share of the contracted asset to nothing. This type of opportunism is possible as long as contracts are incomplete (e.g. as long as contracts fail to cover all possible future contingencies so gaps and ambiguities in the terms are always possible). Moreover, this type of opportunism is possible even when there are no shifts in relative power between the contracting parties.

In theory these holdups are deterrable by anchor-based enforcement – the threat of withholding the opportunist's random variable benefits. In practice such deterrence depends on the adjudication costs of distinguishing between intentional and unintentional disputes. If the adjudication costs are less than the cost of reverting to power-based enforcement, then it is economical to use anchor-based enforcement to punish opportunists who create disputes

intentionally. In this case the threat to adjudicate is credible and such opportunism will be deterred. If the adjudication costs are more than the cost of reverting to power-based enforcement, however, then it is uneconomical to use anchor-based enforcement. In this case the threat to adjudicate is not credible. If opportunists create disputes with such high adjudication costs, and it is not clear to the victims whether the opportunists created them intentionally or unintentionally, then the outcome will be a reversion to power-based enforcement that makes everyone poorer. To avoid this outcome, actors have an incentive to avoid creating disputes with high adjudication costs, even unintentionally, by adhering to contract terms *as written* and ignoring gaps and ambiguities in the terms to the extent possible.

### **5.3 The transcendent value of adhering to contract terms as written**

By using anchor contracts, actors avoid costly competitions in raising relative power and save considerable resources. This savings gives them a strong incentive to adhere to the terms of the anchor contracts *as originally specified*. By doing so they avoid the risk of unintentionally provoking disputes over gaps and ambiguities in the terms – disputes for which the adjudication costs may be so high as to trigger a reversion to power-based enforcement and costly competitions in raising relative power that make everyone poorer. That vicious spiral is avoided by adhering to contract terms as written. This is the transcendent value of adhering to the terms as written.

When actors employ a strategy of adhering to contract terms as written, it requires them to ignore both momentary shifts in relative power and worldly changes that might otherwise expose gaps and ambiguities in the contract terms. Their ignorance of these facets of the world further reinforces the stability of asset distributions among contracting parties. The distributions remain unchanged despite power shifts and subtleties of contractual interpretation.

### **5.4 Anchor contracts resolve the tension between two basic logics of action**

There are two different logics that an actor may employ in choosing his actions. If he employs the *logic of consequences*, he chooses between two options by comparing their consequences and choosing the option that yields him the highest net benefit. If he employs the *logic of*

*appropriateness*, by contrast, he consults a socially agreed rule and chooses the option that is deemed appropriate by that rule *even if it does not yield him the highest net benefit*. In other words, he adheres to a contractual logic of appropriateness – adhering to contract terms as originally written – regardless of momentary power shifts *or* contractual subtleties.

There is a fundamental tension between these two logics of action. This tension is resolved by anchor contracts – which place a high material value on adherence to contract terms as originally written. By following the contractual logic of appropriateness – adhering to mutually agreed rules – actors avoid costly competitions in raising relative power that would make everyone poorer. No actor follows the contractual logic of appropriateness out of concern for another actor's welfare though. Each actor complies with the contract terms out of self-interest: If he commits either of the two types of holdup discussed above, then he will be denied his random variable benefits in punishment. If he unintentionally creates contract disputes with high adjudication costs, then he will be forced to incur (his share of) the costs of reversion to power-based enforcement. In either scenario, it is his individual cost-benefit calculus that drives his adherence to the contractual logic of appropriateness.

At the same time, his adherence commits him to follow mutually agreed contract terms rather than his momentary self-interest at times when the two imply different actions. At such times his behavior is pro-social rather than self-interested. This is the calculus that connects the logic of appropriateness and the logic of consequences. And it is the reason why legal incompatibility disputes are the main cause of reversion to the world of power.

## **6. Why legal incompatibility disputes are the main cause of reversion to the world of power**

In the world of law, actors adhere to contract terms as written and expect their fellows to do the same. If one actor departs from the contract terms at another's expense and the latter objects, there are only two ways to resolve the dispute: adjudication or a reversion to the world of power. Adjudication works for cases of a clear violation of the contract. It also works for cases of a contractual ambiguity that arose unintentionally and can be resolved by splitting the difference. Adjudication does not work for cases of a legal incompatibility, how-

ever. In such a case neither side can make any concessions because that might set a runaway precedent that would allow the other side to gain the whole asset in the end. Neither side can prove that it did not create a legal incompatibility intentionally, because his demand for his full contractual right without any concessions will always seem intentional to the other side. In short, a legal incompatibility confounds the anchor-based system of governance because that system is based on strict adherence to contract terms and the expectation of strict adherence from one's fellows. Since adjudication does not work, the only other option is to revert to the world of power.

**Observable implication** If the capacity for Legal Incompatibility Management (LIM) is *high*, then social trust will remain high and vehicles of social trust will remain intact. Jurisdictional integration will remain high. Actors will engage in direct dialogue and expressions of social approval toward each other. If the capacity for Legal Incompatibility Management (LIM) is *low*, then social trust will become low and vehicles of social trust will decay. There will be jurisdictional disintegration. Actors will avoid direct dialogue, resort to indirect subversion and express social disapproval toward each other.

**Observable implication** If the capacity for Legal Incompatibility Management (LIM) is *high*, then actors will remain in the world of law. If the capacity for Legal Incompatibility Management (LIM) is *low*, then actors will revert to the world of power.

## CHAPTER 12

# Hard-core assumptions, causal claims and observable implications

Contractual Realism is composed of seven hard core assumptions. Together these assumptions provide a general explanation of conflict, settlement and long-run development in European history. In this chapter I lay out the hard core assumptions, the main causal claims, and the observable implications. I present the hard core assumptions in section 1, the main causal claims in section 2, and the observable implications in section 3.

### 1. Hard core assumptions of Contractual Realism

**Assumption 1a.** Actors divide valuable assets amongst themselves in two different ways.

- At times they divide assets according to agreements whose terms *vary* with changes in their relative power. This induces a competition in raising relative power, however. Such competitions are so costly that they dissipate all of the gains from such agreements.
- To avoid this problem, actors divide assets according to agreements whose terms do *not* vary with changes in relative power. The terms are written to refer to *anchors* in the real world that do not move even when relative power shifts among actors. I call such agreements anchor contracts. Such a contract insures that neither actor can shift the distribution of the asset in his favor by increasing his relative power. The agreement thus avoids costly competitions in raising relative power.<sup>1</sup>

**Assumption 1b.** Since the world is always changing, the anchors sometimes move randomly due to forces beyond any actor's control. The contracts specify how the distribution of assets

---

<sup>1</sup>The idea of a contract whose terms are invariant to shifts in relative power is due to Rowe (1989b). I first invoked his idea in my earliest studies of this topic (Koppel 1997; Koppel 1998a; Koppel 1998b; Koppel 1998c; Koppel 2000b).

will be allowed to shift from one actor to another when the anchors move due to random forces beyond any actor's control.

1. When an anchor moves due to random forces beyond the favored actor's control, he requests or takes a greater share of the asset governed by that contract;
2. As long as the disfavored actor is certain that the anchor moved due to random forces beyond the favored actor's control, he permits the favored actor to receive a greater share of the asset. This share is the favored actor's *random variable benefit*. It is variable because he only receives it periodically when the anchor moves in his favor due to random factors.<sup>2</sup>

**Assumption 2a.** Actors therefore need to establish and maintain institutions that have the capacity to distinguish between intentional and unintentional movements of the anchors. The primary function of these institutions is to detect intentional opportunism: cases where some actor moved an anchor in his own favor intentionally for his own gain.

**Assumption 2b.** The enforcement of an anchor contract requires a *justice bargain*. In this type of bargain, each side commits that whenever he is convicted of moving an anchor in his own favor *intentionally* for his own gain, he will accept the guilty verdict and pay the penalty voluntarily. In return for his compliance with this commitment, each side receives the variable benefit specified in the contract whenever an anchor moves in his favor due to *random forces beyond his control*. If he reneges on the commitment, however, then his variable benefit will be withheld. This threat enables the contract to be enforced without reliance on costly, power-based enforcement assets. That is, it enables *anchor-based* enforcement rather than *power-based* enforcement.<sup>3</sup>

**Assumption 3.** Anchor term contracts are incomplete contracts: they do not specify all possible contingencies that might arise in the future. When an unspecified contingency

---

<sup>2</sup>Most anchor contracts are explicit contracts. Some are implicit contracts, however. While they are not written or signed, the contracting parties still know the terms, consent to the terms, and adhere to the terms as long as adherence is reciprocated by the other contracting party.

<sup>3</sup>This enforcement mechanism is original in this study, as far as I am aware. I first characterized this enforcement mechanism in **Koppel11**.



arises, it is ambiguous....<sup>4</sup>

**Assumption 4a.** Every action has at least two effects: an intended effect and an unintended effect. Although the intended effect may not violate any contract or create any contractual ambiguity, the unintended effect may do so.<sup>5</sup>

**Assumption 4b.** When someone takes an action that has an unintended secondary effect on another actor, it may produce a contractual ambiguity that involves a problem of legal incompatibility. A legal incompatibility is a situation where one term of a contract or law gives one actor the legal right to continue his action, but another term gives the other actor the legal right to be free of the negative secondary effect of that action. The legal incompatibility causes a breakdown in common knowledge about the legal meaning of the contract terms and, therefore, a dispute.

**Assumption 4c.** When a conflict erupts due to a legal incompatibility, each side must destroy the value of the asset at stake in the dispute to insure the other side does not gain on net from the dispute, or else the other side will have an incentive to create similar disputes intentionally in the future.<sup>6</sup>

**Assumption 5.** In such a conflict, each side's strategy is to increase his relative power at the other side's expense on all available margins. These strategies induce an unrestricted competition in raising relative power. Each side invests resources in increasing his power up to the point that the marginal benefit from the last unit of investment is just equal to the marginal cost. In equilibrium, each side loses more than he gains from engaging in the

---

<sup>4</sup>The idea of an incomplete contract goes back at least to Klein, Crawford, and Alchian (1978). I first invoked this idea in my earliest studies of this topic (Koppel 1997; Koppel 1998a; Koppel 1998b; Koppel 1998c; Koppel 2000b).

<sup>5</sup>Koppel (2006a, p. 19).

<sup>6</sup>I first proposed assumptions 4a, 4b and 4c in my earlier studies of this topic (Koppel 1998b; Koppel 1998c; Koppel 1999d; Koppel 2000b; Koppel 2006a; Koppel 2006b; Koppel 2010b; Koppel 2002b).

competition: each side suffers a net loss.<sup>7</sup>

**Assumption 6.** To settle such a conflict in a stable way, each side must agree to a new anchor term contract. Thus there are two kinds of settlement corresponding to the two different ways that actors divide assets listed above (in assumption 1a).

- Settlements based on anchor term contracts will be stable, because no actor has an incentive to increase his relative power to shift the settlement terms in his favor.
- Settlements *not* based on anchor term contracts will *not* be stable, because each actor will have an incentive to increase his relative power to shift the settlement terms in his own favor. And this will induce competitions in raising relative power (as explained in assumption 1 above).

**Assumption 7.** To avoid the recurrence of conflict, it is not enough for actors to establish and maintain institutions that have the capacity to distinguish between intentional and unintentional movements of the anchors. Institutions must also have (1) the capacity to manage the contractual ambiguities and legal incompatibilities that arise when an anchor moves into the gaps between existing contract terms, and (2) the capacity to manage the disputes that arise when it is unclear whether such contingencies arose intentionally or unintentionally. Institutions manage such problems by reforming the existing contracts and forming new ones, and in particular, by increasing the dimensionality of the contracting space to resolve legal incompatibilities.<sup>8</sup>

---

<sup>7</sup>The idea of a wasteful competition in raising relative power is due to Rowe (1989b). I first cited Rowe's idea in my earlier studies of this topic. In the field of applied microeconomics, models of "cutthroat competition" and the empty core identify the conditions under which all firms suffer net losses from unrestricted price competition. I first applied this idea to international military competition in Koppel (1997), Koppel (1998a), and Koppel (1998b). I provide citations to the economic literature on these points in chapters 7, 21 and 22. Economic models of conflict assume that each side invests in a conflict technology until marginal benefit equals marginal cost. I first cited and applied these models in Koppel (2000b, 2002a, 2002b, 2009a, 2010a, 2010b). See citations to that economic literature in chapters 21 and 22.

<sup>8</sup>Many studies of international and domestic institutions assume they play an adjudicative role. However, I am not aware of any studies that define adjudication in terms of the concepts of anchor contracts and justice bargains (as I have done here). See M. R. Garfinkel, McBride, and S. Skaperdas (2012) for a model in which power-based deterrence and "governance based on norms" are substitutes rather than complements. Aside from their study and mine, I am not aware of any other studies that make this assumption. More generally, different subfields of political science study institutions for "contract management" under different names. See the "managerial" school of thought on international institutions, for example. CITE Chayes.

## 1.1 Main causal claims about conflict, settlement and long-run development in Europe

The overall argument of this study is composed of eight causal claims. The first four claims concern the oscillation from the world of law to the world of power. The second four claims, which are mirror images of the first four, concern the return oscillation from the world of power to the world of law.

### 1.1.1 Claims about the oscillation from the world of law to the world of power

**Claim 1a.** The absence or failure of anchor contracts causes costly competitions in raising relative power. Anchor contracts are *absent* when settlements are based on variable term contracts rather than fixed term (anchor) contracts. Anchor contracts *fail* when there are legal incompatibility disputes in an environment with a low capacity for Legal Incompatibility Management (LIM).

**Claim 1b.** The competition for power causes conflict. Conflict is simply a part of the competition for power – one of many ways that actors compete for power. While conflicts are tragic, analytically, the best way to understand the causation of conflict is to understand the causation of competitions in raising relative power. They are caused by the absence or failure of anchor contracts (claim 1a).

**Claim 1c.** The competition for power causes instability, misgovernance and social capital depletion. To compete for power, actors commit unilateral asset seizures – the definition of *instability*. Strong actors commit seizures against weak actors – the definition of *misgovernance*. Both patterns cause widespread insecurity and pervasive distrust – the definition of *social capital depletion*.

**Claim 1d.** The competition for power produces a characteristic structure of misgovernance that depends on the type of disputes that are most salient. When the salient disputes are over *political exclusion*, the equilibrium is a series of unstable balances of power between the opposing sides. When the salient disputes are over *economic exclusion*, the equilibrium is a stable preponderance of power on one side or the other.

### 1.1.2 Claims about the oscillation from the world of power to the world of law

**Claim 2a.** Anchor contracts stop the competitions in raising relative power. Because the contract terms are invariant to shifts in relative power, actors have less incentive to invest in power to shift the terms in their favor.

**Claim 2b.** Settlements based on anchor contracts can terminate conflicts in a stable way. By contrast, settlements based on variable term contracts can cause a temporary cessation of conflict. But the competition for power will continue and the settlement will be unstable.

**Claim 2c.** Anchor based governance is the main condition for stability, good governance and social capital formation. Anchor contracts generate *stability* in the distribution of assets among actors, so the distribution is invariant to shifts or differences in relative power. Strong actors respect the rights of weak actors – the definition of *good governance*. Both patterns generate security of asset ownership and trust that stronger actors will not take advantage of power disparities – the definition of *social capital formation*.

**Claim 2d.** Anchor based governance enables actors to avoid both structures of misgovernance – unstable balances of power and stable preponderances of power.

## 2. Observable implications

The main causal factor in the theory is the capacity for Legal Incompatibility Management, or LIM capacity.

### 2.1 Micro-effects of LIM capacity on disputing practices

**Implication 1.** When institutions **have** LIM capacity:

- 1.1 disputes over legal incompatibility problems will be tractable and reconcilable;
- 1.2 disputes will be settled by facts and law rather than relative power: the weak can win against the strong;

1.3 disputants will prefer nonpartisan judges.

**Implication 1'.** When institutions **lack** LIM capacity:

1.1' disputes over legal incompatibility problems will be intractable and irreconcilable;

1.2' disputes will be settled by relative power rather than facts and law: the strong win, the weak lose;

1.3' disputants will prefer partisan judges.

## 2.2 Macro-effects of LIM capacity on choice of enforcement method

**Implication 2.** When institutions **have** LIM capacity:

2.1 actors will be willing to pay for impartial adjudication services: the demand for adjudication services will be high;

2.2 investment in adjudication capacity will increase: institutions will become stronger in adjudication capacity;

2.3 actors will rely on anchor based enforcement rather than power based enforcement for assets that are potentially protected by either method; and

2.4 investment in power based enforcement capacity will decrease: institutions will become weaker in power based enforcement capacity.

**Implication 2'.** When institutions **lack** LIM capacity:

2.1' actors will not be willing to pay for impartial adjudication services: the demand for adjudication services will be low;

2.2' investment in adjudication capacity will decrease: institutions will become weaker in adjudication capacity;

2.3' actors will switch from anchor based enforcement to power based enforcement for assets that are potentially protected by either method; and

2.4' investment in power based enforcement capacity will increase: institutions will become

stronger in power based enforcement capacity.

## **2.3 Macro-effects of LIM capacity on political, economic and social institutions**

**Implication 3.** When institutions **have** LIM capacity:

3.1 the amount of instability and conflict will decrease;

3.2 strong actors will respect weak actors rights:

(a) in the domestic arena, corruption rates and tax rates on non-elites will be relatively low;

(b) in the international arena, strong states will respect the territorial boundaries and political independence of weak states;

3.3 there will be social capital formation.

**Implication 3'.** When institutions **lack** LIM capacity:

3.1' the amount of instability and conflict will increase;

3.2' strong actors will violate weak actors rights:

(a) in the domestic arena, corruption rates and tax rates on non-elites will be relatively high;

(b) in the international arena, strong states will violate the territorial boundaries and political independence of weak states;

3.3' there will be social capital depletion.

## 2.4 Effects of LI disputes on actors' wealth and power

**Implication 4'.** When legal incompatibility disputes become irreconcilable and actors switch from anchor based enforcement to power based enforcement, actors' **wealth and power** will vary with the dispute type:

4.1' In disputes over *economic* exclusionism:

- (a) the excluders will increase in wealth but decrease in enforcement productivity; (b) the excluders will decrease in wealth but increase in enforcement productivity;

4.2' In disputes over *political* exclusionism:

- (a) the excluders will decrease in wealth and decrease in enforcement productivity; (b) the excluders will increase in wealth and increase in enforcement productivity.

## 2.5 Effects of LI disputes on coalition financing strategy

**Implication 5'.** When legal incompatibility disputes become irreconcilable and actors switch from anchor based enforcement to power based enforcement, actors' **coalition financing strategies** will vary with the dispute type:

5.1' In disputes over *economic* exclusionism:

- (a) the excluders will employ compensation-based finance;
- (b) the excluders will employ coercion-based finance.

5.2' In disputes over *political* exclusionism:

- (a) the excluders will employ debt-based finance;
- (b) the excluders will employ self-finance through fiscal separatism.

## 2.6 Effects of LI disputes on coalition size and duration

**Implication 6'.** When legal incompatibility disputes become irreconcilable and actors switch from anchor based enforcement to power based enforcement, actors' **coalition size and duration** will vary with the dispute type:

6.1' In disputes over *economic* exclusionism:

(a) the excluders will expand their coalition as far as possible by compensating and coopting potential members of the excludees' coalition;

(b) the excludees will expand their coalition as far as possible by coercing members of the excluders' coalition into switching sides.

6.2' In disputes over *political* exclusionism:

(a) the excluders will expand their coalition just enough to force the excludees to capitulate, and then maintain their coalition by borrowing to pay third parties to coerce loyalty;

(b) the excludees, if left alone by the excluders, will separate and build no coalition; but if attacked by the excluders, the excludees will expand their coalition just enough to force the excluders to capitulate, and then disband the coalition.



## 2.7 Effects of LIM capacity on regime type

**Implication 7.** When institutions **have** LIM capacity:

- 2.2 institutions will become stronger in adjudication capacity; (implication 2.2)
- 3.2 strong actors will respect weak actors rights; (implication 3.2)
- 7 regime type will be cooperative and inclusionist.

**Implication 7'.** When institutions **lack** LIM capacity:

- 2.4' institutions will become stronger in power based enforcement capacity; (implication 2.3)
- 3.2' strong actors will violate weak actors rights; (implication 3.2)
- 7' regime type will be coercive and exclusionist, but vary with the type of dispute:

7.1' In disputes over *economic* exclusionism:

(a) the excluders will:

- (i) form a compensation-based coalition; (implication 5.1'a)
- (ii) expand it as far as possible by compensating and co-opting potential members of the excluders' coalition; (implication 6.1'a)
- (iii) thereby establishing a broad-based, multi-party, solvent, long-lived regime;

(b) the excluders will:

- (i) form a coercion-based coalition; (implication 5.1'b)
- (ii) expand it as far as possible by coercing members of the excluders' coalition into switching sides; (application 6.1'b)
- (iii) thereby establishing a broad-based, dominant-party, insolvent, short-lived regime.

7.2' In disputes over *political* exclusionism:

(a) the excluders will:

- (i) form a debt-based coalition; (implication 5.2'a)
- (ii) expand it just enough to force the excludees to capitulate, and then maintain it by borrowing to pay third parties to coerce loyalty; (implication 6.2'a)
- (iii) thereby establishing a narrowly-based, highly leveraged, insolvent, short-lived regime;

(b) the excludees will:

- (i) form a self-financed coalition through a confederation of fiscally separate jurisdictions; (implication 5.2'b)
- (ii) expand it just enough to force the excluders to capitulate, and then disband it; (implication 6.2'b)
- (iii) thereby establishing a number of small, confederated, solvent, long-lived regimes.

### 3. Roadmap to the empirical tests

The empirical cases on which I test the theory are displayed in the table on the following page. The cases divide into two main categories, depending on the actors' capacity for legal incompatibility management (LIM). In some cases the LIM capacity was *high*, while in other cases the LIM capacity was *low*. The cases with low LIM capacity divide into two further subcategories, depending on the type of disputes that caused competitions in raising relative power. In some cases the competitions were caused by disputes over *economic* exclusion, while in other cases the competitions were caused by disputes over *political* exclusion. The table includes cases at both the international and domestic levels, because the patterns of evidence on cause and effect were similar at both levels.

<b>LIM capacity</b>	<b>LOW</b>	<b>LOW</b>	<b>HIGH</b>
Enforcement method	power-based	power-based	anchor-based
Political outcome	instability	instability	stability
Economic outcome	misgovernance	misgovernance	good governance
Social outcome	social capital depletion	social capital depletion	social capital formation
<b>Main cause of disputes</b>	<b>political exclusion</b>	<b>economic exclusion</b>	<b>neither</b>
Swings in relative coalition size	frequent, narrow swings	infrequent, wide swings	neither
Resulting distribution of relative power	unstable balance	unipolar predominance	neither
Empirical cases in domestic politics: 1000-1500	Germany	Italy	England France
Empirical cases in domestic politics: 1500-1800	France	Netherlands	England Germany (Prussia)
Empirical cases in International politics: 1500-1950	Peace of Augsburg (1600-18) Thirty Years War (1618-48) Nine Years War (1689-97) War of Spanish Succession (1701-13) Peace of Utrecht (1713-45) French Rev. Wars (1792-99) World War I (1914-18)	Peace of Augsburg (1555-99) Napoleonic Wars (1800-14) Peace of Versailles (1919-39) World War II (1939-45)	Bretton Woods system (1945-) Peace of Westphalia (1648-89) Peace of Vienna (1815-53)

Part IV

The path of power and the path of  
law in domestic politics

## CHAPTER 13

### From the world of power to the world of law in Medieval England

In the 11th century England was a land of feudal anarchy and violence. By the 13th century it was significantly less violent and more legalistic. This long-run development was enabled by two main changes: A transition from fragile customary rights in land to formal legal rights in land and a switch from power-based enforcement to anchor-based enforcement. In the 11th century, a noble tenant had a customary right to his land, and while the right was usually respected it was not enforceable through the courts. Land disputes were commonly settled by feuds or trial methods that did not distinguish which claimant was in the right and which claimant was in the wrong. Royal sheriffs were often unable to enforce legal judgments against claimants who had lost at trial, especially when the losers were powerful magnates with private armies of their own. By the 13th century all of these elements of the land law had changed. Land disputes were settled by a judicial process based on facts and laws rather than relative power or superstition. A noble tenant's right to his land was enforceable through the courts. A tenant could even win in court against his own lord, a powerful magnate with a private army, requiring the magnate to return the tenant's land. In short, English society had transitioned from a world of power to a world of law.

#### 1. A puzzling comparison

The conventional wisdom holds that England had a strong state and Germany had a weak state in the medieval period. The main evidence for this claim is that private wars among the nobility ceased in England, but continued in Germany. The logic, presumably, is that the English state was strong enough to defeat the private army of any noble who engaged in

a private war, thereby deterring the nobility from waging wars. But the German state was too weak in military power to perform the same functions.

Much of the evidence from medieval England and Germany is puzzling from the conventional perspective, however. This suggests there must be some other explanation for the emergence of peace in England and continuation of war in Germany beside the relative strength of the two states. For example, the German king's army was larger than the English king's army. So the German king should have been more able to suppress private wars among the nobility, not less. Even after the German state disintegrated into regional statelets, each ruled by a regional magnate, the magnates' armies were larger than the English magnates' armies. So each German magnate should have been more able to suppress private wars among his noble vassals, not less. At the same time, the supposition that English kings and magnates enforced the law on wayward nobles through a preponderance of military power is refuted by the evidence. The English king and his sheriffs, each of whom was a regional magnate, were often unable to muster enough military power to punish a noble who had committed a legal violation or provoked a violent conflict against another noble.<sup>1</sup> The king even had to issue writs saying who *else* would try to enforce the law on the wayward noble if the sheriff couldn't or wouldn't do it. All of this evidence contradicts the conventional assumption that the English authorities had more military power than the German authorities had, and this is why peace emerged in England but war continued in Germany.

A defender of the conventional wisdom might object that the English monarchy had enough military power to defeat and deter domestic rivals for its own authority. So it must have been powerful enough to suppress private wars among the nobility by force. It is true that the English monarchy could muster a large enough army, when necessary, to defend its own authority from domestic rivals. This fact does not warrant any inference that peace among the English nobility was due to the monarchy's capacity to muster military power, however. The elementary economics of enforcement suggests the opposite. A state with enough military power to defend its own authority will often *lack* the military power to suppress violent conflicts among domestic non-state actors. For its military power depends not only on the size of its resource base, but also on its valuation of the asset to be divided

---

<sup>1</sup>Morris 1927. Hyams 2001.

by relative power. If it values the asset highly, then it will have an incentive to deploy a large amount of military power to secure its share of the asset. But if it has a low valuation of the asset, then it will only have an incentive to deploy a small amount of military power to secure its share. A state may value the security of its own domestic authority more highly than it values the security of domestic non-state actors vis-à-vis each other. Then it will deploy great military force to secure its own authority but little force to secure domestic non-state actors vis-à-vis each other. This principle of enforcement economics explains why the English state was strong enough to secure its own authority, but not strong enough to impose military punishments on nobles who committed legal violations or provoked conflicts against other nobles.

Second, the English monarchy's capacity to pacify the nobility was not based on military threats backed by a royal army larger than the magnates' armies. It was based on two non-power factors. In the 10th to 12th centuries, the monarchy regulated conflicts among the nobility so that the conflicts served legitimate enforcement purposes. In the late 12th century the monarchy began to improve the regulations to the point that they could prevent conflicts among the nobility. Neither of these two functions – regulating conflict or preventing conflict – required the monarchy to have an army that was larger than the magnates' armies. The anchor-based enforcement system employed by the monarchy did not depend on its relative power. To take an analogy, the umpire of a football game does not need a set of players of his own to throw onto one side or the other to punish violations by the opposing players and thereby deter further violations. His enforcement method does not rely on his relative power.

During the 10th and 11th centuries, the best that the English monarchy could do was to *regulate* conflicts among the nobility so that the conflicts served legitimate enforcement purposes.<sup>2</sup> While it was better at this role than any other monarchy in Europe – and some historians have labeled it a "strong" state for this reason – the monarchy could not muster predominant military power to enforce the law. Therefore, if the English state was "stronger" than other states at this time, this was not because the English king had a larger army than other kings had. It was because the English state had a better capacity to regulate conflicts

---

<sup>2</sup>Hyams 2001. Hudson 2010.



among the nobility to serve enforcement purposes. Or, the English state had a better capacity to improve these regulations later, in the 12th and 13th centuries, to the point that they prevented conflicts among the nobility entirely. And then historians projected this novel source of state "strength" back into the previous centuries anachronistically.

German kings and magnates in the 10th and 11th centuries also had the capacity to regulate conflicts among the nobility to serve enforcement purposes. However, they did not have the capacity to improve the regulations to the point that they would prevent conflicts entirely. They could not discern how to transform the ancient customary laws and trial procedures in the way that the English monarchy did after 1165. This difference between medieval Germany and England had nothing to do with military power.

In light of these patterns of evidence, the difference in long run outcomes – peace in England, war in Germany – cannot be explained by looking for differences in the military strength of the state. There must be some other explanation for it. In this chapter, I show how and why improvements in the English legal and judicial systems enabled a transition from power-based enforcement to anchor-based enforcement. In the next chapter, I show how and why the Germans were unable to improve their legal and judicial systems in a similar way, and as a result they were unable to transition from power-based enforcement to anchor-based enforcement.

## **2. Plan of the chapter**

What enabled the English to transition from power-based enforcement of ancient customary rights to anchor-based enforcement of formal law? The key reform was changes in the judicial process that enabled judges, attorneys and jurors to distinguish between intentional and unintentional causes of legal disputes.

Section 1 describes the conditions of government, law and society that the Normans inherited when they invaded in 1066. It was a world of power in which courts did not distinguish between intentional and unintentional disputes and land disputes were settled by battle according to relative power (or other methods that also produced unjust outcomes).

Section 2 describes the problems encountered by government and society as a consequence

of the inadequacies of court process in land disputes. The main problem was the incapacity of the ancient customary process to distinguish between questions of fact and questions of law.

Section 3 describes the innovations in court and trial process introduced by the king, central courts and local judges and sheriffs to solve the problems.

### **3. Power-based enforcement in England (900-1165)**

#### **3.1 Anglo-Saxon feud culture (900-1066)**

In a world where men stole land, rustled cattle, and committed assaults, the main purpose of feuding was to enforce rights to land, property and personal security. Feuds served this purpose in three ways. After wrongs were committed, victims took private vengeance on wrongdoers through feuds to deter further violations of their rights and maintain their reputations for defending their rights. In cases where it was unclear which side was right and which side was wrong, feuds served to maintain *each* side's deterrence and honor. Finally, victims used the *threat* of initiating a feud to compel wrongdoers to pay compensation rather than suffer a violent feud. In this world the main method of securing "justice" was feuding and threatening to feud.

After a wrong occurred, the victim "appealed" to his relatives and friends to join him in taking vengeance on the wrongdoer. Often the wrongdoer and his supporters fought back. Since both men's supporters were obligated to join them in their feuds, escalation and contagion were a natural part of the process. The initiation and reciprocation of violence took varied forms. Interpersonal violence in small numbers led to brawls, injuries and homicides. Interpersonal violence in large numbers led to battles and private wars. Violence against land and livestock led to mutual "ravaging". Violence against infrastructure led to mutual vandalism and arson. The feuding process included not only the acts of violence but the threats to commit violence. Men threatened violence to extort concessions of land or property from their opponents. When a man lost land or property in either way – by violence or extortionate threats – he took vengeance in return and extended the feud all the same.

Feuding was costly. It was also damaging to a lord's authority and reputation. If his

vassals feuded with each other too much, it was a sign that he had failed to protect them.<sup>3</sup> Every lord had a natural interest in preventing feuds among his noble vassals or limiting them at the least. He could limit a feud's destructiveness by persuading relatives and friends of the antagonists to stay out of it. He could prevent a feud by persuading the wrongdoer and the victim to reach a settlement involving compensation. He might choose one side to compel the other side to settle peacefully. There were several reasons to avoid choosing sides though. Above all, the lord might not have enough military power of his own to tip the balance of power decisively. Even if he did, his entry on one man's side would require him to enter a state of hatred toward the other man – an undesirable move since both men were members of his own retinue.<sup>4</sup> The lord's best option for preventing or limiting a feud, therefore, was to gather his vassals who remained neutral in the matter and solicit proposals for settlement terms they could put to the claimants.<sup>5</sup>

The convention of assembling the vassals, arbitrating the claims, and bidding the opponents to settle and reconcile came to be known as the lord's court. After a wrong occurred, the accuser and the accused were required to submit to the court process and give peace a chance before initiating hostilities. The court process was not a *substitute* for the ancient choice between private settlement and private vengeance through violence, however. It was a *means* of making this choice.<sup>6</sup> There were three possible outcomes to the process. In one outcome, no settlement was reached and the opposing groups proceeded to feud. In a second outcome, the opponents reached a settlement and avoided a feud. In the third outcome, they reached a settlement but the settlement broke down later, leading to a feud. The historical evidence shows that each equilibrium occurred at times. Thus, although violent feud and

---

<sup>3</sup>Hyams 2001, pp. 40–1.

<sup>4</sup>"To persuade a great man to hate as his man hates...will be difficult when the object of hatred is another follower; no lord can easily take over the feud for one of his men against another." Hyams 2001, p. 41.

<sup>5</sup>"The lord and the rest of his men then try to make peace on the basis of their own view of distributive justice, giving to each what they are felt to deserve in order to restore equilibrium within the lordship." Hyams 2001, p. 41.

<sup>6</sup>"Thegns and others viewed the public courts and their justice instrumentally as an extension of a feud based system of private redress for personal wrongs." (36) "Men relied in the first instance on their own strength and the power of their connections to seek redress of wrongs. When they eschewed direct self-help, as in open feud, they still followed much the same logic and patterns. They entered the courts in search of the same vengeance or an honorable settlement that they and their ancestors had expected to achieve through the prosecution of feud." (40) Hyams 2001, pp. 36, 40.

peaceful settlement were alternative outcomes, they were part of a single enforcement system. Men took account of the threat of a feud in decisions to settle. And they took account of settlement offers in decisions to feud. The court process was simply a continuation of the appealing and feuding process with more consideration of right and wrong according to the ancient customary laws.

The decision whether to feud or settle was driven by both relative power and legal factors. The balance of right and wrong according to the customs mattered. The customs were unwritten, however, and the court process for making judgments was not based on facts or distinctions between questions of fact and questions of law. Hence a major part of the decision calculus was relative military power. It was the relative power of the wrongdoers and victims themselves that mattered the most, not the lord's power to impose punishments. He often did not have enough power to do that. The judgments and settlements reached at court were enforced by the military capacities of the victims, the wrongdoers and their supporters to wage feuds. In short, the court was a third-party *adjudicator*, but not a third-party *enforcer*.

Despite the availability of court process, violent feuds were a common occurrence in this culture.<sup>7</sup> Hypothetically, feuding could have been eliminated from the system entirely by meeting the conditions for a deterrence equilibrium based on the relative power of the wrongdoers and victims. The conditions were not met, however. Neither the balance of power nor the balance of right and wrong was stable. The balances were often unknowable. The balance of power was shifting and unknowable at times, because the number of supporters that each man could muster might vary for myriad reasons. The balance of right and wrong was subjective, shifting and unknowable at times due to the archaic nature of the customs and court processes. It was often impossible, in particular, for court participants to distinguish between intentional and unintentional wrongs. In some of these cases, the victim claimed the wrong was intentional and demanded high compensation backed by the threat of severe punishment, but the accused claimed it was unintentional and offered only low compensation backed by the threat of stiff resistance. A peaceful settlement was unlikely in these cases. The resulting feuds might have the positive effect of deterring men from

---

<sup>7</sup>Hyams 2001. Hudson.

committing wrongs intentionally. Men still committed wrongs unintentionally, however. If they were unable to prove it was unintentional to the satisfaction of the victim and his supporters, the outcome was a feud. This scenario caused disputes over wrongful accusation – a major problem that lords strove to solve because it caused or prolonged feuding. Despite their efforts, the problem of wrongful accusation remained intractable as long as the courts lacked the capacity to distinguish between intentional and unintentional wrongs.

In sum, each lord faced a dilemma. On one hand, feuding was necessary to deter men from committing intentional wrongs. On the other hand, feuding was costly and damaging to the lord's authority and reputation. The lord did not have overwhelming power to punish wrongdoers or an effective judicial method for distributing liability between the accuser and the accused. The best he could do was to regulate the feuding system to control its worst excesses. It was difficult to regulate the feuding system, however. For when a feud arose, each combatant had relatives and friends who were obligated to support him in his feuds. So escalation and contagion were the norm. This increased the cost of feuding and the incentive for lords to regulate or prevent feuds.

The king was a lord like any other in these respects. He had the same interest in preventing and limiting feuds among his vassals.<sup>8</sup> He had the same limited options for achieving it, because in the end men needed to feud and threaten violence to enforce their rights. The king was different in a few respects. He had a bit more military power than other lords. His obligation to regulate feuding was formalized in a coronation oath. And his regulatory aspirations and exhortations were formalized in royal "laws".<sup>9</sup> These differences did not make that much difference in practice though. The king's military power was just as inadequate to prevent a feud by choosing a side as the next great lord's power was. This problem was particularly acute in his dealings with his most powerful vassals. Since he relied on them for his authority, he could not afford to alienate any of them by choosing sides in their feuds. Knowing this, they maneuvered to escape or reduce responsibility for their wrongs by finding enough allies to make the king's power inadequate for coercive enforcement.<sup>10</sup> Royal

---

<sup>8</sup>Hyams 2001, pp. 40–1.

<sup>9</sup>Hyams 2001, p. 36.

<sup>10</sup>Hyams 2001, pp. 34–6.

”justice” was unequal as a result. It was driven by relative power factors – both the king’s lack of overwhelming power and his subjects’ disparities in power. The king was simply not powerful enough in military terms to impose a royal monopoly on the use of violence or the maintenance of armed forces.<sup>11</sup>

### **3.2 Continuities in feud culture in the first century of Anglo-Norman rule (1066-1165)**

When the Normans invaded in 1066, they inherited this system of government and justice. The courts continued to use the ancient trial procedures for much of the next century. The first decades after the invasion were quite unsettled in many regions, because the violent conquest of landed estates continued to play out at the local level, and it took time for hierarchies of power to become clear.<sup>12</sup> But by the reign of Henry I (1100-1135), the royal government was respected throughout the land for its effective justice backed by strong sheriffs.<sup>13</sup> The impression of good justice may have been due to Henry’s modest improvements in the writ system<sup>14</sup>, his sheriffs’ judicious choices of which cases to consider and enforce, or simply the settling down of the post-conquest chaos. The writ system of Henry’s reign continued to suffer from serious liabilities, however. Before analyzing these problems in detail, consider the underlying rules for dividing land among men and the types of disputes they caused.

### **3.3 Land division rules, exceptions and disputes**

#### **3.3.1 Land was divided among actors according to anchor contracts (but there was no anchor-based enforcement system)**

Land was apportioned in discrete parcels. The parcels changed hands from one holder to another in a number of ways: grants, transfers, inheritance, court settlements and the use of

---

<sup>11</sup>Hyams 2001, p. 3.

<sup>12</sup>Dalton 1994.

<sup>13</sup>Green 1986.

<sup>14</sup>Green 1986.

force. When a parcel of land changed hands, its borderlines remained intact. The borderlines did not change in response to marginal shifts in relative power between one landholder and another. If a landholder's power increased by enough to seize his neighbor's parcel in its entirety, then the parcel might change hands. But smaller shifts in relative power between one tenant and another did not cause marginal shifts in the borderlines between holdings. In this sense, land was divided among its holders by borderline contracts whose terms did not shift with changes in relative power. So titles to land were anchor contracts in the sense defined in this study (**background condition 1**).

### 3.3.2 The land laws had exceptions that occurred randomly – due to forces beyond anyone's control

The main rules governing land ownership had exceptions that occurred more or less randomly (**background condition 2**). For example, when a landholder died the land usually passed to his closest relative. Yet there were legitimate exceptions to this rule which called for the land to pass to the dead landholder's lord instead. These exceptions occurred more or less randomly from the vantage of both the lord and the closest relative. Consider some examples.

- If the landholder died without close relatives, the land escheated to the lord.
- If the original grant had been only for the lifetime of the deceased landholder, then the land returned to the lord.
- If the original grant had been made in the wrong way, then the land returned to the lord.
- If the heir's claim to receive the land was invalid for some reason, then the land returned to the lord.

These exceptions were frequent enough that the lord stood to benefit from a justice system that would authorize his windfall in such circumstances.<sup>15</sup> The key question in each instance was whether the exception was truly warranted. The answer usually turned on a *fact*. The

---

<sup>15</sup>Inheritance was not the only area in which the rules had exceptions that occurred randomly and sometimes provided the landlord with a windfall. The other main rules that governed landownership also had exceptions of this kind.

adjudication system of the early 12th century was not based primarily on facts, however. Hence disputes arose.

### 3.3.3 Sources of land disputes and tasks of the judicial system

Disputes over land arose from four main sources, each due ultimately to the fact that the land laws had exceptions. Each of the four sources of dispute created a task for the judicial system to accomplish if disputes were to be settled justly.

First, it was not always clear whether the grounds for a particular exception really existed, and therefore whether the exception was really warranted. The first task of the judicial system, therefore, was to determine whether the grounds for an exception really existed by gathering the relevant facts. Second, even when the grounds for an exception existed, it was not always clear whether the grounds had arisen randomly or been created intentionally by a landholder in opportunistic pursuit of another man's land. The second judicial task was thus to determine whether the grounds for an exception had arisen intentionally or unintentionally. Third, random factors beyond anyone's control generated gaps and incompatibilities in the land laws that provoked disputes (**background condition 3**). The third task of the judicial system was thus to fill in the gaps and resolve the incompatibilities by defining the grounds for exceptions more clearly. Fourth, opportunists took advantage of the gaps and incompatibilities in the land laws for their own gain, again provoking disputes. The fourth judicial task was to determine whether such disputes had arisen intentionally or unintentionally – and to prohibit the actions that opportunists were taking intentionally that provoked them.

The first and second tasks required the courts to have the full facts of cases that came before them. The third and fourth tasks required the court to develop means of distinguishing between questions of fact and questions of law, and means of preserving case settlements for later use as precedents in settling questions of law. In the first century after 1066, however, the adjudication system was lacking in these capacities. The resulting enforcement problems that inspired reforms in all three areas. The next section considers the incapacities and enforcement problems. The following section considers the reforms and shows that they created the key conditions for anchor-based enforcement to be feasible and effective.



### 3.4 Liabilities of the ancient customary law and trial procedures

#### 3.4.1 Verdicts were not based on facts

The trial process for resolving land claims did not routinely look into the facts of a case. Although the process did allow some of the facts to be discovered, it ultimately prevented the settlement of disputes based on facts.<sup>16</sup> The role of the king, sheriffs and jurors in the process was simply to listen to the men who claimed their land had been seized without cause, render the judgment, and execute it. Once judgment was reached by the ancient trial methods – oath, ordeal or battle – the sheriff executed the judgment by awarding the land to the winning litigant.<sup>17</sup> If the loser was currently in possession of the land, the king issued a writ saying that the loser was ordered to return the land to the winner or else the local sheriff would seize it and return it to the winner.<sup>18</sup>

---

<sup>16</sup>Baker 2002: 4-6. "The procedure in contentious matters was calculated to avoid reasoned decision-making. If the parties could not be persuaded to make a 'love-day' – to settle amicably – then resort could be had to proof by oath, which might have to be backed up by a physical test. In order to put the defendant to this hazard, the plaintiff was required to establish a *prima facie* case. In this he would be supported by his 'suit' (*secta*), the group of followers whom he brought with him, and in a complex case there might be a detailed factual enquiry at this stage. The suit had some affinity with witnesses, and they may have been subject to examination as to competence, but their testimony was only part of the interlocutory process and did not dispose of the matter. If the defendant was allowed the benefit of proof by oath, he proceeded to swear on the holy evangels to the truth of his case, in very general terms and without possibility of cross examination. In the form of proof known to later generations as *wager of law*, he was expected to bring with him some neighbors as 'compurgators' or 'oath-helpers' to back up his word. If this lesser kind of proof was deemed inappropriate, usually because of the gravity of an accusation and the unreliability of the party's word, the oath might have to be proved by the physical test of an ordeal...."

"[Although the ordeal was abandoned eventually, this] did not affect *wager of law*. . . . Proof by simple oath and compurgation therefore survived, and passed into the common law [that emerged in the 12th century and after]. But it worked on the same inscrutable system as the ordeal. There was no question of going behind it into the facts of the case, let alone of having to weigh whether the oath was true, since that was settled conclusively by the compurgators. *The oath, like the ordeal, was intended to obviate a human decision on the factual merits of the case.* This is commonly summarized by saying that, under this old system, judgment preceded proof: once it was adjudged that one of the parties should swear or perform a test there was no further decision to make, except whether he had passed it.... [T]he proof [by oath] was the end rather than the beginning of the process of resolution [of a dispute].... Rights and liabilities could not be worked out in detail, because law stopped short at what we now call procedure." Emphasis added.

<sup>17</sup>Baker 2002: 12-13; 66. "Execution of a judgment was...obtained by judicial writs addressed to the sheriff. . . . In real actions the judgment was usually to recover seisin [possession] of the land demanded, and [in] the writ of execution. . . the operative words were 'cause A to have seisin'."

<sup>18</sup>Baker 2002: 54. "When an executive writ commanded a final result, it was indeed sometimes intended to execute a judgment already given rather than to initiate proceedings. But where a writ ordered the recipient do something 'justly', this was taken to mean that it should be done only if its justness was established by due enquiry [according to the customary trial procedure described above]." (54)

### 3.4.2 Trials did not distinguish between questions of fact and questions of law

The law enforced by the king and his sheriffs was not royal law, but local custom. The royal writs instructed the sheriff to "do right" according to the local custom. By its nature custom did not distinguish between fact and law. The role of custom was simply to state the correct or "normative" behavior in each circumstance. For example:

"In circumstance  $C1$  we do behavior  $B1$ . In circumstance  $C2$  we do behavior  $B2$ ."

The custom did not distinguish formally between the justifying circumstance  $C1$  and the normative behavior  $B1$ . Each custom was stated simply as an amalgam:  $C1|B1$ . That is, each custom stated the correct behavior and the circumstances under which it was correct without distinguishing the behavior from the circumstance in any sense that a trial judge or jurors could recognize.

This feature of the customary laws did not prevent disputes from being resolved successfully *as long as the facts of the case were perfectly clear*. For example, if the facts boiled down to either  $C1|B1$  or  $C2|B2$ , then the verdict was "not guilty". And if the facts boiled down to  $C1|B2$ , then the verdict was "guilty". But if the facts were ambiguous, then disputes would remain contentious. For no matter which of the two verdicts was issued, one of the two opposing litigants would be unsatisfied with it.

The problem is illustrated by a classic example from the early common law. If a landholder had an heir, then he would inherit the land after the landholder's death. In this case the dead landholder's lord had the right to collect a heavy inheritance tax from the heir when he took ownership of the land ( $C1|B1$ ). But if the landholder had no heirs, then he could grant his land to a church in perpetuity ( $C2|B2$ ). In this case the lord lost his right to collect the inheritance tax forever, because a church would never die, so its land would never be inherited by any heirs. As long as landholders who had heirs never granted their lands to churches, there was no problem. But some landowners with heirs did grant their lands to churches simply to deprive their lords of the inheritance taxes. The landholder would grant his land to a church on the condition that the church would lease it back to the landholder and his heirs on the same terms they previously held it from their lord. This

type of grant deprived the lord of his inheritance taxes while allowing the landholder and his heirs to retain possession and use of the land as though they still owned it.

### **3.4.3 Courts did not preserve case precedents in writing**

The customs did not distinguish between facts and laws. And the ancient trial procedures did not include any way of distinguishing between questions of fact and questions of law. So ambiguous cases could not be settled by answering questions of law in a way that distributed rights and obligations clearly and justly. This problem was exacerbated by the lack of any system for preserving case precedents in writing.

”[I]n the absence of any centralization to ensure uniformity from place to place, or records to ensure consistency over time, the old way of doing things could never have generated a body of law comparable with that of ancient Rome or Serjeant Catesby’s England. *Rights and liabilities could not be worked out in detail*, because law stopped short at what we now call procedure.”<sup>19</sup>

In ambiguous cases the distribution of rights and obligations between the opposing parties was likely to be disputed. If the case was settled without working through the ambiguity justly, the losing litigant might resist the adverse verdict. At a time when landowners maintained private armies, this possibility posed problems for sheriffs.

### **3.4.4 Trials did not distinguish between intentional and unintentional sources of disputes**

To draw this distinction, the judicial process would need three capacities, but it lacked all three. First, trials would need to consider the key facts that proved intentionality (or lack of intentionality). The pre-trial process was not designed to generate the relevant facts, however, nor was the trial process designed to consider them. Second, trials would need to distinguish between questions of fact and questions of law to identify disputes that had arisen unintentionally – due to gaps or contradictions in the law. The trial procedure was not designed to make this distinction, however. Third, courts would need to preserve case

---

<sup>19</sup>Baker 2002: 4-6. Emphasis added.

precedents in writing to ensure that when new legal rules were formed to remedy gaps or contradictions in the law, they would be applied consistently after that. Courts did not preserve case precedents in writing, however. As a result of these three failings, trials did not distinguish between intentional and unintentional sources of disputes (**causal factor 2'**). Without this distinction anchor-based enforcement was not feasible, so actors could only use power-based enforcement (**observable implication**).

### 3.5 Enforcement depended on relative power – and varied with it

The effectiveness of the writ system depended on the sheriff's capacity and willingness to carry out the writs, as well as his costs and risks of doing so. He could not claim expenses (and in this sense he was not a royal employee at all). And he was liable for damages if he made mistakes. So "the temptation for him to do nothing was considerable."<sup>20</sup> The temptation to do nothing might also arise because the losing litigant might reject the adverse verdict. In such cases the winning litigant (and his retinue) would call on the king or the sheriff (and his retinue) to ride with them to enforce the verdict by arms.<sup>21</sup> If they prevailed, it redounded to the king's good reputation as a powerful provider of justice. They did not always prevail, however, and sometimes they chose not to ride in the first place in the expectation of defeat. The more powerful the losing litigant was – in an era when landowners maintained private armies – the less likely was the sheriff to be able to defeat him by force. In a number of cases, neither the sheriff nor any other magnate in the region could enforce the court's judgment on the losing litigant.<sup>22</sup> In response to this problem, the king issued writs which acknowledged that the sheriff might not have the capacity to enforce the judgment and listed other magnates who might try to enforce it instead,<sup>23</sup> if the temptation to do nothing did not paralyze them too. Often enough, these orders proved insufficient too.

---

<sup>20</sup>Baker 2002: 65. "The sheriff could return to a [writ of] *distringas* [a report] that he could find nothing to distrain, or to a [writ of] *capias* [a report] that the defendant was ill or not to be found. Whether he had looked was a question one was not allowed to ask, and so these returns became common fictions for use by under-sheriffs who could not be persuaded to take positive action or who hoped by delay to prise some gift from the plaintiff."

<sup>21</sup>Hyams 2001.

<sup>22</sup>Morris 1927.

<sup>23</sup>Morris 1927.

Thus, the sheriffs of the early 12th century were not paramount enforcers of the law. Although they succeeded in enforcing judgments in many cases – thereby lending the government a reputation for effective justice – there were many other cases in which they did nothing to enforce it or failed in their enforcement effort. So the king’s reputation for providing good justice was in part a product of the sheriff’s selection of cases to consider and enforce. The variability of enforcement presents a puzzle. It may have varied simply because the losing litigants varied in military power: The sheriffs could defeat the weak losers but not the strong losers. Alternatively, the sheriffs may have varied in their enforcement success because some of the losing litigants felt the verdict was unjust in their case. Those losers would have resisted the sheriffs more forcefully than the other losers who acknowledged that the verdict in their case was just. In any case, England in the first century of Norman rule was still a place where justice flowed from the capacity of the king and the sheriffs to commit violence against losing litigants, rather than the effectiveness of the court system. The king and the sheriffs often lacked the capacity to commit *enough* violence against losing litigants, however, to deter them from resistance and bring them to justice peacefully. So injustice and violence still occurred regularly. The writ system inherited from the Anglo-Saxons was insufficient to establish justice.

#### **4. Reforms that enabled a switch from power-based enforcement to anchor-based enforcement**

Three sets of reforms were implemented in the second century of Norman rule (1165-1265). These reforms improved the system’s capacities to supply facts to courts, to identify and resolve questions of law, and to preserve the most important of these resolutions in writing for future use. The key effect of these reforms was to enable the judicial process to distinguish between intentional and unintentional sources of disputes, thereby enabling actors to switch from power-based enforcement to anchor-based enforcement.

## **4.1 Reforms that improved the system's capacity to supply fact to courts**

### **4.1.1 Writs offering defendants the option to claim an exception**

The most basic problem in land disputes was that the defendant sometimes refused to surrender the land to the plaintiff even after the king ordered him to do so. To solve this problem, the king began to issue writs that ordered the defendant *either* to return the land to the plaintiff *or* to come to the local court and explain to a sheriff or royal judge why he should not have to return it. In providing this option, the king and his counselors may have been responding to complaints from the shires that judgments in the local courts were not always just. Whatever the motivation, the new procedure produced a fundamental change in the nature of justice provision by the king, his judges and local officials.

Under the old approach, the king's role was simply to have his sherriff inflict penalties by arms on the defendant to secure compliance with the plaintiff's demand to have his land back. The sheriff's role was merely to preside over a trial according to the ancient customary method – which "was intended to obviate a human decision on the factual merits of the case"<sup>24</sup> – and then use force against the losing litigant to secure his compliance with the judgment. Under the new approach, the king's role was to have his sheriff adjudicate a legal dispute between plaintiff and defendant. The sheriff's role was to conduct the adjudication on the basis of facts. The emphasis on facts embedded within an adversarial procedure set in train a number of major developments.

### **4.1.2 Writs ordering judges to inquire whether the facts warranted the exception**

The court ruling often depended on whether an exception to an existing rule was warranted – and the verdict turned on a fact. Was it really *true* that the previous tenant had died without close relatives? Was it really *true* that the original grant to the deceased tenant had been for his lifetime only? To enable such determinations to be made, the king and the central courts developed a system of royal writs that provided an impetus and a vehicle for improvements in fact gathering techniques.

---

<sup>24</sup>Baker 2002: 5.

The king started to issue writs that instructed the local sheriff or judge to look more deeply into the facts of the case, render an appropriate verdict, and *then* enforce the verdict.<sup>25</sup> As a result, more facts were taken into account in court proceedings. Litigants then anticipated the use of facts at court and began to bolster their initial claims with facts documented in writing. At an even earlier stage, landholders began to preserve evidence of their land titles upon first gaining a piece of land. They recorded land grants and transfers in writing and drafted these documents in increasing detail. As cases became more differentiated in type, the king issued new writs that delineated which facts mattered in which ways in each type.

The new writ procedures increased the effectiveness of the adjudication system at supplying facts to courts (**causal factor 1**). The system thus enabled valid claims by honest plaintiffs to be proven so and false claims by opportunistic plaintiffs to be proven so. Now defendants who were in the right would be found so and left to enjoy their land in peace rather than being harassed by sheriffs loosed on them by the false claims of opportunistic plaintiffs. If sheriffs' orders were obeyed in this system, then it was more likely because the orders were just than because the sheriffs were powerful enough to impose all orders on all defendants – even those who lost at trial but felt they were in the right. At the same time, however, the increased supply of facts to courts made it more difficult for the local sheriffs and judges to do their jobs.

#### 4.1.3 The right of appeal to redress factual error or judicial bias

The new writ procedures increased the frequency of disputes in which the local judge or sheriff's verdict was faulty. Local judges were not perfect. While a verdict at the local level might be correct, it might instead be the result of factual oversight, error or bias by the local judge or sheriff. When a disputant lost at the local level, therefore, he often appealed to the king to reconsider his case. In response, the king instituted a system by which royal judges traveled to the shires regularly to preside over trials. These developments further increased the capacity of local courts to distinguish between valid claims lodged by honest plaintiffs and false claims lodged by opportunistic plaintiffs, thereby taking another step toward solving

---

<sup>25</sup>In some cases, the writ instructed the local sheriff or judge to report the facts back to a royal justice.

the root cause of the enforcement problem – stiffer resistance to court judgments by litigants who had lost at trial but felt they were in the right.<sup>26</sup> As the volume of appeals to the king's courts rose, however, so did the frequency of cases that were not covered clearly by the existing customs and laws, due to gaps, ambiguities and incompatibilities in the laws.

## 4.2 Reforms that improved the system's capacity to identify and resolve questions of law

### 4.2.1 The problem of questions of law

Questions of law were raised by atypical cases, ones that did not fit neatly into the existing categories of law. Atypical cases arose for two related reasons. First, in a changing society, actors were constantly discerning new opportunities to take new actions not covered by existing laws. Second, while the intentions behind these actions may have been benign, their unintended consequences were often to violate the rights of others (**background condition 3**). A major source of atypical cases was inheritance disputes. Lords often took actions motivated by intentions that had nothing to do with their tenants, yet whose unintended effect was to violate their tenants' property rights. When a tenant died and his closest relative was owed the land as the rightful heir, the lord might have understandable reasons for choosing someone else instead. In each case the lord's conscious intention was not to disinherit the relative but to achieve some other goal that had nothing to do with the relative – yet the unintended effect was to disinherit the relative nonetheless.

- A lord might wish to reward another vassal for his loyal military service.
- A lord might wish to honor the deceased man's nomination of a different relative from a second marriage, simply to preserve a reputation for good lordship in the eyes of his other vassals.
- A lord might wish to grant the land to someone else simply to secure his own political advantage vis-à-vis other lords or the king.
- A lord might observe that the deceased man's closest relative was such a distant one

---

<sup>26</sup>Land disputes between the king's tenants in chief were settled in the royal court. Land disputes between other nobles could also reach the royal court on appeal from the shire courts. Royal judges heard cases both in the central royal court and in the shire courts as emissaries from the central court. Sheriffs heard cases in the shire courts as well.



that no rule of succession existed in the case; and then the lord might grant the land away as he wished simply to suit his own interests.

- A lord might wish to grant the land to someone else who was more fit for military service than the closest relative was.

In each case the lord's intention was understandable – and it was not to disinherit anyone – yet the unintended effect was to disinherit someone.<sup>27</sup> Inheritance was not the only area in which lords acted with understandable intentions, yet with the unintended effect of violating tenants' property rights. The other rules protecting landownership also suffered from this problem.

The central problem in atypical cases was that a jury or judge could not reach a definitive verdict simply by answering a question of fact – as in the cases mentioned earlier where the only question was whether an exception was warranted. Now there were also questions of law to be answered. Should the lord have the right to pass over the closest relative for one of the reasons listed above? Or should the closest relative have his deceased ancestor's land? Local verdicts by local fact-finding juries were often insufficient to settle such cases to the mutual satisfaction of both sides. Many of these cases were appealed to the royal courts.

The rising volume of appeals created a dilemma for the royal court judges. The more society changed, the more new questions of law arose in disputes, and the more new answers the royal judges would have to provide. They quickly perceived that if every new question of law were resolved by the production of a new rule or an exception to an existing rule, the judicial system would become overwhelmed. The decision costs, measured by the time and training needed to hear each new type of case and deliberate over each possible exception or novel rule, would mount exponentially. And the mounting number of new rules on the books would multiply the potential for inconsistencies between the rules and unpredictability in the verdicts expected by litigants.

"It was better, said the judges, to suffer a 'mischief' [hardship] in an individual case than the 'inconvenience' [inconsistency or unpredictability] which would follow from admitting exceptions to general rules."<sup>28</sup>

---

<sup>27</sup>Hudson 1994.

<sup>28</sup>Baker 1994: 80.

At the same time, however, the judges could not leave questions of law unresolved, as that would permit disputes to go unsettled and allow disgruntled litigants to take matters into their own hands. The dilemma, therefore, was that the judges need to be shielded from having to answer questions of law, yet all the questions of law needed to be answered at trial in some way. In response to this dilemma, the royal courts developed a new way of conceptualizing lawsuits and a new way of conducting trials.

#### **4.2.2 The forms of action**

To make the court process tractable, judges and court officials developed a simple classification system for the accusations that plaintiffs made against defendants. Complaints were classified into different "forms of action" depending on the type of complaint it was. Each form of action was specified in precise language that clarified exactly where the line was between legal and illegal behavior, and exactly what line the plaintiff claimed the defendant had crossed to the point of warranting a suit. For example, there was a form of action for a tenant's complaint that he had been wrongly evicted from his land by his lord. There was another form of action for a landholder's complaint that a neighbor had trespassed on his land and caused economic damage. There were about 20 forms of action in total.

This system helped to shield the judges from having to answer questions of law by limiting the number of new dispute types that reached the courts. The system ruled out new dispute types in two ways. First, if a plaintiff could not couch his complaint in terms of one of the forms of action, then he had no right to make his complaint at all. Second, once the plaintiff chose a form of action, he was obligated to fit the facts of his complaint into the precise language of that form, or else he had no right to make his complaint at all. The courts could reject complaints on either ground – the plaintiff had failed to couch his complaint in terms of a form of action, or the plaintiff had failed to fit his facts into the language of the form he had chosen.

These rules show that the forms of action constituted a set of anchor contracts between the members of English society – and between the government and society. The forms were anchor contracts in that their terms were invariant to shifts in relative power between the actors. The set of forms and the language of each form remained the same from case

to case, regardless of the relative power of the litigants and regardless of their power or weakness relative to the government. Occasionally the central court officers created a new form of action to acknowledge a new way that members of society were harming each other's interests so that the victims could obtain relief at court. But these were the exceptions that proved the rule. The forms of action were more or less fixed and if a plaintiff could not find a form for his facts then he could not find relief at court.

Because the forms of action were anchor contracts, they help litigants to avoid competitions in raising their relative power to maneuver real-world events into gaps in the law to extract material concessions from each other. The result was a reduction in the number of new dispute types that reached the courts and forced judges to answer new questions of law.

### 4.2.3 The pleading rules

When a dispute reached a royal court for litigation, the opposing attorneys were not always in consensus on the issue that was actually to be tried. Yet the judge needed them to be in consensus on it. So they were required to follow strict rules to reach consensus on it. The purpose of the rules was to reduce each case to a single question of fact that could be put to the jury for settlement based on its knowledge of the case. There were both formal and informal pleading rules.

**Formal pleading rules.** The opposing attorneys were required to proceed through a sequence of alternating moves in which the options were limited at each move. The plaintiff's attorney moved first by couching the complaint in terms of a form of action. Then the defendant's attorney made his move in response, but he had only three options:

1. He could deny the facts as alleged by the plaintiff. This would reduce the case to a question of fact that could be put to the jury.
2. He could accept the facts, but deny that his client had committed any violation of law. This would produce a question of law.
3. He could accept the facts and admit that his client had violated the law, but allege that it was justified in self-defense because the plaintiff had violated *another* law. This would start the whole process over again in reverse with a new complaint couched in

terms of another form of action. In this case the original plaintiff's attorney would have the same three options in reverse in his new role as a defendant's attorney (for the original plaintiff).

Theoretically, this game of two moves could be repeated *ad infinitum*. In practice it was limited to four rounds.<sup>29</sup>

The defense attorney's three options served to identify the three problems that a trial needed to solve to reach a satisfactory verdict. The first option identified questions of fact. The second option identified questions of law. The third option identified problems of legal incompatibility.<sup>30</sup> By distinguishing between the problems, the pleading procedure enabled the attorneys to tackle each problem separately. It also enabled the judges and court officials to improve their methods of supporting the attorneys in their effort. And ultimately, it enabled the court system as a whole to solve the three problems in society that were propelling issues of fact, law and legal incompatibility into the courtrooms. Consider the three options and their positive effects in turn.

The first option identified questions of fact. If the opposing attorneys disagreed only about the facts, then the case could be put to the jury to settle the questions of fact. In practice, judges were reluctant to allow multiple questions of fact to be put to a jury, because that would multiply the possibilities for mixed verdicts, contradictory verdicts or inconclusive verdicts. Judges therefore pressured attorneys to boil down the evidence to a single question of fact on which a jury could issue a definitive ruling. To support the attorneys in that effort, the judges and court officials created an inquest procedure to uncover the relevant facts of each case. An inquest was not simply a matter of the king instructing the local sheriff to look into the facts (as mentioned above). Rather, the courts formulated new rules for empaneling juries of fact-finders, charging them appropriately for their task, and holding them accountable for the performance of that task. The emergence of the inquest procedure helped to solve the broader social problem of intentional violations of the law. As long as the

---

<sup>29</sup>Each move in each round had a different name, perhaps to enable the judge to enforce the rules correctly, or perhaps simply to enable him to keep track of where the attorneys were in the process in case they got lost in negotiations.

<sup>30</sup>Recall that a legal incompatibility is a situation in which one law gives one actor the right to an asset, but another law gives another actor the right to the same asset.

judicial system lacked the capacity to supply the relevant facts to the courts, law violators could get away with it. That problem was solved by the inquest procedure.

The second option identified questions of law. Judges were reluctant to settle questions of law themselves. They had an important reason for preferring that the opposing attorneys do it. Judges knew that if they settled the questions of law themselves, over time they would establish many new precedents, some would be inconsistent with others, and eventually this would paralyze the judicial system. By contrast, if the opposing attorneys settled the questions of law, then each case would be shoehorned into one of the *existing* precedents at trial, and few *new* precedents would need to be created to reach trial verdicts. Judges therefore pressured attorneys to resolve questions of law themselves. To support them, judges improve their capacity to serve as neutral arbiters and sounding boards. In the words of one legal historian, the judge's "role was more that of umpire and adviser" than decider of tough questions of law.<sup>31</sup>

The emergence of this division of labor between judges and attorneys helped to solve the broader social problem of gaps and ambiguities in the law that masked the difference between intentional and unintentional violations of the law. As long as the gaps and ambiguities remained unresolved, it could never be clear whether a violation was intentional or unintentional. For the unintentional violator could always claim legitimately that he didn't know his action was illegal. Yet the intentional violator – who exploited gaps and ambiguities in the law deliberately for his own gain – could make the same claim. Judges had to solve this problem without creating a stream of new precedents that would paralyze the judicial system. Their solution was to force attorneys to resolve questions of law informally between themselves.

The third option identified problems of legal incompatibility. If the defendant admitted he had violated the plaintiff's right, but claimed it was justified because the plaintiff had earlier violated his right according to a different law, this was not simply a question of gaps or ambiguities in the law. It was a problem of legal incompatibility. Judges were reluctant to solve legal incompatibility problems themselves. They had an important reason for preferring that the opposing attorney do it. Judges knew that if they solved legal incompatibility

---

<sup>31</sup>Baker 2002: 79

problems one way or the other, they would likely leave one of the opposing litigants in each trial feeling the verdict was illegitimate. In a warlord society where the disgruntled litigant could easily take his cause of action "out-of-doors" and start a private war against the other litigant, it was important not to leave litigants disgruntled with trial outcomes. By contrast, if the opposing attorneys settled questions of law through bilateral negotiation, then at worst, one litigant might be unhappy with how the *jury* had settled the *factual* questions. But neither litigant could fault the *judge* for how the questions of *law* had been settled, since their own attorneys had done that. Thus the judges' preference *not* to settle legal incompatibility problems themselves reinforced the legitimacy of the judicial system – a significant achievement at this early stage in its development. Judges therefore pressured attorneys to resolve legal incompatibility problems. In response to this pressure attorneys developed a system of informal pleading rules to structure their bilateral negotiations.

**Informal pleading rules.** Whenever opposing attorneys negotiated over legal incompatibility problems, they had to discriminate between competing definitions of their clients' interests and decide which definitions mattered most to their clients. Since the stakes were usually high on both sides, each side had to deploy a combination of threats, concessions, and persuasion if the two sides were to reach consensus. They were more likely to reach consensus if each could persuade the other to see the matter a little differently. Yet the scope for persuasion was still limited by hard material interests. Over the long run, such bargaining led to the emergence of a characteristic *way* of doing legal business in the royal courts. It was an oral tradition that took practitioners years to learn – and the tradition evolved over decades and centuries. An attorney needed knowledge of this tradition to practice law, but could only gain it from deep immersion in the legal culture of informal pleading over years.

This culture shaped the way that disputes were settled in the royal courts and thereby solved a broader problem in society. When men conducted land transactions, it was not always clear which contract terms might engender disputes and how those disputes might be handled in the royal courts. This became clearer as the legal culture of the courts evolved and deepened. Then men were able to structure and document their land transactions in ways that reduced the likelihood of disputes arising from gaps, ambiguities and incompatibilities in the law.

## 5. Sources of the reforms' effectiveness

### 5.1 Sequential complementarities

The reforms of the royal writ system, mentioned above, were central to the increase in judicial effectiveness. Some of the main reforms were adopted in sequential fashion, because each reform solved one problem but created another problem, which then had to be solved by further reform. After the reforms were completed, they formed a *complementary set* in which each reform only worked because the other reforms were in place as well. This pattern is evident from a legal historian's description of how the process unfolded.

**Intention of the reforms.** "Henry II...chose to ensure that any bargains which lords made with their tenants [for land possession] would be binding.... Henry needed to see that any disputes which arose concerning such bargains were solved without disruption in an institutionalized fashion. He needed...to provide [legal] remedies to lords when tenants failed to perform their services, and [legal] remedies to tenants if lords failed to keep their part of the bargains.

**Problem 1.** "[T]he most blatant feudal deprivation would be displacing an able tenant for an abler or more-favoured one at the lord's will.

**Reform 1.** To prevent such events... Henry established the principle of *nemo tenetur respondere sine brevi*, that no man need answer for his free tenement without the king's writ. This would prevent the tenant being disseised without judgment.

**Problem 2.** But if the person the lord wanted [to put] in [possession of the land] was the better [skilled] knight, then the trial by battle which would decide the issue would in the end oust the sitting [less skilled] tenant.

**Reform 2.** To overcome this, Henry allowed the tenant a choice of how the issue was to be tried; Henry allowed the tenant to elect between [trial by] battle and the grand assize [a court procedure], taken in the king's court to try the issue of [which litigant had the] greater right [to the land according to the original contract terms].

**Problem 3.** Lords could, however, get around the problem of the grand assize by disseising the sitting [less skilled] tenant so that it was [this one] who was forced to demand the land from the new [more skilled] tenant, and the new tenant would then have the choice [of settlement method] and could elect for [trial by] battle [rather than the grand assize].

**Reform 3.** To overcome this device...Henry introduced the assize of novel disseisin which allowed the recently disseised [less skilled] tenant to be resealed of the land [by court order], and gave him back his right to elect [either trial by battle or the grand assize] if the [more skilled] disseisor subsequently brought an action against him for the land.

**Problem 4.** Finally, if the lord chose to put in [possession of the land] his own man after the tenant's death instead of seising [the tenant's] heir, novel disseisin would not be available, so the heir would again be at the mercy of the new tenant's election [of trial by battle rather than the grand assize].

**Reform 4.** To overcome this device...was one of the functions of the assize of mort d'ancestor, which protected the heir's right to be seised if his ancestor died seised as of fee and the claimant was his next heir.

**Net effect of sequential complementarities.** Throughout all this...the governing policy was not to protect possession or to control the customs of the lord's courts, but rather to ensure that the bargains struck between lords and tenants were honored. It was the honoring of such bargains that transformed precarious rights under feudal compact into secure rights in real property."<sup>32</sup>

Each problem arose because lords were exploiting a gap in the law to do an "end run" around the underlying intent of the reforms. The lords were superior in relative power to the tenants. Without legal protection the tenants would be dispossessed by the lords. By filling the gaps in the law, the monarchy gave the tenants legal protection that rendered their shortage of power irrelevant.

---

<sup>32</sup>Cambrian Law Review 45, 1979. I broke the quotation into paragraphs and attached the paragraph labels in boldface text just for clarity.



## 5.2 Anchor-based enforcement

What happened to a man who defied an adverse verdict in a land dispute? The king could seize his land or impose a fine on him. Some men resisted these penalties, however. In that case it would be costly for the king's agents to prevail over the man by force – especially if he was a large landholder having a strong power base in his region. A cheaper method of enforcement was available in these cases. The king could deny the man access to the royal court system, thereby precluding him from recovering parcels of his own land that he had previously let out to sub-tenants – even when recovery was warranted by one of the exceptions listed above. To deny the man access, the king simply took the man's lands "into his hand". This need not require actually seizing the man's lands physically. "After all, land could be in the king's hand theoretically and still be in the tenant's possession practically."<sup>33</sup> Indeed it was the king's tenants-in-chief, the most powerful lords in the kingdom, from whom the king was *least* likely to be able to seize their lands physically. Rather, to take the man's lands into his hand, the king simply rendered a judgement to this effect and notified the royal judges not to hear any claims brought by the man to recover lands previously let out to his own vassals. "Persons with claims to lands [taken into the king's hand] were unable to seek restoration from the holders [such as their own vassals], since one whose lands had been seized by the king could neither plead nor be impleaded in the royal courts."<sup>34</sup>

This system of enforcement was simultaneously cheaper and more effective than the use of force. It was cheaper because it conserved on physical enforcement costs precisely where these costs were the highest: The more land a lord held, the more powerful he was, and the more costly it would be for the king's agents to prevail over him by physical force. At the same time, the system was more effective than using force because its effect was greatest against the strongest lords. For the more land a lord held, the more sub-tenants he had. The more sub-tenants he had, in turn, the more frequently he could expect to recover parcels of land from his sub-tenants (under the exceptions listed above) as long as he retained access to the royal courts. By the same measure, though, the more sub-tenants he had, the higher the penalty he would pay (in lost recovery rights) from losing access to the courts.

---

<sup>33</sup>R. V. Turner 1968, pp. 65–6.

<sup>34</sup>Turner 1968: 65.

The king likely understood this calculus. In effect the king presided over a judicial economy. He had to *ration* his enforcement assets to maximize the level of compliance with his courts' verdicts. If physical enforcement against a powerful lord would be too costly, the king had to impose the alternative penalty of denying the lord access to the courts. To make the threat to deny access credible, in turn, he had to make access to the courts a *valuable* right. He had to insure that if a disputant prevailed in a land dispute at court, then he would receive the land he was owed – or some other land of equal value. That is, the king had to ration *patronage* to keep the justice bargain over which he presided in balance. The key to his role, then, was to reward patronage to those subjects who complied with court verdicts but deny patronage to those subjects who defied court verdicts.

The king had two ways to deny patronage. He could deny a man access to the law courts that judged when it was warranted. Or he could deny a man access to his "grace" by labeling the man as having "fallen from grace".

## 6. Micro effects of the reforms

### 6.1 Disputes were settled by facts rather than power and conflict

Altogether these practices created a growing momentum in transforming the legal process from one based on relative power to one based on factual evidence. The more clear it became that cases would be decided on the basis of facts, the more effort each player in the process exerted to produce a factual foundation for deciding cases. As a result, the royal courts became an arena where a disputant could prevail based on factual evidence even if he was physically weaker than the opponent (**implication 1f**). X evidence.

### 6.2 All actors developed a preference for neutral judges

Men at all ranks of English society recognized the value of this system, because even powerful men knew there were others still more powerful than they. Yet all men learned from experience that the system's value rested on the neutrality of the judges (**implication 1g**). When a disputant lost his case at the local level but felt it was due to oversight or bias by the local judge, he would appeal to the king's court on the ground of "default of justice".

Such appeals put pressure on local judges to reduce their errors and biases *and* pressure on the central courts to issue new rules to increase neutrality by local judges (**implication 1h**). Even if these rules were motivated simply to reduce the volume of appeals coming from the local level – to avoid an overload at the central courts – the rules could still be effective.

## 7. Macro effects of the reforms

### 7.1 Political stability

Prior to the legal and judicial reforms, lesser lords were vulnerable to having their land seized by more powerful lords. Their vulnerability was highest in the decades after the Conquest (1066-1100) and the "Anarchy of King Stephen's Reign" (1135-1165). Yet even in the reign of Henry I (1100-1135), the lesser lords were still vulnerable. The reforms gave the lesser lords more legal protection against land seizures by the greater lords. This security stabilized the distribution of land and the distribution of political authority. Since lords did not have to wage violent conflicts over land disputes, they did not have to compete so intensely for relative military power. And they did not have to usurp taxing and spending rights from other lords to maximize power. They could allow the existing lines of fiscal authority to remain intact.

#### *Contrast with political instability in medieval Germany*

In the absence of legal and judicial reforms, German nobles continued to rely on ancient customs and judicial procedures to settle land disputes. Since these practices did not support anchor-based enforcement, the nobles resorted to power-based enforcement to protect their land rights. When land disputes arose and became irreconcilable, nobles used violent conflict to punish perceived violations of their rights and deter further violations. This induced a competition in raising the relative power to defend land and deter rivals from expanding territorially. In this environment land was divided among nobles based on relative power.

The competition for power gave lords an incentive to usurp taxing and spending rights from overlords and neighboring lords. The result was political instability.

## 7.2 Social capital formation

The legal and judicial reforms promoted social capital formation by creating shared values, interests and identity among nobles. Because land was divided among them on terms that were independent of relative power, the resulting allocation constituted a shared value – a mutual commitment that transcended the ebbs and flows in their relative power. As agricultural productivity and incomes increased, nobles of all ranks gained a shared interest in the rules for dividing land and settling land disputes. The social approval that nobles gave each other for complying with the rules created a sense of shared identity.

The increase in social capital formation was evident in three measures: jurisdictional integration, direct dialogue, and markers of social approval. Jurisdictional integration between the county courts and the royal courts increased as a result of the ascent of appeals to the royal courts and the descent of procedural rules to the county courts. This development increased the amount of direct dialogue between greater and lesser lords in their land disputes with each other and their efforts to monitor court officials and check abuses of judicial authority. Closer interactions between greater and lesser lords, mediated by the court system, led to the emergence of new markers of social approval for the lesser lords who were now legally enfranchised. As they transitioned from the military profession of knight service to the agricultural profession of estate management and ownership, they came to be known as "gentleman" and their class as the "gentry".

### *Contrast with social capital depletion in medieval Germany*

In the absence of legal and judicial reforms, German nobles continued to rely on the ancient customs and judicial procedures. Since these forms do not support anchor-based enforcement, nobles resorted to power-based enforcement to protect their land rights. When land disputes arose and became irreconcilable, nobles used violent conflict to punish perceived violations of their rights and deter further violations. The result was a competition in raising the relative power to defend land and deter rivals from expanding their land holdings. In this environment land became divided among nobles based on relative power.

These developments caused social capital depletion. Without any mutual commitment

to stationary land boundaries, the main value the nobles shared was the pursuit of power, even at each other's expense. Their main shared interests were simply to form alliances to aggregate power and to regulate the violent conflicts to limit their worst side-effects. As agricultural productivity and incomes strained under the pressures of conflict, the stakes in the competition for relative power rose. Social animosities increased and noble society fractured into competing identities of "robber barons" and "robber knights".

The social capital depletion was evident in three measures: jurisdictional separation, indirect subversion, and markers of social disapproval. German lords of all ranks created separate jurisdictions with independent courts and prohibited their vassals from appealing court verdicts to any higher lord – be it a regional magnate or the king. As a result any lord with his own jurisdiction could exploit his power advantage over his vassals with impunity. He could commit land and tax theft against peasants and other lords and then manipulate his court proceedings in his favor. To contend with such rapacity, disgruntled overlords, neighboring lords, vassals and peasants abandoned direct dialogue through court process and resorted to indirect subversion. Mounting distrust led to increased perceptions of social distance and the use of pejorative labels of social disapproval. Higher lords were regarded as robber barons, local lords were regarded as robber knights, and peasants were regarded as "trees that need pruning".

## TABLE OF CONTENTS

## CHAPTER 14

### From the world of law to the world of power in Medieval Germany

#### 1. Background conditions

In the 10th century, the east Frankish kings established a large empire in Germany through a combination of conciliation of their own dukes, diplomacy with the magnates of west Francia, and repulsion of the foreign invaders from the north and east.<sup>1</sup> In these military campaigns against the Vikings, Slavs and Saracens, each German duchy was led by its duke rather than the king and, in this sense, was a separate polity. In compensation for the dukes' service, the German kings distributed land among them by means of written charters. Each duchy was populated by a different German tribe and ducal government was based on the tribal customs of the ancient germanic war bands.

In the 11th century, the empire arrived at its natural limits of expansion and the pressure of foreign invasions relaxed. Infighting among and within the German tribal duchies led to the disintegration of royal authority.<sup>2</sup> A number of smaller duchies and counties emerged with more or less independent sovereignty. In each of these lordships political authority was based on land ownership and usage rights. The overlord, a duke or count, delegated his authority to local lords through feudal land grants. The local lord promised to provide military service to the overlord, and in return, received the rights to all of the land, incomes and governmental authority within his district. Ownership of land was thus valuable not simply for the income from agricultural production, extracted in the form of rents, but also for the income from market taxes and tolls, judicial profits and feudal incidents accruing to the local lord as sole political authority holder in his district. All of this income was placed at stake by disputes over land and jurisdiction. So owners and vassals fought hard over land

---

<sup>1</sup>Reuter 1991: 139-147, .

<sup>2</sup>Reuter 1991. Arnold CITE date. 1991?

disputes.

### 1.1 method of dividing goods between actors

Land was apportioned among nobles in discrete parcels. The parcels changed hands from one noble to another through transfers, grants, inheritance and the use of force. Nobles transferred land for compensation by mutual agreement. When a noble died, his heirs inherited the land. If there were no heirs, the overlord took possession and regranted the land to another man. Overlords also regranted the land of vassals who had fallen out of favor. Land was also be usurped by force. When a parcel of land changed hands, its borders remained intact. In short, land was divided by means of borderline contracts whose terms did not change in response to marginal shifts in relative power. A noble did not alter the borderline between his parcel and his neighbor's parcel every time his relative power increased marginally or momentarily. Of course he might usurp the neighbor's *entire* parcel if his power increased by enough to do it. But alterations in the borderlines did not arise from the marginal and momentary fluctuations in relative power that occurred all the time. In this sense, title to real property in medieval Germany was distributed by anchor term contracts, as the term is defined by the theory of justice bargains.

Titles to parcels of land were obtained by recognition of the local community. Recognition was provided either by a written charter from the community's lord or by verbal acclamation of the members of the community.<sup>3</sup> The lord's decision to recognize a person's title often followed from the community members' decision and acclamation. They might give it in an assembly in the local court. Alternatively, if an inquest was needed, they would view the land in question and render their decision and acclamation *in situ*.

### 1.2 causes of disputes

Disputes over title to land arose from the fact that there were many exceptions to the basic rules of land tenure and usage.<sup>4</sup> Often it was clear that a particular exception was warranted

---

<sup>3</sup>For charters, see Arnold 1985: 140-1, 146. For acclamations by the community, see von Repgow 1235 [1999]: 158, 172.

<sup>4</sup>von Repgow 1235 [1999]: Book IV.



by the facts of the case. At times, however, novel contingencies arose in which it was not clear whether a particular exception was warranted. If the written charter recording the terms of title was unclear or there was no written charter, such legal ambiguities could cause disputes. Two causes of dispute were prevalent (similar to Medieval England). First, legal ambiguities arose from the random forces of birth and death – beyond anyone’s control – that altered the family lines of inheritance.<sup>5</sup> Second, gaps and shortcomings in the titling system itself provided nobles with opportunities to create ambiguities intentionally in search of parcels of land that they could not gain otherwise. For example, a landholder who had transferred his land might try to retain possession of it, nonetheless, by denying that he had transferred it.<sup>6</sup> To take another example, a landholder might grant his land to another man to prevent the land’s reversion to his lord or inheritance by his child after his death.<sup>7</sup>

### **1.3 main tasks of the judicial system**

Courts hearing of land cases had two main tasks (similar to England). First, they had to determine whether exceptions to the land law were warranted by the facts. Second, they had to determine whether one of the litigants had acted opportunistically. The judicial system in Germany did not develop in the same way that England’s system developed in the medieval period, however.

## **2. The persistence of customary law through the medieval period**

After the political fragmentation of the 11th century, each duke or count gained full control of the court in his district. Trial procedures in these courts followed the ancient customary

---

<sup>5</sup>von Regow 1235 [1999]: Book IV, 163. Arnold 1991P: 239. (”.”) Arnold 1997: 70. (”Although the rhythm of local politics was dictated by feuds, the territorial contours of aristocratic power were also affected by the demise of dynasties.... Aristocratic inheritance explains a great deal about political change in the provinces.... Feuds were motivated and fuelled by these dynastic accidents. When the landgraves of Thuringia failed in 1247, their enormous inheritance was claimed by Margrave Henry of Meissen, but after years of war, he could not prevent the other chief claimant, Henry of Brabant, from seizing Hesse and installing himself there as landgrave. Similar feuds disturbed Franconia and Bavaria [and] Lower Lotharingia.”)

<sup>6</sup>Dobozy 1999: 157.

<sup>7</sup>Dobozy 1999: 163 par. 60; 155. ([Landholder] ”tries to void his lord’s or his child’s claim [to his land]...by granting it anew or transferring it.” CHECK IT.)

rules. A panel of *schoffen* – usually translated as *assessors*<sup>8</sup> – rendered judgment on the basis of the ancient methods of proof: trial by water, ordeal, battle or compurgation (ritual oath-giving). The first three methods did not rely on the facts of the case at all. The fourth method, oath-giving, relied on the facts only in theory, not in practice. Precise formulas dictated the number of oaths that had to be given by each litigant and his supporters depending on the type of case it was. Precise formulas dictated the wording with which each oath was to be recited. Many of the formulas were purely procedural: they contributed nothing to the fact value of the oath.<sup>9</sup> The formulas also dictated which litigant should receive the favorable judgment as long as he and his supporters gave the prescribed number of oaths in the prescribed wording. In this trial procedure, the assessors' task was merely to determine whether the oaths had been given in the prescribed numbers and wording and to apply the formula dictating which side won the case.

Alternatively, the assessors might listen to the claims and oaths of the opposing litigants and then act as "law-finders". To "find the law", the assessors simply stated what the facts of the case amounted to in legal terms – *in toto* – saying either "custom has been violated" or "custom has been respected". This procedure did not distinguish between the facts of the case, on one hand, and the legal rules invoked by those facts, on the other hand. Each of the litigants' claims or oaths was an *amalgam* of the facts of the case and the rules invoked by those facts. For example, "The lord confiscated his vassal's fief in a way that he should not have done according to the customs governing this situation." Or: "The lord confiscated his vassal's fief in the way that he should have done according to the custom governing this situation."

The ancient trial procedures and methods of proof remained standard in Germany throughout the medieval period. While trial by battle and ordeal fell out of use in the 1300s, ritual oath-giving continued to be used into the 1400s.<sup>10</sup> As a method of proof, the practice of

---

<sup>8</sup>Nicholas 2009: 170-1.

<sup>9</sup>Nicholas 2009: 177. Dobozý 1999, 12: "Legal procedure was experienced as formal spoken language accompanied by prescribed gesture. Oral practice continued in all procedures of the law court, including pleas, contracts, payments, transfers, *oaths*, and appeals." Emphasis added.

<sup>10</sup>Dawson 1960: 97. Nicholas 2009: 169, 176.

oath-giving was "irrational".<sup>11</sup> The persistence of this method along with the other customary procedures hampered legal development in a way that increased the rate of conflict over legal disputes in medieval Germany.

[The forms of proof were irrational – trial by battle, ordeal and unsupported oath. The trial procedure was oral and emphasized form and ritual over factual substance and legal evaluation.]

[The judicial system relied on the ancient customary process and methods of proof.]

[The king's main goal was to bring the opposing parties to a compromise settlement through a combination of moral suasion and physical threat. The emphasis was on compromise and reconciliation between the parties rather than determining which party was right and which wrong. In this judicial system the king was valued not primarily for his neutrality, knowledge of the law or deeper wisdom, but rather for his ability to cajole, shame and coerce the recalcitrant party into submission and compromise. As a result litigants endorsed the king for his partisanship when he was on their side and rebuked the king for his partisanship – or even rebelled – when he was on the opponent's side. This approach to adjudication continued to predominate long after the Ottonian empire fell. In the new, decentralized political system that emerged in the 11th century, each duke or count had his own court and the power to restrict or prohibit appeals to any higher court. The judicial methods of these courts remained customary and suffered from the same liabilities of customary justice already described.]

### **3. Causes of the prevalence of conflict in dispute processing (implication 1.1)**

#### **3.1 cause 1: ineffectiveness of the system at supplying facts to courts**

The system of oath-giving was ineffective at supplying facts to the assessors because it did not require or allow many facts to be supplied. The formulas prescribing the number and content of the oaths limited the number of facts that each claimant *needed* to support his

---

<sup>11</sup>Baker 2002 CITE pages.

case. The formulas also limited the number of facts that he was *allowed* to supply. The oath-givers "were not cross-examined and were seldom expected to testify to specific facts that they had observed."<sup>12</sup> They simply recalled facts from memory and supplied them orally to the assessors, who considered them without aid of written documents.<sup>13</sup>

In the late medieval period, the courts began to use the testimony of witnesses, in the modern sense of the term, as a source of factual evidence.<sup>14</sup> However, this source of facts had a number of limits. First, scholars are not even in agreement that witness testimony was used as a source of factual evidence.<sup>15</sup> Second, witness testimony was limited to criminal cases.<sup>16</sup> Hence witnesses did not displace the customary method of oath-giving as the main form of proof.<sup>17</sup> "So long as this procedure survived, the conduct of trials imposed no heavy burden on the *schoffen* in assembling and evaluating the evidence."<sup>18</sup> Third, the court system had no structure of petty and grand juries akin to the English structure.<sup>19</sup> So judges and assessors could not authorize inquests to ascertain the additional facts needed to resolve a

---

<sup>12</sup>Dawson, 1960: 97-98.

<sup>13</sup>Dobozý 1999, 12: "Evidence was also oral, as it had to come from a guarantor, an eyewitness, or a witness who recalled information from memory, since written records were rare. Legal practice was thus forced to remain oral as long as a support system of written documentation and court records was largely unavailable."

<sup>14</sup>Nicholas 2009: 176-7. Dobozý 1999: 22, "In the old system [of the Carolingian and Ottonian eras] people had not been concerned that compurgation and cleansing oaths might be unsuccessful means for discovering facts. The new approach attested to in the Saxon Mirror [ca. 1230] signals an attempt to ascertain factual information by placing increasing dependence on witnesses, whose testimony was understood as a means of establishing facts."

<sup>15</sup>See next note.

<sup>16</sup>Dawson 1960: 97-98, "There has been some dispute as to whether the 'witnesses' referred to in 13th century sources were merely oath helpers, supporting in general terms the assertions of the party who produced them, or were on the contrary expected and encouraged to disclose specific facts from their own direct observation. Planck, discussing Saxon procedure as described in the Saxon mirror, insisted that the 'witnesses' were merely oath helpers, even though they might in particular cases refer in detail to facts or events which showed that the claim they supported was just.... The same conclusion was reached by Kleinfeller as to the practice in other districts of Germany in the 14th century.... However, a more recent author has argued at length that witnesses able to testify to facts or events within their own knowledge were used, indeed required, in many districts of Germany as early as the 13th century. The sources he quotes, however, are almost all concerned with criminal cases."

<sup>17</sup>Nicholas 2009: 176-7.

<sup>18</sup>Dawson 1960: 97-98.

<sup>19</sup>Nicholas 2009: 171.

dispute definitively.<sup>20</sup> Fourth, even the limited set of facts one court's assessors did consider could not be preserved for use in any retrial there or elsewhere, since the method of oath-giving was an oral procedure. As a result of all of these deficiencies, the judicial system left little written evidence of the cases it tried – either for its own appellate courts or for later historians.<sup>21</sup>

In civil matters some key facts might be supplied to the courts by another route: written documentation of transactions. The assessors and other court officials could certify that they had witnessed certain commercial dealings, land grants, land transfers, and the settlements of frontiers after feuds.<sup>22</sup> Such transactions were becoming more frequent as legal title to offices and lands became a kind of currency that nobles traded in an effort to amass resources and military power to wage their feuds. As a result the drafters of such documents felt some pressure to condense and standardize their language.<sup>23</sup> The certifying documents could be used as evidence in court proceedings. Yet even here, the power of court officials to generate facts for use in trials remained limited. Written accounting of lords' legal rights to their domains, fiefs, toll stations, and other income-generating properties only became common after 1400; and even then the accounts were less sophisticated than those of English lords and royal government officers.<sup>24</sup>

In sum, the system had a limited capacity to supply facts to courts. This was a major obstacle to legal development because it hindered judges and assessors from dealing effectively with ambiguities in the customs. To understand this problem it is necessary to consider the nature of customary law and the difference between customary and formal law in how each deals with facts.

---

<sup>20</sup>Nicholas 2009: 177. Dawson 1960: 98-9. "...[T]o go outside their own membership and tap sources of information that might be available in their local communities required greater powers of coercion than most *schoffen* courts could mobilize. The jury of trial, descendant of the Frankish royal group inquest, was used for limited purposes in a few districts. Perhaps it could have been used more widely if political collapse had been arrested."

<sup>21</sup>Arnold 1985: 187, "[As far as the] surviving [written evidence], there is nothing in Germany to compare with the rich fiscal and judicial records of the Angevin and Capetian realms...."

<sup>22</sup>Dawson 1960: 98-9. DuBoulay 1983: 11. Arnold 1985: 146, 196-8.

<sup>23</sup>DuBoulay 1983: 11. "[S]cribes were learning to be briefer in recording the necessary descriptions."

<sup>24</sup>Nicholas 2009: 138. Cf. DuBoulay 1983: 12-3, "The keeping of financial accounts also came later in Germany than in England, and there are few from the 14th century even in the greater principalities."

### 3.2 cause 2: ineffectiveness of the system at distinguishing between questions of fact and questions of law

The oral court procedures used in the medieval period up to 1450 did not distinguish between fact and law.<sup>25</sup> Hence these procedures could not distinguish between questions of fact and questions of law. This incapacity prevented the judicial system from handling ambiguous cases in a definitive and precedent-setting way. The root of the problem was the trade-off between inventing new law to settle ambiguous cases, on one hand, and the need to limit the number of new laws to keep the judicial process affordable, on the other. This trade-off is reduced considerably by a judicial process that distinguishes between questions of fact and questions of law (as shown in the chapter on Medieval England). By contrast a system that fails to make this distinction runs afoul of one end of the trade-off or the other.

In medieval German courts, civil disputes often ran into ambiguities in the customs. In the case of an ambiguity, the court's assessors had three options. First, they could issue a single, authoritative verdict favoring one litigant or the other, regardless of the injustice it might cause. The problem with this solution is that it left the litigant suffering the injustice with no recourse except to resort to self-help – potentially through feuding. Second, the assessors could say, 'If *these* are the facts *in toto*, then the verdict is guilty. But if *those* are the facts *in toto*, then the verdict is not-guilty.'<sup>26</sup> The problem with this solution is that it was no verdict at all: it simply returned the dispute to the litigants to resolve in some other way – either by appealing to another court or by feuding. Third, the assessors could invent a new sub-category of the custom in dispute to cover the novel contingency that generated the ambiguity. That is, they could say, 'If *some* of the facts point to a verdict of guilty while *others* point to a verdict of not-guilty, then the guilty verdict may be rendered *only* if the first set of facts is proven, but otherwise not.' The problem with this solution is that it would generate a new sub-category of custom (e.g. a new exception to the existing custom) every time an ambiguous case arrived in a court. The number of customs would grow so large that no assessor or court of assessors could remember them all. Even if all the customs

---

<sup>25</sup>Baker 2002: 27. ("Germany had never had a national system of law, and Germanic customs were obscured by the late survival of irrational modes of proof. When the archaic oral procedures gave way at the end of the fifteenth century to written procedures, *which separated fact from law*, there was no law ready to fill the scientific vacuum." Emphasis added.)

<sup>26</sup>Dawson.

were written down, the book would be out of date immediately because of the constant flow of new ambiguous cases into the courts.

The customary judicial process – to “find the law” in the facts – could not do anything beyond these three options. The root of the problem lay in the very nature of customary law. A custom is a statement of the right action to take in a specific situation. That situation is defined by certain facts. The customary action is right in that situation *not* because the situation is governed by any set of abstract legal rules that distinguish between right and wrong. The action is right simply because it is the customary or normative action to take in that situation. It is simply “how things are done” in that situation. The facts of the situation *together with* the normative action to take in that situation form an inseparable *amalgam* of fact and law (or more precisely, fact and “right”). For on one hand, the facts of the situation *include* the customary “right” action to take in that situation. On the other hand, the customary “right” action to take is no more than a subset of the facts of the situation – that subset which describes the normative action to take in that situation. Yet this action is no more than another fact of the situation. In short, “how things are done” in a particular situation is simply a matter of how things were always done in that situation customarily. And that amalgam of fact and law is “the custom”.

Now it becomes clear why the medieval German assessors who confronted an ambiguity in a custom had limited options. On one hand, they could not inquire more deeply into the facts of the case due to the system’s limited capacity to supply facts to courts (as explained above). On the other hand, they could not inquire into the nature or foundations of the law *as a set of abstract legal rules* because this is not what the customs were.<sup>27</sup> The assessors could only choose among the three options listed above. Since they sometimes chose the third option – creating a sub-category of custom based on a new exception to an existing custom – the number of customs expanded over time. The system was flexible and adaptable in this sense. Yet its growth was neither systematic nor, as a result, precedent-setting.

The growth was not systematic because each novel settlement of an ambiguous case stood as a new addition to the customs in itself, no matter how similar or different it was

---

<sup>27</sup>Schulze 1992: 13 (as cited in Dobozý 1999: 15). (“[This is] a culture in which the development of laws and legal settlements are not at all as yet predetermined by legislation and legal theory and also do not depend upon any written formulation of law as code, concept, or system.”)

from previous additions.<sup>28</sup> One addition had no more legal significance than another. By implication, the *omission* of one of these new additions from a written compilation of the customs did not have any more significance than the omission of another. An omission did not reflect a gap in the *law*, but rather just a failure to include a particular type of *dispute* in the compilation.<sup>29</sup>

The written compilations of customs produced [in the 13th century ] did not constitute a systematic body of law to which the assessors could refer in facing new ambiguous cases.<sup>30</sup> Since the body of customs was always changing with new accretions in response to new ambiguities, no compilation could be definitive for long.<sup>31</sup> Hence, while such a compilation could provide the assessors with guidance in reaching a verdict, it was not an authoritative source of law.<sup>32</sup>

---

<sup>28</sup>“Until someone attempted, by study and analysis, to add up and coordinate its important results, full reporting of the [assessors’] work case-by-case would probably have added to the prevailing confusion. Again we encounter the truism that ingenious solutions to particular cases do not add up to a body of law unless someone can be found to do the adding.... Neither in the working methods of the assessors nor in their output itself was there much to inspire analytical study. The assessors did not surround themselves with an organized bar.... [T]hey had neither time, training nor disposition to engage in extensive analysis or to prepare their judgments in a form that would facilitate study by others.”Dawson 1968: 173-5.

<sup>29</sup>Dobozy 1999: 13. (“Eike [the author of the *Saxon Mirror*] acknowledges this concept and signals its implications in the first prologue [of the book]: ‘I request the support of all law-abiding people who desire justice. If they encounter a legal dispute that I have omitted from this book because of my limited knowledge, I request that they reach a determination according to justice to their best knowledge and discretion.’ Eike is not concerned about having omitted a part of the law; rather, he warns that he may have omitted an example of a particular type of dispute and is asking his readers to apply legal custom based on their community’s knowledge which is not exhausted in the book. In fact, the text [of the *Saxon Mirror*] never uses the term *recht* to refer to a written document.”)

<sup>30</sup>Dawson? Page? As a result of these limits, the assessors judgements held little value as sources of legal development.

<sup>31</sup>Dobozy 1999: 13. (“[C]ustomary law, like all human institutions, is constantly in flux. Consequently, it is best to view the written customal with more than 450 manuscript versions as participating in that flux.... Each *Saxon Mirror* manuscript freezes one moment in the evolution of the judicial system and is not an exhaustive nor necessarily accurate record of custom. What is more, it need not be. Lacking the precise formulation of a code, their unsystematic, open-ended organization allows for much of the flexibility of the [oral] legal tradition as practiced by the judiciary.)

<sup>32</sup>Dobozy 1999: 12. (“...[C]ustomals are not to be read as we read modern written codes, that is, with the expectation of a fixed, systematic presentation.... [They] exhibit a great deal of variation.... [The writers and users of the *Saxon Mirror*] treated the lawbook as a flexible *guide* that they could adapt to their own needs because they frequently made changes that deviated from acknowledged local or territorial custom outside of Saxon. The lesson here is that a customal is best understood today in the context of its oral traditional foundation. This means one must know what not to expect: a customal is not a legislative code of law in the modern sense, *has no normative function*, is not fixed, and is not an exhaustive account of legal custom.” Emphasis added.)



+++++ [What they could not do was to determine whether the new exception was needed due to random factors beyond any litigant's control or to the intentional actions of one litigant.]

The right of appeal was often limited because the court of the local lord or the regional overlord had full sovereign jurisdiction (e.g. that lord prohibited appeals to the royal courts).

### **3.3 cause 3: ineffectiveness of the system at preserving verdicts for use as precedents in future cases**

Court verdicts were based on customary law. These laws were collected in writing in three forms. However, none of these forms provided any basis for a system of case-based law. [was updated to reflect novel verdicts that resolved ambiguities in the existing law.]

#### **3.3.1 local custumals**

Beginning in the 1200s, each rural court composed a written record of its customs.<sup>33</sup> These "custumals" included both statements of the court's customary law and records of the pleas and verdicts rendered in specific cases. Neither source could enable the development of case-based law, however. There were several reasons for this. First, each verdict was the product of the collective judgment of the local community in assembly, rather than a bilateral negotiation between opposing counsel or a decision by a single judge.<sup>34</sup> Since the case records were dictated and composed by villagers who were untrained in the law, their language was too terse and colloquial to reveal the information necessary to form a system of case-based law.<sup>35</sup>

Second, each verdict was *sui generis*. It had no binding authority on subsequent cases.<sup>36</sup>

---

<sup>33</sup>Toch 1986: 677-682. Dobozy 1999: 28: beginning in the late 1100s

<sup>34</sup>"...*Recht* also meant the duty of community members to know legal custom and to participate in the legal system in order to maintain justice and the community.... Everyone participated in all aspects of legal custom, thus implying that the participatory process itself is what creates justice.... Without the contribution of the [court] members and a consensus on a judgment, there is no customary law." (Dobozy 1999, 11.)

"[C]ustomary law [of this kind] is a social phenomenon, like language: it exists an individual application but is invalid unless the community of speakers – in this case the members of the judiciary – acknowledges its meaning.... [T]hey must determine which custom applies to the specific situation...." [14] (Dobozy 1999: 11-4.)

<sup>35</sup>CITE.

<sup>36</sup>"The majority had to agree to each judgment before any final determination could be reached in the case. Since legal judgments required the consensus of the members of the judiciary, it is clear that the

Third, each court was independent in that there was no superior court at the regional level, or if there was, the local court's suitors did not have the right to appeal to it. Hence there was no hierarchy of judicial officials that could impose uniformity on the trial process, the recording procedures or the language. As a result the customals remained strictly local in perspective and static in content.

"[T]hey expressed traditional ideas and dealt with the same recurrent themes.... Wholly unprocessed by lawyers' logic and filled with uncountable diversities, they were the raw material for a legal system, grist for a mill that did not exist.... [T]hey do not reveal much conscious innovation of the kind that often appeared in the manorial by-laws in England."<sup>37</sup>

### 3.3.2 Regional compilations of local customals

The customary laws of Germany were collected into three written digests in the 1200s.<sup>38</sup> These digests were distributed widely and used as a source of law. The most widely known, the *Saxon Mirror* "clearly met a need that was widely felt for a reasoned and orderly statement of the elements common to the customs. It had a significant effect in marking a

---

responsibility for legal decisions rested with the community and not with the individual. Consequently, the analogy with language illustrates that the system evolved [in a way] requiring the consensus among the judiciary for each judgment because custom exists only in the community. If consensus fails, then customary law fails because it did not apply to a given situation. *Each time a case was tried, the custom had to be reestablished, and thus individual settlements and judgments had no normalizing function.*" (Dobozy 1999, 14. Emphasis added.)

"...[C]ustomals are not to be read as we read modern written codes, that is, with the expectation of a fixed, systematic presentation.... [They] exhibit a great deal of variation.... [The writers and users] treated the lawbook as a flexible guide that they could adapt to their own needs because they frequently made changes that deviated from acknowledged local or territorial custom outside of Saxony.... [A] customal is best understood today in the context of its oral traditional foundation. This means one must know what not to expect: a customal is not a legislative code of law in the modern sense, *has no normative function*, is not fixed, and is not an exhaustive account of legal custom." (Dobozy 1999, 12. Emphasis added.)

<sup>37</sup>Dawson 1968: 154-6. Cf. cite. Some of the customals produced by the German city courts were more articulate and refined in the areas of commercial and property law. These customals were sought and adopted elsewhere for their perceived usefulness. Yet they suffered from the same drawbacks as sources for the development of case-based law.

<sup>38</sup>The *Saxon Mirror*, collecting the customs of Saxony, was written in the 1230s. The *German Mirror* was written in the 1260s. The *Swabian Mirror*, collecting the customs of Swabia, was written in the 1270s. (Nicholas 2009: 164.) The Saxon mirror is concerned primarily with the procedures and law applied in the counts' courts rather than royal courts or princes' courts. The counts' courts had jurisdiction over all civil cases involving nobles. (Dobozy 1999: 17-8.)

path through the undergrowth.”<sup>39</sup> It included both manorial customs governing relations between lords and peasants and feudal customs governing relations between lords. Yet it had fundamental shortcomings. It neither reflected nor created a single, truly common law for Germany.<sup>40</sup> It neither resolved ambiguities in the underlying customs nor created any system for resolving them by means of specific verdicts in key cases.<sup>41</sup> As a result, it did not inspire the development of a body of formal case law.

### 3.3.3 City courts

Difficult cases from all kinds of courts were sent to the *schoffen* in the city courts for definitive resolution. These courts kept written records of their decisions starting in the 1300s. Since some of the writers tried to condense the cases in a usable way, the records “have all the external marks of having been intended as precedent-books.”<sup>42</sup> Yet these records had limited value as sources of case-based law. The city *schoffen* were not trained lawyers or jurists, so their adjudication method was no more systematic than that of other courts. Information on each case was received mainly through written reports sent by the litigants, their attorneys or the officials of the referring court. The city *schoffen* did not conduct adversary hearings using oral arguments, written briefs or rules of procedure to resolve questions of fact. They had to accept the facts as reported, despite any errors, gaps or ambiguities.<sup>43</sup> Their decisions, in turn, left little scope for the accumulation of legal understanding among the litigants, attorneys or officials of the referring court. Often the decision simply stated the opposing claims and the proposed verdict without summarizing the key facts of the case. The *schoffen* often gave the verdict without stating the reasons for it explicitly. If the facts fell into gaps in the law, the *schoffen* often simply treated the facts “in a conglomerate way”, thereby returning the dilemma to the litigants. The decision said, in effect, ‘If *these* are the facts, then this is the verdict. If *those* are the facts, then that is the verdict.’<sup>44</sup>

---

<sup>39</sup>Dawson 1968: 153-4.

<sup>40</sup>DuBoulay 1983: 86.

<sup>41</sup>Dawson 1968: 154.

<sup>42</sup>Dawson 1968: 172.

<sup>43</sup>Dawson 1968: 158-160, 167, 174.

<sup>44</sup>Dawson 1968: 164, 169-70, 174.

The city *schoffen*'s courts were not appellate courts, so these judges had neither the authority nor the incentive of a true appellate jurisdiction. They could not require a difficult case to be referred to them. Their decisions did not establish precedents with binding authority.<sup>45</sup> Powerful counts often overruled their decisions and prohibited subjects in the county from sending cases to them.<sup>46</sup> As a result, the city *schoffen* had no incentive or authority to systematize their decisions into a body of formal case law.

"Until someone attempted, by study and analysis, to add up and coordinate its important results, full reporting of their work case-by-case would probably have added to the prevailing confusion. Again we encounter the truism that ingenious solutions to particular cases do not add up to a body of law unless someone can be found to do the adding.... Neither in the working methods of the *schoffen* nor in their output itself was there much to inspire analytical study. The *schoffen* did not surround themselves with an organized bar.... [T]hey had neither time, training nor disposition to engage in extensive analysis or to prepare their judgments in a form that would facilitate study by others."<sup>47</sup>

By the mid-1400s, dissatisfaction with the legal system in general, and by implication with its most respected judges, the city *schoffen*, led to widespread demand for legal reform.<sup>48</sup>

### 3.3.4 Lack of true appellate jurisdiction

In the late 1200s, the nobles began to form Estates and Diets to order their relations with the ruling lords. However, these organizations were not courts of law or legislatures with the right to settle disputes, set precedents and promulgate statutes, as the parliaments of France and England were. Rather they were administrative bodies for the regulation of the magnate's tax collection and spending activities.<sup>49</sup> Before the mid-1400s, none of the German kings established a body of laws or court procedures, as the English and French

---

<sup>45</sup>Dawson 1968: 168, 173.

<sup>46</sup>Dawson 1968: 175.

<sup>47</sup>Dawson 1968: 173-5.

<sup>48</sup>"In the course of time complaints began to be raised against them for the inconsistencies in their rulings and for suspected partiality." Dawson 1968: 173-5. Dawson cites German scholarship on the complaints voiced in the 1500s and later. Dahmer cites German scholarship on the complaints voiced in the 1400s?

<sup>49</sup>Nicholas 2009: 141.

kings had done since the 1100s.<sup>50</sup>

#### 4. Prevalence of conflict in dispute processing (implication 1.1a)

When actors could not settle their disputes peacefully, they resorted to violent feuding. Actors of all kinds used feuding as a means to defend their rights – nobles, ministeriales, freeholders and peasants.<sup>51</sup> Most feuds were sparked by disputes over civil matters.<sup>52</sup> Organized violence was viewed as a legitimate means of settling disputes over land, political jurisdiction, commercial rights, debts and other civil matters.<sup>53</sup> Since disputes of this kind occurred frequently, feuding broke out frequently.<sup>54</sup>

Once two individuals failed to resolve a civil dispute peacefully and resorted to feud, it usually escalated into a "private war" between their armed retinues. The members of each retinue were obligated to provide military support to their leader.<sup>55</sup> Once the opposing retinues took the field, the feuding often descended into indiscriminate violence.<sup>56</sup> Responsibility for the consequences was difficult to pin down. While the feud was a method of enforcing individual rights, its practice blurred the line between the principal perpetrator,

---

<sup>50</sup>Nicholas 2009: 163.

<sup>51</sup>For peasant feuding, see Cite. For non-noble freeholder feuding, see Dobozy 1999. For noble and ministeriale feuding, see Arnold 1985.

<sup>52</sup>Arnold 1985, 245-6. ("The majority of feuds appear to have been motivated by disputes over property rights, and to have had their material cause in the social position of ministerials as armed landowners with acquisitive tastes.")

<sup>53</sup>Nicholas 2009: 172. Arnold 1985. (225-6: "[T]he contemporary aristocratic outlook...regarded feuds as a legitimate means of redress." 245-6: "In certain circumstances, acts of violence were acceptable to the noble orders in Germany, and their evil consequences to some extent mitigated by customs governing the conduct of feuds.... Violence was a double-edged instrument in knightly hands. Its acceptable basis lay in custom permitting vengeance, redress and self-defense in noble society.") DuBoulay 1983: 71-3. ("[Feuding] was an accepted instrument of politics and law...More often the cause of war was a debt, real or imagined, a territorial claim following a partitioned inheritance, a marriage or an ambiguous grant [of land], or simply the friction between neighbors, especially where mercantile wealth was in some way open to plunder.")

<sup>54</sup>Nicholas 2009: 175. Arnold 1985, 243. ("Feuds in Westphalia were incessant....")

<sup>55</sup>Arnold 1985, 246, "The [ministerials'] use of violence was complicated and extended by the medieval German nobleman's expectation of support in all such actions from his relatives, friends and armed following. Even when the feud originated in the grievances of an individual, it became the collective act of a kindred operating from more than one castle."

<sup>56</sup>Arnold 1985, 242, "Medieval German society paid a high price for the military mentality of the free nobility and of the ministerials in the form of the extensive devastation arising from their feuds.... [F]euds might deteriorate into large-scale knightly marauding." Arnold 1985, 225, "[M]inisterials pursued destructive feuds which could degenerate into extensive banditry over many years."

on one hand, and his accomplices, on the other.<sup>57</sup> Hence private war, as a means of settling disputes, generated a dilemma for its practitioners. They wanted to reap its benefit – the maintenance of deterrence against any competitor who might violate their rights – while avoiding its unintended cost – indiscriminate mayhem and carnage.

## 5. Prevalence of conflict in dispute processing (implication 1.1a)

When actors could not settle their disputes peacefully, they resorted to violent feuding. Actors of all kinds used feuding as a means to defend their rights – nobles, ministeriales, freeholders and peasants.<sup>58</sup> Most feuds were sparked by disputes over civil matters.<sup>59</sup> Organized violence was viewed as a legitimate means of settling disputes over land, political jurisdiction, commercial rights, debts and other civil matters.<sup>60</sup> Since disputes of this kind occurred frequently, the feuding broke out frequently.<sup>61</sup>

Once two individuals failed to resolve a civil dispute peacefully and resorted to feud, it usually escalated into a "private war" between their armed retinues. The members of each retinue were obligated to provide military support to their leader.<sup>62</sup> Once the opposing retinues took the field, the feuding often descended into indiscriminate violence.<sup>63</sup>

---

<sup>57</sup>Nicholas 2009: 175.

<sup>58</sup>For peasant feuding, see Cite. For non-noble freeholder feuding, see Dobozsy 1999. For noble and ministeriale feuding, see Arnold 1985.

<sup>59</sup>Arnold 1985, 245-6. ("The majority of feuds appear to have been motivated by disputes over property rights, and to have had their material cause in the social position of ministerials as armed landowners with acquisitive tastes.")

<sup>60</sup>Nicholas 2009: 172. Arnold 1985. (225-6: "...[T]he contemporary aristocratic outlook...regarded feuds as a legitimate means of redress." 245-6: "In certain circumstances, acts of violence were acceptable to the noble orders in Germany, and their evil consequences to some extent mitigated by customs governing the conduct of feuds.... Violence was a double-edged instrument in knightly hands. Its acceptable basis lay in custom permitting vengeance, redress and self-defense in noble society.") DuBoulay 1983: 71-3. ("[Feuding] was an accepted instrument of politics and law...More often the cause of war was a debt, real or imagined, a territorial claim following a partitioned inheritance, a marriage or an ambiguous grant [of land], or simply the friction between neighbors, especially where mercantile wealth was in some way open to plunder.")

<sup>61</sup>Nicholas 2009: 175. Arnold 1985, 243. ("Feuds in Westphalia were incessant....")

<sup>62</sup>Arnold 1985, 246, "The [ministerials'] use of violence was complicated and extended by the medieval German nobleman's expectation of support in all such actions from his relatives, friends and armed following. Even when the feud originated in the grievances of an individual, it became the collective act of a kindred operating from more than one castle."

<sup>63</sup>Arnold 1985, 242, "Medieval German society paid a high price for the military mentality of the free

Responsibility for the consequences was difficult to pin down. While the feud was a method of enforcing individual rights, its practice blurred the line between the principal perpetrator, on one hand, and his accomplices, on the other.<sup>64</sup> Hence, as a means of settling disputes, private war generated a dilemma for its practitioners. They wanted to reap its benefit – the maintenance of deterrence against any competitor who might violate their rights – while avoiding its unintended cost – indiscriminate mayhem and carnage.

### 5.1 Effect of conflict on the role of courts in dispute settlement

The county courts in medieval Germany served two main functions in civil matters. Nobles conducted and witnessed property transactions there. And they processed disputes over property and jurisdiction there. While some transactions were documented in writing, the disputes were processed orally.<sup>65</sup> The plaintiff lodged his claim orally, the defendant pledged his innocence orally, and the supporters of the opposing sides offered their corroborating oaths orally. When a case raised legal ambiguities, the verdict might seem unjust to one or both sides. If one side was dissatisfied with the verdict – or materially harmed by the other side’s flouting of the verdict – he was obligated to issue an oral “defiance” to the opponent before launching a feud.<sup>66</sup> In the event of a feud, the initial hostilities were often a part of the judicial process itself, intended to force the opponent into a higher court or a compromise settlement.<sup>67</sup> Success in either aim depended on military strength, of course.

Both during the initial court proceedings and the subsequent feud, each side would engage in polemical campaigns against his opponent outside of the court of law proper. Through pamphlets and letters addressed to the regional nobility, he would try to undermine his ad-

---

nobility and of the ministerials in the form of the extensive devastation arising from their feuds.... [F]euds might deteriorate into large-scale knightly marauding.” Arnold 1985, 225, “[M]inisterials pursued destructive feuds which could degenerate into extensive banditry over many years.”

<sup>64</sup>Nicholas 2009: 175.

<sup>65</sup>Reuter 1991: 218. (“Counts still had the judicial functions they exercised in the Carolingian period, but these were increasingly fragmented by grants of immunity; the county court seems to have been more important as a place for the public carrying out of property transactions than as a court proper, though this may just reflect what has survived as written record.”)

<sup>66</sup>After 1186 it became obligatory to issue the defiance in writing as well. Zmora 1997: 28-9.

<sup>67</sup>Zmora 1997: 30.

versary's reputation by accusing him of other legal violations, disloyalty, lying or obstruction of the judicial process itself.<sup>68</sup> A legal dispute thus provoked a battle on three fronts. On the first front, the court of law, the battle began with the trading of claims, denials and counterclaims. On the second front – the audience of regional noblemen who read the letters and pamphlets – there was "a battle for 'public opinion' in which each side sought to undermine the legitimacy of the other's actions"<sup>69</sup>. On the third front, the military field of operations, the battle continued through raids and counter-raids. Events on each front affected strategy and outcomes on the other fronts. In particular, a nobleman's decision whether to defend his legal rights was often driven by considerations of military strength:

"Bishops were installed or elected, counts succeeded their fathers or uncles, and with the titles came clusters of rights that the men in question exercised or defended or exceeded in accordance with their inherited expectations, or according to the emergencies of the day. [Such] *political decisions* thus depended to a great degree upon the force of individual personality.... In medieval Germany, where law and order under the Crown, the *Landfrieden*, and the local jurisdictions were hard to maintain, the *political decisions* taken by such personalities were often of a military nature, because the feud was a normal annual event."<sup>70</sup>

Courts in medieval Germany were thus arenas for duels based on relative power (**implication 1f**). For there were actually three types of court, and the court of battle affected outcomes in the other two courts fundamentally. In the courts of law proper, disputes were tried according to customary oral procedure. These courts alone were rarely sufficient to serve justice, however.<sup>71</sup> In the courts of public opinion, disputants undermined each other's legitimacy in the eyes of their peers through social disapprobation. And in the court of battle, disputants leveraged their military power in the service of their legal goal: to obtain the most favorable settlement and distribution of the rights that were in dispute.

Attempts were made to institute rules that might govern the feuding. The rules were limited in ambition and effect, however. They were designed only to regulate the feuding, not to eliminate it. As long as a feud met the criteria for being *legitimate*, the violence was

---

<sup>68</sup>Zmora 1997: 18.

<sup>69</sup>Zmora 1997: 28.

<sup>70</sup>Arnold 1992: 171-2. Emphasis added.

<sup>71</sup>"... '[O]rdinary' courts were of such limited effectiveness that they required constant attempts to provide peace and justice by other means...."DuBoulay 1983: 85.



not punishable by law.<sup>72</sup> Permitted forms of violence included devastation of the enemy's lands, moveable property and peasants.<sup>73</sup> Uses of feuding that were considered illegitimate included attacks on one's own lord, attacks on members of one's own retinue, and attacks in support of unprovoked robbery, homicide or arson.<sup>74</sup>

The emperors and dukes periodically tried to create a Public Peace or *Landfriede* to end the feuding. However, the *Landfrieden* did not apply to civil disputes or jurisdictional and land disputes covered by feudal law.<sup>75</sup> In any case, there is no evidence that the *Landfriedens* were successful in stopping the violent conflicts.<sup>76</sup> "Since *Landfrieden* were voluntary associations which envisaged the imposition of order by force of arms, they did little more than provide a new mechanism for aristocratic violence under the pretext of pursuing peace-breakers."<sup>77</sup> Contemporary observers concluded that the magnates "had no intention of surrendering their hereditary right and tradition of feud at the emperor's behest."<sup>78</sup>

The main effect of feuding over the long run was to slow the economy and bankrupt many of the lords. To meet the growing need for fighting men, the nobility elevated non-noble freeholders and even unfree peasants into the military class. These men were left unfree, however, creating a class of warriors – the *ministeriales*. While the *ministeriales* served in administrative functions, their main function was military service.<sup>79</sup> In return for service in their lords' feuds, the *ministeriales* received grants of land, office and local jurisdiction in feudal tenure. Over time these grants became hereditary and in some cases

---

<sup>72</sup>Nicholas 2009: 172. Some public authorities restricted or forbade feuding [but mostly cities?]. Nicholas 2009: 173.

<sup>73</sup>The private armies of the lords preferred to attack the adversary's peasants and land rather than his army. (Mortimer 1986: 98.) "Where the ministerials' declared motives approximated to current standards, never closely defined, about legitimate redress through feuds, then the region also faced devastation, laying waste by fire, intimidation and extortion, violent deaths, and the seizure of cattle, treasure, movables, and captives." (Arnold 1985: 246-7.)

<sup>74</sup>Arnold 1985: 246.

<sup>75</sup>"Cases concerned with ordinary civil matters or feudal relationships were supposed to be excluded from the Public Peace and dealt with by the ordinary courts of the local jurisdictional lord." DuBoulay 1983: 77.

<sup>76</sup>Arnold 1985: 242-3. Mortimer 1986: 98. Contrast with DuBoulay 1983: 77, "There is continuing debate about [the Public Peace movement's] effectiveness."

<sup>77</sup>Arnold 1985: 16.

<sup>78</sup>Arnold 1985: 15.

<sup>79</sup>Arnold 1985, 1995: 41, 49, 54.

allodial. Such changes in the terms of tenure occurred at times by legal means, at times by simple usurpation. The growing independence of the ministeriales had negative consequences for their lords, because the ministeriales tended to usurp lands and devastate the economy by engaging in destructive feuds [and erecting illegal tolls, market restrictions and trade barriers].<sup>80</sup> As a result, most of the local lords and many of the regional magnates were in financial distress by the mid-1400s.<sup>81</sup> ?<sup>82</sup>?

## 6. Broader effects of a weak judicial system (implication 1.2)

### 6.1 effect on enforcement options

When a noble lost his case in court, but defied the judgment, the main way of punishing him was through criminal outlawry. By declaring him an outlaw, the court authorized any and all to use force against him and kill him if possible.<sup>83</sup> His main avenues of recourse were either to mount a public relations campaign to defend his actions or, if forced, to fight. His opponent usually responded to public verbal attacks in kind, producing a battle of mutual recriminations between the two parties. Each party distributed pamphlets among the region's nobility – his peers – to restate his claims in the case, defend his own honor and impugn his opponent's honor. Without the right to appeal adverse verdicts to a higher court, such public debates served as a substitute forum of quasi-adjudication. However, the quality of debate was lower than that of any court proceeding, and such debates were often just a prelude to open feuding.<sup>84</sup>

Outlawry was originally used to enforce the criminal laws – those against murder, rape,

---

<sup>80</sup>Arnold 1985, 225, "[They] usurped offices, land and revenues [from their own lords]; they pursued destructive feuds..., and they might even engineer conspiracies ending in the expulsion or murder of their lords."

<sup>81</sup>"almost all of the magnates" Nicholas 2009: 139. Compare with Zmora (DATE), who argues that the nobility became stratified into a wealthy minority and a financially distressed majority.

<sup>82</sup>Arnold 1985: .

<sup>83</sup>Arnold 1985: .

<sup>84</sup>Zmora 2011: .

theft and the like. The Saxon Mirror speaks only of criminal outlawry.<sup>85</sup> As an enforcement tool outlawry was not always effective, however. Some ministerials who were condemned to outlawry continued to violate the law at will and to be sheltered, at times, by their own lord's regional adversary.<sup>86</sup>

In the late medieval period, the use of outlawry as an enforcement tool was extended to civil actions at least in Northern Germany. In Southern Germany, the rules of outlawry were milder, involving the forfeiture of legal protections [= civil outlawry ?].<sup>87</sup>

## 6.2 effect on amount of state power needed to secure compliance with law

Feuding was most prevalent in regions where each lord had his own independent jurisdiction and it did not recognize any higher judicial authority. In these regions the large number of independent jurisdictions raised both the likelihood of disputes and the opportunities to ally with neighboring lords to accumulate power for feuding.<sup>88</sup> Yet lawlessness and feuding were also frequent in the regions that were dominated by a large magnate governing an extensive territory through delegation to vassal knights. These knights engaged in expropriations and feuding with each other, with their own lords, and with their lords' enemies.<sup>89</sup> Thus even the more powerful magnates, ones who possessed regional power bases, could not suppress feuding or secure compliance with the law by force. That their superior power was still inadequate to establish order suggests that power alone was not the foundation of compliance with law when it occurred.

---

<sup>85</sup>As far as I can tell from Dobozý's introduction, glossary and index. Dobozý 1999.

<sup>86</sup>Arnold 1985: 244-5. ("[C]ondemnation to outlawry, or proscription, by the imperial court was more serious [than outlawry by a regional or local court], but not necessarily more effective.")

<sup>87</sup>Nicholas 2009: 175.

<sup>88</sup>DuBoulay 1983: 73. ("Private war was a more frequent and natural condition in those numerous parts of Germany which were politically fragmented.... [T]here were innumerable occasions for dispute and many men SP able to easily than by taking sides.... [T]his knightly capacity to organize for feuding in self-protection, this insecurity, was a consequence of the German realm's frontier character as much as SP in anything else, and with frontiers that is to say which were internal, shifting, uncertain [in part due to the frequent feuding].")

<sup>89</sup>Arnold 1985, 243.? Bosl 1978 [1972]: 301. Arnold 1985, 225. ("Ministerials also risked excommunication, condemnation as outlaws, exile from their province, and the death penalty itself [for violating the law or the rules of the feuding?].")

### 6.3 effect on the fate of weak actors

Weak actors fared poorly in the medieval German judicial system. This was evident in a few ways. First, weak actors lacked the right to initiate many types of cases in court, because standing in court was determined by personal status. Nobles, gentry and ministerials were exempt from the jurisdiction of local courts.<sup>90</sup> So peasants [and freeholders?] could not sue them there. This left local villagers who were disappointed by their lord's conduct with no choice but to resort to self-help.<sup>91</sup> Further up the social scale, many of the regional magnates prohibited their noble vassals and ministerials from appealing court cases to the royal court.<sup>92</sup> This left the vassal or ministerial disappointed by an adverse verdict in his lord's court with no choice but to contract military allies and start a feud with his lord.<sup>93</sup>

Second, weak actors had few legal rights to protect them from expropriation by the stronger actors. Nobles were not prohibited by law from raising rents and usurping peasants arbitrarily, as nobles were in England and France.<sup>94</sup> The rights that weak actors did have were not easily enforced, because they could not sue actors up the social scale or appeal cases beyond the local court. As a result ministerials holding local authority often expropriated their peasants and freeholders.<sup>95</sup> The regional magnates also took advantage of their noble and ministerial vassals.... [check].

---

<sup>90</sup>Dawson 1968: 151. Bosl 1978 [1972]: 305-6. Dobozsy 1999: 19. Nicholas 2009: 169-70, 172.

<sup>91</sup>Cite on peasant uprisings in late medieval Germany.

<sup>92</sup>.

<sup>93</sup>Cite Arnold, Zmora.

<sup>94</sup>Nicholas 2009: 225.

<sup>95</sup>Arnold 1985: "There are plentiful indications that, as officials, the ministerials...oppressed the people committed to their charge." (201-2) "Complaints were frequently leveled at ministerials who usurped from the manorial economy revenues, services, and lands to which they were not entitled. Everywhere the military, fiscal and legal functions delegated to ministerials were open to abuse." (238) "Another irritation for landlords was that ministerials usurped manors and offices to which they had not been appointed at all, or took manorial renders, fines, and services to which they had no right, or extorted payments in kind on grounds of custom which they themselves had invented." (239) "Whether their titles were recognized or not, ministerials were tempted to use their castles as basis for extortions in the surrounding countryside." (241)

## 7. Preference for partisan judges (implication 1.3)

[Reuter analysis.] For disputes among the king's own vassals, he was the ultimate judge. The traditional approach of German kings was to force reconciliation and compromise on the parties, rather than to determine which party was right and which wrong. To secure reconciliation the king deployed a combination of moral suasion and physical threat.<sup>96</sup> Consequently each side in a dispute valued the king for his ability to cajole, shame and coerce the other side into submission and compromise, rather than for his impartiality, knowledge of the law or deeper wisdom. Naturally, the king was endorsed by those whose side he took but rebuked by the others. In cases where their interests were threatened seriously by the king's judicial partisanship, they rebelled and elected an "anti-king" who would take their side in future disputes. Such wrangles led to a number of civil wars between the "royal" party and the "anti-king" party.

Each new king thus arrived on the throne needing to maintain enough political support to *stay* on the throne. If he did not favor his supporters in judicial verdicts, then he would be replaced by an anti-king. So he acted as a partisan judge – "favoring adherents and penalizing dissenters" – and this led to a "traditional line of favoring the enemies of his enemies."<sup>97</sup> Whatever the source of the king's partisanship, the result was that those who were dissenters – or enemies of the king for some other reason – often supported a rival who was more likely to render verdicts in their favor.

Royal judicial partisanship emerged in the 10th and 11th centuries under the Ottonian and Salian kings. It was unusual for them to act with judicial impartiality, and as a result, they were not seen as neutral umpires in court.<sup>98</sup> After many civil wars between kings and anti-kings over occupancy of the throne, the higher nobility instituted an electoral system

---

<sup>96</sup>Reuter 1991: 215. ("...[F]or the political elite, self-help in the form of feud was acceptable, though the king might impose reconciliation.... The ruler was supposed by his presence and his *terror* to discourage evildoers.... [M]ost attempts by kings to establish peace were by example and persuasion. Henry II made all Swabians swear to keep the peace.... Henry III on several occasions... urged not only peace but forgiveness on all present...." Emphasis in original.)

<sup>97</sup>Toch.

<sup>98</sup>Reuter 1991: 216. ("The Ottonians and Salians were themselves often too closely linked to their magnates by ties of marriage, family or friendship to look wholly convincing as umpires, and probably this was not even expected of them.... [In one dispute] Henry contented himself with binding both [litigants] over to keep the peace, but this was unusually impartial.")

to decide who would be king. The system had strict rules governing the membership in the college of electors. It took a century – well into the later medieval period – for the system to develop fully.<sup>99</sup> In time the system was used to solve the problem of judicial partisanship by the king: His settlements in disputes over imperial fiefs had to be endorsed by the electors.<sup>100</sup>

The problem of judicial partisanship also arose lower down the political ladder. Nobles who encountered legal disputes would shop for the magnate judge most likely to render a favorable verdict. If a noble was dissatisfied with one magnate's verdict, he would take his case to another magnate's court.<sup>101</sup> Not surprisingly, situations arose in which each side in a dispute preferred to take it to a different court, each side feared the other had "easy access to the [judge's] ear" in its preferred court, and consequently the two sides could not agree on which court to proceed in.<sup>102</sup>

## 8. Social capital depletion (implication 1.1b)

Germany's stock of social capital became depleted in the medieval period. The process occurred in two phases. From 1000 to 1300, the lower nobility broke away from the crown and the upper nobility and descended into internecine conflicts that destroyed the fabric of lesser noble society. From 1300 to 1500, the lower nobility drew closer to the upper nobility in search of protection, yet conflicts over the terms of the rapprochement escalated in a way that ruptured social relations between the groups. Each phase was marked by the three main indicators of social capital depletion: jurisdictional separation, social disapprobation and a switch from direct dialogue to indirect subversion. Consider each phase in turn.

The disintegration of royal authority in the 11th and 12th centuries created a judicial power vacuum. The lower nobility were left without an ultimate arbiter to settle their dis-

---

<sup>99</sup>Toch, 393.

<sup>100</sup>Toch, 397.

<sup>101</sup>Zmora 1997: 19. ("This state of affairs enabled noblemen to play off one prince against the other and provided them with considerable room for maneuver. In explaining why sentences passed by Franconian law-courts were difficult to enforce, one margravian noble counsellor wrote that 'the knighthood... are provided with many princes; when one is not pleased with the prince's or lord's judgment, he goes over to another prince or lord.' ")

<sup>102</sup>Zmora 1997: 28, 30-1.

putes over land, peasants, markets, tolls and the like. They might have brought their disputes to the higher nobility – their regional overlords – since they were tied to their overlords by bonds of homage, protection and military duty. Because the magnates' courts were not effective, however, the lesser nobles fell into violent feuding over their internecine disputes. These disputes destroyed the fabric of lower noble society in three ways. First, the nobles saw little value in remaining within the jurisdictions of the regional magnates, because their judicial services provided such little value. So the lesser nobles established their own jurisdictions – independent of the regional magnates and each other. This occurred throughout western Germany (Franconia, Westphalia and Swabia). This jurisdictional separatism compounded the social distance created by feuding, however, because it reduced further the opportunity for dialogue over disputes.

Second, the feuding was often accompanied by pamphlet wars in which each side in a dispute launched verbal attacks and denunciations against the other side. The disputes, feuds and verbal polemics were raised to an unprecedented level by a civil war which began in 1073 and lasted for fifty years.<sup>103</sup> The combined strategy of feuding and social disapprobation continued to be used throughout the medieval period.

”Feuds were nearly always part of the political process by which they were in turn regulated. Hence they were often as much ethical as violent. Opposing parties invariably set much store on ‘public opinion’ and legitimacy: they used ink no less than sword and torch. Naked power was by and large inconceivable.”<sup>104</sup>

Third, nobles undermined their adversaries through indirect means of subversion. Lords poached military vassals from each other.<sup>105</sup> Lords threatened the peasants of other lords with pillaging attacks to compel them to withhold their rents from their lords.<sup>106</sup> Lords took out loans from vassals and then convicted them on false charges to avoid having to repay the debts.<sup>107</sup> Lords expelled vassals from their lands without good legal cause or court

---

<sup>103</sup>Leyser 1968: XX.

<sup>104</sup>Zmora 1997: 34.

<sup>105</sup>Arnold 1992: 78.

<sup>106</sup>Zmora 1997: 108-9.

<sup>107</sup>Leyser 1968: xx.

warrant.<sup>108</sup> Vassals rendered homage to multiple lords to subvert the capacity of any one lord to subordinate his vassals to territorial control and jurisdiction.<sup>109</sup> When summoned to court to answer charges, nobles at all ranks defied the summonses and holed up in their castles with their followers – either to evade conviction when guilty or to evade wrongful conviction when innocent.<sup>110</sup> In sum, each noble maneuvered subversively from his own local jurisdiction to undermine his opponents through cunning deeds and damaging words. The combination of jurisdictional separatism, social disapprobation and indirect subversion reflected a deep rupturing of the social fabric of noble society.

By 1150 there was increasing mutual distrust and social distance between the lower nobility and the regional overlords from whom they had separated jurisdictionally. By the 1200s, however, some of the lesser nobles were seeking protection from the higher nobility.<sup>111</sup> And by the 1300s some of the lesser nobles started to band together to defend themselves and increase their bargaining power vis-à-vis the regional magnates. They formed knightly societies that served multiple functions. In some cases the society established an internal jurisdiction to resolve disputes among its members – independent of the jurisdiction of the regional overlord whose court was supposed to serve that function.<sup>112</sup> In many cases the knightly society was intended to oppose the region's overlord in both word and deed.<sup>113</sup> Some of the societies feuded with regional overlords over disputes.<sup>114</sup> Again the combination of jurisdictional separatism, social disapproval and subversion reflected a rupture in the fabric of noble society. Moreover, these were inherent features of the world of power, conflict and high-cost deterrence. Each lord resorted to subversion, verbal disapprobation and feuding because they were necessary tools for defending his rights in a world of separate jurisdictions with no effective third-party adjudicators. In turn the conflicts induced competitions in raising relative power that simply intensified the separatism, disapprobation and subversion.

---

<sup>108</sup>von Repgow 1235 [1999]: 157.

<sup>109</sup>Zmora 1997: 96, 116-117. Arnold 1985: 240-1. Arnold 1992: 101.

<sup>110</sup>Arnold 1985: 240-2.

<sup>111</sup>.

<sup>112</sup>.

<sup>113</sup>.

<sup>114</sup>.



## 9. Higher conscription, tax and corruption rates (implication 1.1c)

There were many independent lordships in medieval Germany. Each lord held a sovereign jurisdiction, so he held final authority to set the military conscription rate on his noble vassals and the tax rate on his peasants. By prohibiting appeals from his law court to higher courts, he could violate his vassals' and peasants' rights with impunity. The only constraints on his extraction and corruption were the capacities of his subjects to rebel with violence or to switch allegiance to a neighboring lordship, perhaps with the aid of neighboring lords and their subjects. As a result, the conscription, tax and corruption rates were relatively high in medieval Germany.

This is evident from both contemporary sources and later histories. Testifying to the high rate of military conscription, chroniclers observed that almost every lord had his castle from which he swooped with his armed retinue to devastate the lands of neighboring lords. The problem in the chroniclers' view was how to *disarm and demobilize* these armed lordships.<sup>115</sup> Modern historians have concluded rather that German lords *needed* to maintain large private armies at all times simply to defend their rights.<sup>116</sup> With respect to taxation, lords were not legally prohibited from raising rents and enserfing SP peasants arbitrarily. Hence they often taxed and expropriated their subjects unilaterally, and the victims had no legal recourse to any higher judicial authority. Lords also extorted rents from merchants through tolls and ransoms.<sup>117</sup>

The patterns were quite the reverse in England. The number of knights who maintained arms and were fit for battle decreased steadily from the 12th to 14th centuries. As fewer

---

<sup>115</sup>.

<sup>116</sup>Arnold 1985: 14. DuBoulay 1983: 71-3. ("[L]arge numbers of German knights and their followers spent their time fighting.")

<sup>117</sup>Bauml 1969: 156. ("The commercial traffic engendered by the productivity of the relatively well-governed towns and overland trade with foreign countries were continuously endangered by the social chaos prevailing in the countryside. The disruption of feudal order placed countless knights, even holders of a single castle, in an independent position. With the relative decrease in the value of land and increase in the value of money the caravans of merchants on the roads, and their ships on the rivers, became objects of attraction for the robber-barons. The countryside, particularly along the Rhine and Danube, was dotted with their strongpoints, toll-gates, and navigational obstacles designed to aid in the extortion of money or goods from the merchant as the price of his passage. There was no law which could provide shelter to the merchant. He had to be prepared either to pay heavy sums to the robber-barons along the way as ransom for himself and his goods, or have his caravan accompanied by an armed convoy.")

disputes were settled through conflict and the economic returns to investments in agriculture increased, the lower nobility increasingly turned from military to civilian pursuits.<sup>118</sup> From the perspective of the royal government and the higher nobility, the problem was to insure the availability of mounted knights in the face of this structural shift. In [date] the government issued a statute, the *Assize of Arms*, which required [ ]. In the fiscal arena, the English lords still extracted sizable rents from their peasants. Yet [as the lords earned their income increasingly from agricultural markets – the turn to "high farming" – they gained an incentive to curb the tax rate before it got so high as to inhibit the peasants' productivity; as a result, rental rates were constrained to an extent.]<sup>119</sup> Moreover, peasants had legal recourse to the royal courts in the event that their lord raised rents illegally or exorbitantly.<sup>120</sup>

## 10. Effect of weak judicial system on scale and scope of the state (implication 1.5)

---

<sup>118</sup>Coss?

<sup>119</sup>.

<sup>120</sup>.

## **CHAPTER 15**

### **Worlds of power and law in the democratization battles of the 19th century**

Democratization efforts in Europe in the 19th century produced a variety of outcomes. In the largest countries the outcomes ranged from substantial progress in Britain and France to welfare state authoritarianism in Germany to agrarian populist dictatorship in Spain. The Contractual Realist theory explains these outcomes differently than previous theories did. In section 1, I outline the explanation offered by the Contractual Realist theory. In section 2, I provide brief case studies of Britain, Spain, and Germany.

#### **1. Anchor based governance and democratization**

The Contractual Realist theory provides a straightforward explanation of democratization successes and failures. The theory predicts that democratization efforts will succeed in political systems governed by anchor contracts and anchor based enforcement. But efforts will fail in systems where anchor based governance has broken down due to the emergence of contentious issues. These predictions are corroborated by the evidence from 19th-century England, France, Germany and Spain.

Table 15.1: Conditions for democratization in 19th century Europe: Consensus on legal regimes for agricultural and industrial property rights

Consensus on legal regime for <u>industrial</u> property rights	Britain (1840-1860)	Britain (1820-1840) Britain (1860-1900) France (1820-1845) France (1875-1900)
Dissensus on legal regime for <u>industrial</u> property rights	Germany (1815-1900...) Spain (1815-1900...)	France (1845-75)
	Dissensus on legal regime for <u>agricultural</u> property rights	Consensus on legal regime for <u>agricultural</u> property rights

## 2. Brief case studies

### 2.1 Britain

Democratization began well with a Reform Act of 1832.<sup>1</sup> In the 1840s, however, the emergence of the Chartist Movement ruptured the existing anchor contract with respect to land ownership. The Chartists called into question the existing legal regime for landed property and claimed the right to receive a substantial land reform that would redistribute significant

<sup>1</sup>In writing this case study I relied on several secondary sources: Garrard, Tolz, and White (2000), Claeys (2001), Garrard (2002), and Roberts (2003).

amounts of land from its current owners. The current owners responded in defense of their existing ownership rights. The outcome was a competition in raising relative power. Current owners opposed any further expansion of democratic voting rights, because it would give voting power to elements of society that favored the Chartist Movement, and they might use it to enact property redistributions at the expense of current owners. Efforts toward further democratization could not succeed in this environment. No further expansions of voting rights were enacted between 1832 and 1870.

By the 1870s the Chartist Movement had receded from political prominence. Its calls for land reform dropped off the political agenda for most segments of the population. Once again the existing legal regime for landed property was generally accepted. Current owners did not have to fear that a further expansion of voting rights would give political power to elements of society that would use it to redistribute landed property. Settlement of the land rights issued opened the way for further expansions of voting rights through acts of parliament.

## **2.2 Spain**

Throughout the 19th century Spanish society was deeply divided over the issue of land rights.<sup>2</sup> The dispute arose from the constitution of 1812 which was enacted during the Napoleonic wars under the umbrella of French revolutionary hegemony. The Constitution emancipated the peasantry from feudalism and established a relatively liberal regime. While the emancipation eliminated the worst abuses of feudalism, it also raised the question of how far peasant rights extended in agrarian society. In time this generated a dispute over land ownership, because some peasant leaders favored changes in the agrarian legal regime that would amount to a redistribution of land in favor of landless peasants.

This dispute intersected with constitutional politics. The vast majority of Spanish society – those outside the aristocratic, commercial and administrative classes – had never enjoyed access to political participation and power. The constitution of 1812 granted access. After the Napoleonic wars ended, however, a conservative reaction replaced the constitution of

---

<sup>2</sup>In writing this case study I relied on several secondary sources: Carr (1982), Lynch (1981), Ringrose (1996), Shubert and Alvarez Junco (2000), and Smith (2000).

1812, in part, to prevent the land redistribution it might allow. Following this reaction Spanish politics devolved into a competition in raising relative power – with the matter of land rights being one of the most contentious issues dividing Spanish conservatives and liberals in that competition. Each side’s goal was to increase its power by enough to achieve its goal. The liberals’ goal was to expose the illegitimacy of the existing legal regime for land rights and enact a land redistribution through political action. The Conservatives’ goal was to prevent the kind of land redistribution envisioned in the constitution of 1812. The contention over land continued unabated throughout the 19th century.

Spain experienced a number of changes in constitutional regime and leadership over the course of the century, swinging back and forth between conservative and more moderate regimes and leaders. Yet Spanish conservatives opposed any expansion of voting rights because it might result in the enactment of a land redistribution (just as the conservatives in England had opposed it at the height of the Chartist movement in mid century). As a result, there were no major expansions in voting rights of the kind enacted in England in the later 19th century. Although mass suffrage was granted for elections to some government positions, there was no move toward mass suffrage in elections to a national parliament exercising unlimited popular sovereignty and legislative authority.

## **2.3 Germany**

German society also experienced deep division over the matter of land rights.<sup>3</sup> In the 1830s and 1840s the number of landless peasants and workers swelled for several reasons. On the agrarian side, large numbers of peasants were landless due to the long history of feudal subordination and expropriation, especially in the leading German state, Prussia. After 1815 agricultural production was rationalized on the basis of new technologies and financial practices, but the negative side effect was to render landless peasants redundant on the rural estates, unemployed and homeless. Their migration to the cities and factories lowered labor costs and facilitated German industrialization. Yet the negative side effect was to create a large urban working class with neither land ownership rights nor workers’ rights.

---

<sup>3</sup>In writing this case study I relied on several secondary sources: Snell and Schmitt (1976), Blackbourn and Eley (1984), and Blackbourn (2003).

Some elements of the peasant and urban working classes advocated fundamental changes in the legal regimes for landed property and industrial assets. The conservative elements of Prussian society therefore expected that if voting rights were extended to landless workers, they would use their new political power to redistribute wealth and income to themselves. This was one of the most contentious issues dividing German conservatives and socialists in the 19th century. Each side jockeyed for power to defend and advance its interests on this issue throughout the century. The large, radical segment of the Prussian working class never abandoned its quest for fundamental changes in the legal regimes for landed property and industrial assets. In this sense the dispute was never resolved. Consequently, there were no major expansions of voting rights of the kind enacted in England in the 19th century. While mass suffrage was granted for elections to certain government positions, there was no move toward mass suffrage in elections to a national parliament exercising unlimited popular sovereignty and legislative authority.

At the same time, Prussian workers and peasants were granted certain employment and welfare benefit rights in the second half of the century. This observation is consistent with the Contractual Realist theory. Worker rights constituted an anchor contract. Workers had both an obligation to work for their pay and the right to exceptions under certain circumstances that were beyond their control (such as sickness, old age, death of the family's male wage earner). As long as the Prussian welfare state could distinguish between intentional and unintentional causes of a worker's need for one of these exceptions, the anchor contract was enforceable by anchor based enforcement methods. And labor peace remained feasible, at least on the factory floors, if not on the political parade grounds.

The Prussian case thus provides a double demonstration of the explanatory power of the Contractual Realist theory. On one hand, the economic and social interests of *workers* were *well* represented by their advocates in the Prussian welfare state. On the other hand, the political interests of *citizens* were not well represented in the Prussian parliament. The Contractual Realist theory explains both the success of democratization at the level of worker representation and the failure of democratization at the level of citizen representation. The success was due to the capacity of the Prussian state to form an anchor contract in the area of worker rights. The failure was due to the incapacity of the Prussian state to form an anchor

contract and a justice bargain in the area of land rights and industrial asset ownership.



Part V

From the path of law to the path of  
power in international politics: The  
outbreak of hegemonic wars

## CHAPTER 16

### Theories of hegemonic war: A review and critique

Since 1550 Europe has been rocked by a series of catastrophic wars. Scholars agree on the basic facts of these wars. Each war erupted from an international dispute. Once the war started one state expanded territorially and tried to establish hegemony over other states. Those states responded by forming a coalition to beat back the expansionist state and establish a stable peace again. These wars were later dubbed "hegemonic wars" in recognition of the fact that in each war one state tried to establish hegemony over other states.

While the historical facts of hegemonic war have not changed, the explanatory puzzles identified by scholars have changed many times. Since theorizing about hegemonic war began, there have been at least four distinct generations of thinking about its causes, conduct and consequences. Each generation identified a different explanatory puzzle as the main one to be resolved in explaining hegemonic war. In the first generation (1600-1945), theorists were puzzled simply by the question of why these wars occurred. Their answer was that they were caused by relative power factors. Every succeeding generation accepted this answer as correct and focused on the task of solving deeper puzzles within this school of thought. In the second generation (1945-1985), theorists were puzzled by the question of *why and how* relative power factors cause war. Their answer was that a favorable shift in relative power might give one state the expectation of reaping a net gain from waging a war. In the third generation (1985-1995), theorists were puzzled by the question of how domestic political systems could permit states to wage a war that would impose large net losses on both sides. Their answer was that some domestic actors have an innate preference for power rather than security, and some domestic systems suffer from distortions in the way they process information about relative power, formulate strategy based on relative power expectations, and distribute the war costs arising from relative power threats. In the fourth generation

(1995-2005), theorists were puzzled by the question of why two rational states – ones that do not suffer from any domestic distortions – would ever wage a war when they could settle a dispute peacefully and avoid the net war losses on both sides. These theorists gave two answers. On one hand, uncertainty about relative power might give one state an incentive to bluff that its power has increased and another state an incentive to initiate a war to deter such bluffing. On the other hand, an impending shift in relative power might give one state the expectation of reaping a net gain from a war and another state an incentive to initiate war first to prevent the power shift from occurring in the first place.

In each generation the main assumption was that Europe's hegemonic wars were caused by relative power factors. Over time this idea grew into a foundation stone for the explanation of *all* wars. As one theorist wrote,

"The [conventional] realist theory of international political change is based on what can be called the law of uneven growth.... According to [conventional] realism, the fundamental cause of wars among states and changes in international systems is the uneven growth of power among states. Realist writers from Thucydides and Mackinder to present-day scholars have attributed the dynamics of international relations to the fact that the distribution of power in an international system shifts over a period of time.... Throughout history the primary means of resolving the disequilibrium between the *structure* of the international system and the *redistribution of power* [in the international system] has been war, more particularly, what we shall call a hegemonic war."<sup>1</sup>

By expanding this explanation of Europe's hegemonic wars into an explanation of all wars, Conventional Realists made an important move. They claimed, at least implicitly, that whatever causal mechanisms they might identify in Europe's hegemonic wars would lie at the root of the rest of the wars in human history. In other words, they saw Europe's hegemonic wars as *paradigmatic cases* that teach general lessons about war. The aim of the Conventional Realist paradigm was to draw these lessons by producing a series of theories that explained the myriad facts of Europe's hegemonic wars.

---

<sup>1</sup>Gilpin (1981, pp. 94, 197). Underlining and italics added for emphasis. Also see Morgenthau (1978 [1948]).

In this way the axiom that wars are caused by relative power factors became the yardstick by which all theorizing was measured. This development followed logically from the widespread academic agreement on three basic facts of international life. Power is a driving factor in international relations. The possibility of conflict is ever present in international relations. And hegemonic wars are the paradigmatic examples of conflict. Working from these facts, the axiom that wars are caused by relative power factors became central to a wide range of research on international relations and related topics. While some studies accepted the axiom and other studies focused on the apparent exceptions to it, almost all studies assumed that it was the essential axiom that needed to be accepted or else refuted. This axiom of the Conventional Realist paradigm became the industry standard.

At its deepest level the Conventional Realist paradigm is based on just three hard-core assumptions. States *always* divide material assets among themselves according to relative power, in peace as well as in war. Conflict is *always* likely enough that states need to maintain military power at all times, in all places, on all issues. And each state is *always* jockeying for a relative power advantage – if not to expropriate assets from other states then to avoid becoming the target of expropriation by other states. These assumptions generate the positive heuristic that drives theory development in this paradigm: Write theories in which conflict is caused by relative power factors. Conventional Realists have developed five protective belts of theory along these lines.

- Theories of war due to shifts in relative power;
- Theories of war due to uncertainty about relative power;
- Theories of war due to an actor's preference for power rather than security;
- Theories of war due to a relative power advantage conferred by military technology;  
and
- Theories of war due to domestic distortions in the way that a state forms preferences about relative power, processes information about relative power, and allocates war costs arising from relative power shifts.

In the following sections, I evaluate these theories against the Lakatosian standards established in the previous chapter. I describe the anomalies encountered by these theories and discuss the measures taken by their defenders to save the theories. I show that the defenders amended the theories repeatedly by adding new *ceteris paribus* conditions, auxiliary hypotheses and observational theories that might account for the anomalies. I evaluate their amendments in the light of the new school of thought proposed in this study, Contractual Realism, and conclude that their amendments were flawed. They constituted degenerative problemshifts. Finally, I show that Contractual Realism offers a progressive problemshift according to the same standards.

## 1. The first generation (1600-1945): Theories of war due to discretionary shifts in relative power

### 1.1 Shifts in relative territorial power

The earliest theory of European wars due to relative power factors focused on shifts in territorial power. The theory was developed to explain cases in which (1) a great power expanded territorially by conquering minor states in its region, (2) another great power or coalition of powers felt threatened by its expansion, and (3) they launched a "counterbalancing" war to reduce the expansionist power to its original size and "restore the balance". The main hypothesis is that one state's territorial expansion poses a threat – in and of itself – because it increases the state's power to the point where it can expropriate other states.

"[In 1612] Sir Francis Bacon in his essay 'Of Empire' was outlining specific policy guide-lines for maintaining a balance of power. Bacon urged that 'princes do keep due sentinel, that none of their neighbours do overgrow so (by increase of territory, by embracing of trade, by approaches or the like), as they become more able to annoy them than they were'. Bacon argued that Henry VIII, Francis I of France and the Emperor Charles V had followed such a policy with respect to each other with the result that 'none of the three could win a palm of ground, but the other two would straightways balance it, either by confederation, or if need were, by a war; and would not in any wise take up peace at interest'. (Bacon

quoted in Maurseth, 1964: 121). Bacon went on to argue that danger to the balance of power justified preventive war."<sup>2</sup>

After the settlement of the first few hegemonic wars – in 1609, 1648, 1714 and 1815 – observers looked at the wreckage and sought to explain why the catastrophe had happened. Their main hypothesis was that each war arose from a shift in the balance of territorial power.<sup>3</sup>

This hypothesis seems to be supported by the historical evidence from four of the seven hegemonic wars in Europe between 1550 and 1950:

1. **The Thirty Years War (1618-48).** In the 1620s Austria conquered some minor principalities in Germany, and in response Sweden and France launched a counterbalancing war.
2. **The Nine Years War (1688-97).** In 1688 France conquered some minor principalities in western Germany, and in response England, the Dutch Republic, Savoy, Spain and some other German principalities waged a counterbalancing war.
3. **The War of the Spanish Succession (1700-1714).** In 1700 France conquered the Duchy of Milan and the Southern Netherlands, and in response Austria launched a counterbalancing war. After France committed further aggrandizement, England and the Dutch Republic entered the war on Austria's side.
4. **World War II (1939-45).** Between 1936 and 1939 Germany conquered Austria, Czechoslovakia and Poland. In response France and Britain declared a counterbalancing war and the United States eventually entered on their side.

The hypothesis is not supported by the historical evidence from the other three hegemonic wars in this period, however. In each case the war erupted *before* any of the great power opponents had conquered any state, large or small.

1. **The Eighty Years War (1568-1648).** This war erupted before Spain conquered any lands in Northern France, Western Germany or England.

---

<sup>2</sup>Underlining added for emphasis. M. Sheehan 1996, pp. 36–7.

<sup>3</sup>See J. J. Sheehan (1989) and Haslam (2002) for reviews of the earliest theories of great power competition and conflict in European history.

2. **The French Revolutionary Wars (1792-1815).** This war erupted before France conquered any lands in Italy, Germany or Belgium.
3. **World War I (1914-18).** This war erupted before Austria conquered Serbia or Germany conquered Belgium.

In response to these anomalies, theorists switched the focus from shifts in territorial power to shifts in military manpower.

## 1.2 Shifts in relative military manpower

Here the main hypothesis is that when one state increases its military manpower – even within its own borders – this poses a threat in and of itself. For again, it increases the state's power to the point where it can expropriate other states. The classic statement of this hypothesis was offered by Hans Morgenthau in *Politics Among Nations* (first published in 1948).

”[Nation] A tries to increase its power in relation to [Nation] B to such an extent that it can control the decisions of B and thus lead its imperialistic [expropriation] policy to success. B, on the other hand, will try to increase its power to such an extent that it can resist A's pressure and frustrate A's policy, or else embark upon an imperialistic policy of its own.... This balancing of opposing forces will go on, the increase in the power of one nation calling forth an at least proportionate increase in the power of the other, until the nations concerned change the objectives of their imperialistic policies – if they do not give them up altogether – or until one nation gains or believes it has gained a decisive advantage over the other. Then either the weaker yields to the stronger or war decides the issue.”<sup>4</sup>

This hypothesis seems to explain the three cases that were anomalies for the previous theory. In each case one great power increased its military manpower *within its own borders*, but this was seen as a threat by another great power or coalition of powers.

---

<sup>4</sup>Underlining added for emphasis. Morgenthau 1978 [1948], pp. 173–4.

1. **The Eighty Years War (1568-1648).** In the 1570s Spain deployed a large army to its province of Southern Netherlands to fight the secessionists in its other province of Northern Netherlands. This force was perceived as a threat to the English homeland, so England entered the war on the secessionists' side.
2. **The French Revolutionary Wars (1792-1815).** In 1791 it was reported that Austria was raising an army to invade France. When these reports arrived in France, the average Frenchman perceived it as a threat to the homeland – and French policy toward Austria hardened. Within a few months Austria was *actually* raising an army and France soon declared war on Austria.<sup>5</sup>
3. **World War I (1914-18).** In the summer of 1914, each of the continental powers mobilized its army in order to be prepared in case its opponent *might* initiate a war. Yet each side perceived the other side's mobilization as evidence that it *would* initiate a war. So each side initiated war first to avoid being overrun.

Although this theory seems to explain these cases, in fact these cases are anomalies for the theory. The theory assumes that one state's increase in military power is perceived as a threat to commit expropriation against another state. The first state's threat to commit expropriation is only credible, however, if it expects to reap *net gains* from committing expropriation. If it expects to reap net losses, then it will not initiate a war of expropriation in the first place. And it will seek a quick end to any war initiated by the other state (with the goal of preempting it from committing expropriation). Thus, a war of this kind is expected to be long and costly *only* if the first state expects to reap a net gain from the war. If the historical evidence shows that the war was long and costly, but the first state reaped a net loss, then such evidence is an anomaly for the theory. In fact, each of the three wars listed above was long and costly, but the state whose military manpower had increased prior to the war suffered huge net losses in the end. Thus these cases are anomalies for this theory.<sup>6</sup>

---

<sup>5</sup>Austria also declared war on France for similar reasons, before even receiving news of the French declaration of war.

<sup>6</sup>It might seem unlikely that a state would fight a long and costly war and still reap a net gain from it. There are examples however. The Boer War was long and costly for Britain but yielded a net gain by securing control of southern Africa. Some of Britain's wars in South Asia in the 19th century were long and costly, but they yielded a net gain by securing the British Empire in India. The Indian Wars fought by the United States government in the 19th century were long and costly, but they yielded a net gain by securing



In response to these anomalies, Conventional Realists of the next generation shifted the focus from *discretionary* shifts in relative power due to deliberate policy decisions to *non-discretionary* shifts in relative power due to factors beyond anyone's control. Before turning to these theories, however, consider some questions raised by the first generation theories.

### 1.3 Questions asked and answered by Conventional Realists to generate further theoretical development in their tradition

1. Assuming that power shifts are the root cause of war, then what is it about a power shift that *motivates* the rising or declining state to initiate war?
2. What are the *returns to power* for the rising state? And for the declining state?
3. What are the *returns to war* for the rising state? And for the declining state?
4. How should theorists conceptualize and measure power, the returns to power, and the returns to war in order to answer these questions? What observational theory should they use?

In a moment I will identify the observational theories and *ceteris paribus* conditions that Conventional Realists developed to answer these questions and I will evaluate whether they constituted progressive problemshifts.

### 1.4 Questions not asked by Conventional Realists

1. Are shifts in relative territorial power or military manpower a *root cause* of war or only an *intervening variable* that mediates between some deeper root cause and the outcome of war initiation?
2. If these kinds of power shift are only an intervening variable, then how does the deeper root cause generate a power shift as an epiphenomenal event in the process of causing a war?
3. How should theorists conceptualize and measure power, the returns to power, and the returns to war to answer these questions? What *new* observational theory should they use? How does it differ from the observational theory used by Conventional Realists

---

much of North America for the United States. The medieval period also offers examples. The wars of the Germanic tribes against the Slavs were long and costly, but yielded net gains in territory and wealth. The reconquest of Spain was long and costly, but also yielded net gains in territory and wealth.

to conceptualize and measure these factors?

These questions are answered by the theory that I develop and test in this study: Contractual Realism. The answers are summarized briefly in the remainder of this chapter and provided in Parts II, III and VII of this study.

## **2. The second generation (1945-1985): Theories of war due to non-discretionary shifts in relative power**

Theorists of this generation were no longer puzzled by the question "Why did hegemonic wars occur?" They accepted the previous generation's answer: These wars occurred because of shifts in relative power between states. The new puzzle was: "Why do shifts in relative power cause war?" The second generation theorists tried to solve this puzzle by starting with the fact that hegemonic wars were the most destructive wars in modern European history.<sup>7</sup> They derived two simple intuitions from this fact. First, no rational state would be willing to expend so much blood and treasure on a war unless it expected to gain a great deal in return. Second, only a particularly large shift in relative power between the largest states in the system could give a state the expectation of gaining so much. This perspective seemed to explain why the largest states in each war:

- conquered each other as well as smaller states,
- manipulated the political systems of the conquered lands for their own benefit, and
- tried to replace existing international institutions with new ones they could control.

Second generation scholars developed five main theories that would explain these patterns as an effect of shifts in relative power. Each theory focused on a different role of state power in international security competition. While the theories contradicted each other on the question of why power shifts cause war, they shared an implicit assumption that second generation theorists did not call into question. In each case at least one state went to war because its power had increased to the point that it expected to reap *net gains* from a war. To put this assumption another way: At least one state expected that because its power had increased, the *returns to state power* would yield a *net benefit*. Each theory focused

---

<sup>7</sup>Levy 1985.

on a different source of returns to state power and claimed that it was the *driving* source of returns to power in international security competition. The theories disagreed on several points:

1. What is the level of relative power that causes war?
2. What is the source of a state's benefits from war? Do the benefits accrue from imposing direct rule on other states or from taking control of international institutions?
3. What is the source of the incentive to initiate war? Does the incentive arise from an increase in the state's own power or a decline in some other state's power?

The first two theories agreed that the benefits from war accrue from imposing direct rule on other states. They disagreed on the level of relative power that causes war.

- Balance of power theory assumed that peace is preserved by a rough balance of power between states. War erupts when one state grows from a position of rough equality with others to a position of superiority over them.<sup>8</sup>
- Power transition theory assumed that peace is preserved by one state's superiority over others in relative power. War occurs when another state grows from a position of inferiority to a position of rough equality with the leader.<sup>9</sup>

The next two theories assumed that the benefits of war accrue from taking control of international institutions rather than imposing direct rule on other states. These theories agreed on the level of relative power that causes war. War occurs when a second tier state grows from a position of inferiority to a position of rough equality with the leading state in the system. These theories disagreed, however, on the source of the incentive for war.

- Long cycle theory assumed that war ensues from a competition among numerous rising states to take control of international institutions.<sup>10</sup>
- Hegemonic stability theory assumed that war erupts because the leading state's incentive to maintain the peace (to reap the gains from controlling international institutions)

---

<sup>8</sup>Morgenthau 1978 [1948]; Waltz 1979; Niou and Ordeshook 1986; Kennedy 1987; James 1995.

<sup>9</sup>Organski 1958; Organski and Kugler 1980; Gilpin 1981; Kugler and Lemke 1996.

<sup>10</sup>G. C. Modelski 1978; G. Modelski 1987; Goldstein 1988.

declines with its falling relative power.<sup>11</sup>

The seminal contributions of the second generation were made from 1945 to 1985. Yet these theories continue to be accepted, used and further developed by scholars to this day under new names. For example, the main assumptions of power transition theory continue to be developed under the rubric of "unipolar stability" theory.<sup>12</sup> The main assumptions of hegemonic stability theory continue to be developed under the rubric of "international hierarchy" theory.<sup>13</sup> And balance of power theory continues to be developed under the rubrics of "soft balancing" theory and "unipolar illusion" theory.<sup>14</sup>

While the second generation theories made substantial progress in charting the many roles of power in state strategies and international outcomes, they suffered from many logical and empirical problems. These problems fell into three categories. First, there were problems with the assumption that one state expected to reap net gains from each hegemonic war. Second, there were problems with the assumption that a second state initiated war to prevent the first state from reaping net gains. Third, there were problems with the assumption that relative power shifted before each hegemonic war. The next three sections treat these topics.

## **2.1 Problems with the assumption that one state expected to reap net gains from each hegemonic war**

This assumption has three flaws. First of all, the intuition that a state would only incur large war losses in return for even larger gains does not withstand scrutiny. When a state really grows in relative power to the point that it can reap net gains from war, the ensuing conflict will not be long and costly. It will be short and cheap. For the state that is targeted for attack will have nothing to gain by fighting. This is why all of the other wars in European and colonial history that were driven by a relative power advantage were neither long nor

---

<sup>11</sup>Charles Poor Kindleberger 1973; S. D. Krasner 1976; Keohane 1984; Snidal 1985; Webb and S. D. C. Krasner 1989; Brawley 1993; Lake 1993.

<sup>12</sup>Wohlforth 1999; Wohlforth 2007.

<sup>13</sup>Lake 2007.

<sup>14</sup>Layne 1993; Layne 2006.

destructive.<sup>15</sup>

Second, the intuition that large war costs become worthwhile in exchange for even larger gains cannot explain the actual conduct of hegemonic wars in European history. For suppose that a state initiates a war of conquest expecting net losses in the short run but net gains in the long run after the war is concluded and the conquests are consolidated. This logic still does not explain why the state would pursue conquests involving such large war losses that it must choose between regurgitating its conquests, on one hand, and running its war machine into the ground and being over-run militarily, on the other hand. Either outcome belies the logic that short-run war costs are compensated by long-run gains. In fact, every conqueror in every hegemonic war reached this fateful impasse. This fact requires explanation. Yet the second generation theories cannot make sense of it, because they assume that the hegemony seeker expected to reap net gains from territorial expansion.

Third, these theories conflate the *causal* roles of power in international security competition with the *other* roles of power, namely, as an intervening variable, a subsidiary variable or an epiphenomenal variable. To see this point, consider the possibility that a state's territorial power might be an intervening variable. From this perspective there are two very different factors that might cause a state's territorial power to expand. On one hand, territorial expansion can be an optimal means to *maximize a state's net gains* after an increase in its underlying sources of power (such as population or military technology). On the other hand, territorial expansion can be an optimal means to *minimize a state's net losses* after the eruption of an international dispute over valuable (non-military) assets that are unrelated to state power. In this case the expansion of a state's territorial power is an intervening variable that is driven by deeper causes *which are unrelated to power*: International disputes over valuable (non-military) assets. My theory employs this approach to resolve the puzzle of territorial over-expansion in hegemonic wars in European history.

My theory thereby provides new answers to the questions on which the second generation theories disagreed.

1. Question: *What is the level of relative power that causes war?*

---

<sup>15</sup>See the second chapter in Part VII of this study for a formal proof of this proposition. For historical analyses of power-based conquest within Europe, see Greengrass (1991), and beyond Europe, see Parker (1988) and Abernethy (2000).

My theory's answer: War is not caused by a change in the level of relative power between states. War is caused by the development of a new domestic activity in one state that increases its income but generates international spillover effects that reduce a second state's income. The second state has an incentive to initiate war against the first state to force it to stop its new domestic activity. The second state has this incentive regardless of whether the spillover problem causes relative power to shift in its favor or against it (or not at all).

2. Question: *What is the source of a state's benefits from war? Do the benefits accrue from imposing direct rule on other states or from taking control of international institutions?*

My theory's answer: The second state, e.g. the one suffering the income losses from the spillover problem, has an incentive to adopt whichever strategy enables it to avoid these income losses, whether it be imposing direct rule on other states or taking control of international institutions.

3. Question: *What is the exact source of the incentive to initiate war?*

My theory's answer: The second state has an incentive to impose more war costs on the first state than it gains from its new domestic activity that is causing the spillover problem (e.g. to induce the first state to stop its domestic activity). The first state has an incentive to impose more war costs on the second state than it would incur by retracting its demand to stop the activity and accepting the income losses from the spillovers passively (e.g. to induce the second state to retract the demand and accept the income losses passively).

This approach provides a completely new answer to the question of why one state would expand territorially by a large margin – and the question of why another state would fight back. It has nothing to do with relative power. It is all about income, or more precisely, the motive of income loss avoidance. Each state is driven to wage war by the expectation that it would incur even greater income losses by settling for peace instead. Specifically, if the second state were to settle for peace, it would have to continue suffering the income losses from the spillover problem. If the first state were to settle for peace, it would have to give up its new domestic activity and lose the income gains from that activity. At the same time, once the war has begun, relative power is the primary factor driving state strategies and international outcomes. How can it be that relative power plays *no* role in each state's decision whether to

go to war over the dispute or to settle it peacefully, yet relative power plays a *primary* role in driving the states' strategies in the war and the international outcomes that result? This is a puzzle that would not have occurred to second generation theorists, because they assumed that the wars were both initiated and conducted in response to relative power factors. My theory resolves this puzzle by integrating income factors and relative power factors within the same model (as we will see in later chapters).

## **2.2 Problems with the assumption that a "declining" state initiated war to prevent a "rising" state from reaping net gains**

There was a variant of second generation theory that seemed to take account of the motive of income loss avoidance.

- Preventive war theory assumed that when a state is declining in relative power, it will wage war against the state that is rising in relative power to secure its interests quickly, before the rising state can gain enough power to harm its interests.<sup>16</sup>

This theory assumes that if the rising state were to increase in relative power by a large enough margin, then it would gain an incentive to harm the interests of the declining state. Hence the declining state has an incentive to initiate war before the rising state's power increases by that much. According to this logic, the rising state will only have an incentive to fight back if its power had already increased by enough (before the war erupted) to win. But if its power had not yet increased by enough to win, then it will not have an incentive to fight back. By implication, if war actually occurs, it must have been because the rising state's power had increased by enough that it expected to win. This is the logic that suggests that preventive war theory should be included in the second generation. It is a theory of war caused by a shift in relative power.

Among preventive war theorists, there is a school of thought which assumes that whenever a state's income falls for any reason, this automatically creates an adverse shift in relative power at the same time, and therefore any prospect of income losses is threatening for relative power reasons (as well as income reasons). If this were true, then the seven hegemonic wars

---

<sup>16</sup>Niou and Ordeshook 1987; Levy 1987; Powell 2006; Levy 2008; Chan and Tessenman 2009; Krainin 2017.

between 1550 and 1950 would confirm the predictions of the theory of preventive war due to shifting power. (For each war was caused by an international spillover problem that caused income losses for one state.) The evidence from these wars contradicts the predictions of that theory, however. The theory predicts military victories in situations where defeats actually occurred; and it predicts military defeats in situations where victories actually occurred. Examples of both kinds are provided in section 4 of this chapter, where the theory of preventive war due to shifting power is reconsidered in its game theoretic form.

Not only is the theory falsified by the evidence, but it also suffers from a deeper logical problem. The theory conflates two very different ways that one state might harm another state's interests.

- The first state might increase in power relative to the second state.
- The first state might develop a new domestic activity that increases its own income but generates international spillover effects that reduce the second state's income.

All examples of the second scenario are lumped into the first scenario by the theory of preventive war, because it attributes war to a shift in relative power in both scenarios. (This is why the theory makes the wrong prediction in so many of the empirical cases.) In fact the second scenario can only be understood logically – and used to explain historical cases – by recognizing that it is *not* a problem of war due to shifting relative power. It is a different type of problem entirely. The problem is that when one state develops a new domestic activity that increases its income but lowers a second state's income through an international spillover effect, the second state has an incentive to initiate war against the first state to force it to stop the domestic activity. To be precise, the second state has an incentive to wage a war that will impose more war costs on the first state than it would gain from continuing to engage in its new domestic activity, thereby inducing it to stop that activity. The second state has this incentive regardless of whether the spillover problem causes relative power to shift in its favor or against it (or not at all). In response, the first state has an incentive to wage a war that will impose more war costs on the second state than it would incur by retracting its demand to stop that activity – to induce it to retract its demand and accept the income losses from the spillovers. The first state, too, has this incentive regardless of whether the spillover problem shifts relative power in its favor or against it (or not at all).



Thus we have two competing hypotheses for what caused the hegemonic wars in Europe between 1550 and 1950:

- My theory hypothesizes: Each war was caused by the motive of income loss avoidance.
- Preventive war theory hypothesizes: Each war was caused by the incentive for preventive war to stop a power shift.

The seven hegemonic wars in this period provide a "natural experiment" that can distinguish between these two hypotheses. For these wars were caused by three different types of international spillover problem, and each type had a different effect on relative power between the state generating the international spillover effect and the state suffering the income losses from it. One type of spillover problem shifted relative power in favor of the state generating the international spillover effect. A second type shifted relative power in favor of the state suffering the income losses from it. A third type had no effect on relative power between the two states. The consequences of these power shifts refute the predictions of preventive war theory, but confirm the predictions of my theory. Specifically, these power shifts explain many aspects of both states' strategies and the international outcomes that resulted – but only under the assumption that the power shifts played *no causal role* whatsoever in each state's decision whether to go to war over the international spillover problem or to settle the dispute over it peacefully. In other words, these power shifts acted as intervening variables that were driven by a deeper cause – the international spillover problem and the income losses it imposed on each state (or threatened to impose). The motive of income loss avoidance was the sole cause of each state's decision to go to war over the dispute rather than to settle it peacefully.

### **2.3 Problems with the assumption that relative power shifted before each hegemonic war**

Theorists of the second generation assumed that hegemonic wars were caused by *non-discretionary* shifts in relative power – long run shifts due to factors beyond anyone's control. In choosing this focus, they claimed, at least implicitly, that each hegemonic war became unavoidable because the preceding power shift was unstoppable. From this vantage the analytical goal was simply to show *how and why* an unstoppable power shift would motivate

a war. Theorists proposed two different logics based on two different causes of power shifts. In each branch of theory, however, there was a problem with the underlying assumption that power had actually shifted prior to each hegemonic war. It is refuted by the empirical evidence in at least four of the seven cases. Consider each branch of theory in turn.

### 2.3.1 Power shifts due to industrialization

One set of theorists focused on the impact of industrialization on relative power standings (Organski 1958; Organski and Kugler 1980; Kugler and Lemke 1996). The main hypothesis was that large states which industrialized between 1750 and 1900 gained the economic and military power to set the rules of the international system. But these large states did not industrialize simultaneously. So the first one to industrialize *actually* set the rules of the international system. And when the second one followed, it gained the economic and military power to challenge the first one for the authority to (re)set the rules. The outcome was war. From a Lakatosian perspective, this theory added a new *ceteris paribus* condition to the Conventional Realist program of proving that wars are caused by shifts in relative power.

If a hegemonic war was preceded by a power shift due to industrialization, then the war was caused by that kind of power shift. In other cases of hegemonic war where other causes were operating, however, that explanation does not apply.

In effect these theorists divided the data set of hegemonic wars in Europe into two subsets: the wars since 1750 that are explained by *exogenous* power shifts due to industrialization and the other wars that are explained by *discretionary* power shifts or other factors entirely. The question naturally arises: Is this *ceteris paribus* condition really necessary to explain these wars? I contend that it is not necessary. The theory that I offer in this study – Contractual Realism – explains all of the hegemonic wars in Europe between 1550 and 1950 as consequences of a single, deeper root cause that is unrelated to relative power. Thus it explains all of the cases that the industrialization theory explained *and more*, but with fewer *ceteris paribus* conditions.

### 2.3.2 Power shifts due to uneven economic growth

A second set of theorists tried to generalize from trends in industrial growth to trends in economic growth generally (G. C. Modelski 1978; Gilpin 1981; Kennedy 1987; Levy 1987; Goldstein 1988; Copeland 2000). The main hypothesis was the *law of uneven growth*.

”The realist theory of international political change is based on what can be called the law of uneven growth.... According to realism, the fundamental cause of war among states...is the uneven growth of power among states.... [T]he distribution of power in an international system shifts over a period of time.... [Each shift causes an] increasing disjuncture between the existing governance of the system [by the traditional dominant power] and the redistribution of power in the system. Although the hierarchy of prestige, the distribution of territory, the rules of the system, and the international division of labor continue to favor the traditional dominant power or powers, the power base on which the governance of the system ultimately rests has eroded because of differential growth and development among states.... Throughout history the primary means of resolving the disequilibrium between the [political] structure of the international system and the redistribution of power has been war, more particularly, what we shall call hegemonic war.”<sup>17</sup>

Gilpin claimed that the law of uneven growth explains the six hegemonic wars in Europe since 1618.<sup>18</sup> The historical evidence does not support this claim, however. In four of the six cases, there was no clear trend in long-run economic growth rates that might warrant coding one of the opposing states as the ”rising” state and the other one as the ”declining” state. Between 1555 and 1618, economic growth was not significantly lower or higher in Austria than it was in the other states that would eventually oppose Austria in the Thirty Years War. Between 1648 and 1688, each of the major states destined for war – France, England, Austria and the Dutch Republic – experienced solid economic growth in some sectors but stagnation in other sectors. Although one might construct a weighted average of the sectors in each state to generate an overall growth rate for each state, it is not clear what the weights should

---

<sup>17</sup>Underlining added for emphasis. Gilpin 1981, pp. 94, 187, 191, 197, 201–2.

<sup>18</sup>Gilpin 1981, p. 200.

be. While one weighting might produce overall growth rates that show a rising state and a declining state, another weighting might not. Moreover, even if one adopts the conventional wisdom on the rise and decline of the European powers, the evidence confounds the theory. In each of the hegemonic wars between 1688 and 1715, *each* of the opposing coalitions was composed of a great power on the rise and a great power on the decline. While France was on the rise, its ally Spain was on the decline. While England was on the rise, its ally the Dutch Republic was on the decline. Finally, between 1715 and 1792, each of the major states destined for another war – France and Austria – experienced solid economic growth in some sectors but stagnation in other sectors. The conventional wisdom is that France was the rising power and Austria was the declining power. Yet there is plenty of evidence in support of the opposite view. The French economy became increasingly involuted over the course of the century due to the inefficiencies of tax farming and the excesses of the manorial regime. By contrast, the reforms of the Austrian Emperors after 1750 put Austria on a path of increasing rationalization in government administration, the military and the economy. While Austria was no juggernaut by Prussian or English standards, it is simply not true to say that Austria was a power in decline compared to France between 1750 and 1792.

These empirical anomalies raise fundamental questions about the theory that war is caused by uneven economic growth over the long run. What is the observational theory underlying this substantive theory? How does it conceptualize and measure state power, the returns to power, and the returns to war? To answer these questions, consider a classic statement of the substantive theory.

"As its relative power increases, a rising state attempts to change the rules governing the international system, the division of the spheres of influence, and, most of all, the international distribution of territory. In response, the dominant [but declining] power counters this challenge through changes in its policies that attempt to restore equilibrium in the system. The historical record reveals that if it fails in its attempt, the disequilibrium will be resolved by war.... By launching a preventive war the declining power destroys or weakens the rising challenger while the military advantage is still with the declining power.

"[The declining power has] the perception that a fundamental historical change

is taking place and the gnawing fear...that time is somehow beginning to work against it and that one should settle matters through preemptive war while the advantage is still on one's side.... The alternatives open to a state whose relative power is being eclipsed are seldom those of waging war versus promoting peace, but rather waging war while the balance is still in that state's favor or waging war later when the tide may have turned against it. Thus the motive for hegemonic war, at least from the perspective of the dominant power, is to minimize one's losses rather than to maximize one's gains. In effect, a precondition for hegemonic war is the realization that the law of uneven growth has begun to operate to one's disadvantage."<sup>19</sup>

The underlying observational theory is composed of the following assumptions:

1. A state's power is directly correlated with its level of economic output (GDP). The higher its economic output is, the more power it has.
2. A state's economic output level may fluctuate in the short run due to many factors. But only a state's long-run average output really matters. By implication, only changes in its power due to changes in its long-run average output really matter. Thus a state's "power" is measured and coded by its long-run average output level.
3. This long run economic source of power is the most important source of power in international relations. It is the main determinant of relative power between states. By implication, changes in relative standings in long run economic output are the main determinant of shifts in relative power among states, and thus the main cause of wars due to shifting relative power.
4. Like an individual state's economic output, states' *relative* economic output may fluctuate in the short run due to many factors. But this doesn't matter. Only the relative standings in long run average economic output really matter. Hence only *changes* in these standings really matter.
5. Every state would *benefit* from rewriting the rules of the international system in its own favor. But only one state has the *capacity* to rewrite the rules in its favor: the state with the highest (long run average) level of economic output. This state is coded

---

<sup>19</sup>Underlining added for emphasis. Gilpin 1981, pp. 94, 187, 191, 197, 201–2.

- as "the dominant state". Because this state has the highest economic output level, it has the *lowest marginal cost* of rewriting the international rules. This fact alone is necessary and sufficient to give a state the capacity to rewrite the international rules. In other words, only the state with the lowest marginal cost of doing it can do it. And once a state has the lowest marginal cost of doing it – or at least the expectation of soon having the lowest marginal cost – this is sufficient for that state to do it.
6. States' relative standings in economic output only change slowly over decades. Hence rivalry for the position of dominant state breaks out only once every few decades.
  7. Rivalry breaks out whenever one state has the highest *level* of economic output (GDP), but not the highest *rate of growth* of its level. Another state has a lower *level* of economic output, but a higher *rate of growth* of its level. The first state is the "dominant but declining" state. The second state is the "subordinate but rising" state. Thus, in coding the raw evidence from each inter-war period into "observations", one state is coded as the dominant but declining state in that period, and another state is coded as the subordinate but rising state in that period.
  8. The *dominant* state's main returns to power come from its capacity to set the rules of the international system in its favor, especially the distribution of territory among governments.
  9. The *rising* state's main returns to power come from its capacity to displace the dominant state and *reset* the rules of the international system in its favor, especially the distribution of territory among governments.
  10. The *dominant* state's main returns to war come from its capacity to wage war to prevent the rising state from displacing it.
  11. Thus, in coding the raw evidence from each hegemonic war into "observations", one state's actions to change territorial boundaries are coded as "the returns to power from resetting the rules of the international system in the rising state's favor". And another state's actions to change territorial boundaries are coded as "the returns to power from stopping the rising state and reasserting the dominant state's authority to set the rules in its favor".

These observational assumptions form the 'empirical basis'<sup>20</sup> for the main claim of the substantive theory of war due to uneven long run growth:

As uneven growth shrinks the gap in long run average output between the dominant state and the rising state, the dominant state gains an incentive to initiate preventive war against the rising state.

The logic of this theory is impeccable. But is the theory supported by the historical evidence? States did indeed redraw territorial boundaries in each of the hegemonic wars between 1550 and 1950. But why? Did the incentive to redraw territorial boundaries arise from uneven economic growth over the *decades* prior to the war? Or did the incentive arise from some other factor? How can the *motivation* behind territorial boundary changes – the cost-benefit calculus that drove statesmen to adopt this strategy – be inferred from the historical evidence? Does the observational theory summarized above produce valid causal inferences on this question? Or is a new observational theory needed?

Two bodies of evidence speak to the question of the motivation behind territorial boundary changes. The first body of evidence suggests that the theory summarized above fails to prove motivation. In four of the seven cases of hegemonic war, the data on long run growth rates do not indicate a "rising" state and a "declining" state in the decades before the war (see above). This fact alone suggests that these four cases are anomalies for the theory of war due to uneven long run growth.

The second body of evidence suggests that a new observational theory is needed. In *every one* of the seven cases of hegemonic war between 1550 and 1950, there was an international spillover problem that caused *short run* changes in economic output and income in each of the main opposing states.<sup>21</sup> And these changes created strong incentives for states to adopt all of the strategies mentioned above: waging war, redrawing territorial boundaries, and rewriting the rules for international political and economic relations. In each case, prior

---

<sup>20</sup>According to Lakatos, the 'empirical basis' is the set of empirical "observations" that is created by applying certain conventional coding rules, previously agreed and accepted by the scientific community, to the raw evidence.

<sup>21</sup>In the standard model of a national economy, total output *equals* total income. That is, the total revenue from the sale of all output is paid back to the domestic suppliers of labor and capital as income. (This simple model abstracts away from international trade.)

to the war, one state had developed a new domestic activity that *increased* its economic income but also generated an international spillover effect that *decreased* another state's economic income. This gave the second state an incentive to wage war against the first state to stop its new domestic activity *and* the first state an incentive to fight back to defend its new domestic activity. According to existing international laws, the first state had the right to continue its new domestic activity *and* the second state had the right to be free of the spillover effects and the income losses caused by that activity. Hence the two states' rights were incompatible and the international dispute over the spillover problem became irreconcilable. Each state perceived that if it were to accept a compromise settlement that gave the other state some of the right it claimed, that would give the other state a net gain from the current dispute and hence an incentive to create more disputes in the future for its own gain. So each state perceived that it had to make sure the other state suffered a net loss from the issue in the end. That is, each state perceived that it had to impose more costs on the other state through war than the other state might gain from any compromise settlement in the end.

Each state's decision to wage war rather than settle the dispute peacefully was driven by this short run calculus – regardless of long-run trends in relative power over the preceding decades. This short run calculus also drove each state's decisions to increase military manpower, form alliances, redraw territorial boundaries and rewrite international rules. The strategy of redrawing territorial boundaries was driven by the spillover problem in two ways. First, the state suffering the income losses from the spillover problem sought to stop the first state's new domestic activity by any means necessary – including the seizure of territory. In turn the first state sought to defend its domestic activity against the second state's onslaught by any means necessary – including the seizure of territory. Second, each spillover problem changed the relative productivity of military and civilian labor within each state. This changed each state's *power* – defined as the number of men it had an incentive to transfer from its domestic economy into its military – and hence its manpower strategy for the war. In cases where the spillover problem *increased* its military labor productivity (relative to its civilian labor productivity), the state preferred to increase military manpower through *domestic mobilization*. By contrast, in cases where the spillover problem *decreased* its military labor productivity (relative to its civilian labor productivity), the state preferred to increase



military manpower through *alliance formation*. If it lacked natural allies, then it was driven to conquer new territories and establish a coercive hegemony over them to secure allies.

These two mechanisms explain *all* of the territorial strategies adopted by the main opponents in *all* of the hegemonic wars since 1550. Crucially, the impact of the spillover problem in each of the seven cases was a *short run* impact. Each state made its decision to wage war over the issue rather than settle it peacefully within weeks, months or, at most, a few years after the spillover problem first arose. Thus it was short run changes in economic income that drove state strategies and international outcomes, not long run changes. Similarly, the impact of the spillover problem on the relative productivity of military and civilian labor in each state was a *short run* impact. Thus it was short run changes in relative power that drove statesmen's decisions to adopt strategies of territorial hegemony, among other things, not long run changes.

I will specify and test this theory of hegemonic war in Parts II through V of this study. The underlying observational theory is based on the following assumptions – which contrast sharply with the assumptions of the observational theory behind the hypothesis that hegemonic wars were caused by uneven long run growth.

1. In each case there is an international spillover problem that threatens to impose income losses on the two states involved. One state has developed a new domestic activity that increases its income, but that activity emits negative spillover effects that decrease another state's income. The first state is coded as "the emitting state". The second state is coded as "the harmed state".
2. These income effects are *short run* effects which are *independent* of long run trends in economic output and income.
3. Each state's returns to military manpower come from its capacity to impose costs on the other state through war to resolve the spillover problem in its favor. By imposing costs through war, each state can (1) extract territorial concessions from the other state and then (2) use the territorial gains to extract political concessions from the other state to solve the spillover problem in its favor. The harmed state uses its military and territorial power to force the emitting state to stop its new domestic activity emitting the spillovers. In turn, the emitting state uses its military and territorial power to

force the harmed state to stop interfering in its new domestic activity.

4. Each state makes its decision whether to wage war over the spillover problem or settle the problem peacefully by comparing the expected net benefit from war to the expected net benefit from peace.

- Each state's expected benefit from peace is its current national income at the existing distribution of territory (which includes its peace dividend, e.g. the part of national income generated by allowing the military reservist corps to continue working in the civilian economy).
- Each state's expected cost of peace is that by capitulating in the dispute over the spillover problem, it would continue to suffer income losses. If the harmed state capitulates, it would continue to suffer the income losses caused by the spillovers themselves. If the emitting state capitulates, it would give up its new domestic activity and thereby lose the income generated by that activity. The "expected cost of peace" is coded from the raw historical evidence on statesmen's expected losses from capitulating in the dispute over the spillover problem.
- Each state's expected cost of war is determined by the relative productivity of its military labor and civilian labor.
- Each state's expected benefit from war comes from its capacity to resolve the spillover problem in its favor by force. The *harmed state's returns from war* arise from its capacity to stop the emitting state's new domestic activity. The *emitting state's returns from war* arise from its capacity to stop the harmed state from interfering in its new domestic activity. The "expected returns from war" is coded from the raw historical evidence on the income losses that statesmen expected to *avoid* by solving the spillover problem by force.

From a Lakatosian perspective, the question is whether the observational theory behind this new explanation of hegemonic war receives more or less support from the historical evidence than the observational theory behind the older hypothesis that hegemonic wars were caused by uneven long run growth. A central aim of Parts II through V of this study, therefore, is to derive as many observable implications as possible from this new setup – this new combination of a substantive theory and an observational theory – and to present the

historical evidence that corroborates the implications.

From this new perspective, it is worthwhile to reevaluate the observational theory associated with the hypothesis that hegemonic wars were caused by uneven long-run growth. Was that observational theory a progressive or degenerative problemshift? The answer is suggested by assumptions 5 and 6 of that observational theory (recapitulated from above).

5. .... Only one state has the *capacity* to rewrite the rules in its favor: the state with the highest (long run average) level of economic output. This state is coded as "the dominant state" ....

6. States' relative standings in economic output only change slowly over decades. Hence rivalry for the position of dominant state breaks out only once every few decades.

Assumption 6 seems to be corroborated by the historical evidence. Relative standings in economic output *did* change only in the long run – over decades. And since 1550 hegemonic wars only broke out once every 50 to 100 years. But what is it that turns this empirical evidence on the *infrequency* of hegemonic wars into a confirmation of the claim that the wars were caused by uneven long-run growth? It is assumption 5: Only the state with the highest economic output has the capacity to rewrite international rules. Together these two assumptions turn the observation that hegemonic wars were infrequent into confirming evidence for the theory. But why is it that only the state with the highest economic output has the capacity to rewrite international rules? What reason is there to believe this assumption? Is it simply that, by coding the evidence on territorial boundary changes in this way, one might confirm the theory? In other words, is this way of coding the historical evidence on boundary changes simply an example of "harmonizing theories and factual evidence by the use of conventionalist stratagems" – the tendency that Karl Popper and Imre Lakatos warned about?

"Popper agrees with the conventionalists that theories and factual propositions can always be harmonized with the help of auxiliary hypotheses: he agrees that the problem is how to demarcate between scientific and pseudoscientific *adjustments*, between rational and irrational changes of theory. According to Popper,

saving a theory with the help of auxiliary hypotheses which satisfy certain well-defined conditions represents scientific progress; but saving a theory with the help of auxiliary hypotheses which do not [satisfy such conditions] represents degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, 'conventionalist stratagems'."<sup>22</sup>

## 2.4 Net assessment of the second generation: Contributions, failures and proposed remedies

The underlying assumption of the second generation theorists remained that of the first generation: Hegemonic wars were caused by shifts in relative power. If power had not shifted, then war would not have occurred. Because power shifted, war did occur – and it would have occurred regardless of any other factors that were present or absent. The second generation theories advanced the debate over hegemonic war by asking why shifts in relative power might cause a long and costly war, and why one state might pursue territorial expansion with the goal of establishing hegemony over other states. While these theories elaborated numerous roles of state power in hegemonic war, however, they conflated the *causal* roles of power with its other roles. By contrast, my theory reconsiders the question of territorial expansion and hegemony seeking without buying into the second generation assumption that these wars were *caused* by power shifts. My theory identifies the expansion of a state's territorial power as an intervening variable that is driven by deeper causes that are unrelated to power. My theory generates a range of observable implications about territorial expansion and contraction during hegemonic wars, all of which are corroborated by the historical evidence.

The defining feature of the hegemonic wars in Europe between 1550 and 1950 was territorial conquest and consolidation on a massive scale. This is how these wars got their name. Later generations of theorists would abstract away from territorial conquest and consolidation as an object of explanation in order to focus on other puzzles, as we will see below.<sup>23</sup>

---

<sup>22</sup>Lakatos70.

<sup>23</sup>The fourth generation theories abstracted away from territorial variables as an object of explanation in order to focus on the logical underpinnings of theories of war due to relative power factors.

My theory reopens the debates about the value of territorial conquest that the second generation theorists started, and renews these debates on foundations that are more logically and empirically defensible. By explaining territorial expansion in a new way, my theory provides an opportunity to understand the historical facts that made hegemonic wars distinctive from other wars.

## **2.5 The evolution of power shift theory: From *types* of power shifts to *causes* of power shifts to the *reasons why* power shifts make war unavoidable**

The theory that wars are caused by power shifts has gone through several generations of evolution. In each generation Conventional Realists charted new analytical territory but missed a key region of that territory. Each of the missed regions is critical for understanding and explaining the history of hegemonic wars in Europe.

In the first and second generations, theorists explored different *types* of power shift and developed hypotheses about how each type might cause war (sections 1 and 2 above). They explored shifts in territorial power, shifts in military manpower, and shifts in the economic foundations of power. However, they did not consider changes in the relative productivity of a state's military labor and civilian labor – and the impact of these changes on the state's incentive to transfer men from its domestic economy into its military. This factor, too, shifts relative power between states. In every hegemonic war since 1550, this type of power shift *drove* the first two types – shifts in military manpower and territorial power – even though none of these wars was *caused* by any type of power shift at all.

In the second generation, theorists explored different *causes* of shifts in relative power and developed hypotheses about how each cause might generate war (section 2 above). They explored industrialization as a specific cause of power shifts in the long run and uneven economic growth as a more general cause of power shifts in the long run. However, they did not consider how fluctuations in economic income in the short run can alter a state's labor productivity in the military or civilian sector, thereby altering the state's incentive to transfer men from its domestic economy into its military. Such short run income changes also shift relative power among states. In every hegemonic war this short run cause of power shifts exerted a decisive impact on state strategies and international outcomes, while the

long run causes of power shifts had no demonstrable impact.

The power shift theories of the second generation were criticized by two main groups of scholars. Both groups focused on the fact that both sides suffered huge losses in each hegemonic war. Hence the puzzle should no longer be "Why do shifts in relative power cause war?" The new puzzle should be: "Why do states get into wars where both sides suffer huge net losses?" Each group of critics interpreted this puzzle differently and produced a different generation of theories as a result. In the third generation, theorists understood the puzzle as a sign that domestic factors prevented statesmen from responding to power shifts rationally and acting in the national interest to avoid a war. In the fourth generation, theorists understood the puzzle as a sign that even rational statesmen acting in the national interest might still find war unavoidable because of uncertainties about relative power or commitment problems related to relative power. Thus, in both generations theorists shifted the focus from types and causes of power shifts to the *reasons why* power shifts render war unavoidable.

In the third generation, theorists explored how war is caused by domestic distortions in the way that a state processes information about relative power, forms preferences related to relative power, and allocates the war costs arising from its response to shifts in relative power (section 3 below). The analytical goal was to show that if it were not for these domestic distortions, statesmen would act rationally and heed the national interest in avoiding wars that impose large net losses on all sides. However, theorists did not consider the possibility that in each hegemonic war statesmen actually *were* acting rationally in response to an international spillover problem, and the domestic changes in information processing, preference formation and war cost distribution were *caused* by these rational responses to the spillover problem.

In the fourth generation, theorists explored how fully rational states might go to war even if they expected net war losses all around (section 4 below). The analytical goal was to resolve the "inefficiency puzzle" of war. When two states get into a dispute, there is always a potential settlement that would save them the costs of war – and a way to split the savings that would leave each state better off than it would be going to war. So why would the two

states *ever* go to war over a dispute?<sup>24</sup> Theorists assumed that the answer would boil down to some kind of relative power factor in the end – that there is something about power shifts that prevents the peaceful settlement of disputes. Theorists sought a general solution to the inefficiency puzzle of war – one that would apply to all types of power shifts driven by all manner of causes. Two main theories were developed. One theory assumes that the root cause of the problem is *unusually large* power shifts. The other theory assumes that the root cause of the problem is *uncertainty* about whether or not power has shifted. In section 4, I will evaluate these theories against the Lakatosian standard and show they are degenerative. I will also explain what each theory missed in the territory that it explored and demonstrate that what it missed is essential for understanding and explaining the history of hegemonic wars.

### **3. The third generation (1985-95): Theories of hegemonic war due to domestic distortions**

For this generation the new puzzle was: "Why did domestic political systems fail to restrain leaders from waging a war, when everyone understood that the prevailing conditions of relative power would only produce huge net losses?" While a number of theories were developed, they shared an implicit assumption. Not only did domestic political systems fail to restrain leaders from *waging* these wars for so long at such cost, but it was these failures that *caused* the wars in the first place. Thus the third generation adopted an entirely new explanatory approach compared to the first two generations. Rather than claiming the wars were caused directly by international factors such as shifting power, the third generation held the wars were caused by domestic political factors that distorted the policy responses to shifting power. If the domestic political dysfunctions had not existed, the wars would not have occurred. The only remaining question for third generation theorists was to identify the domestic political dysfunction in each case and show how it caused the war. Again the puzzle had changed.

The theories of the third generation focused on a range of domestic political distortions

---

<sup>24</sup>Fearon 1995; Powell 2006.

and tried to show how they caused aggression and territorial over-extension. While these theories made progress in charting the domestic political patterns associated with foreign policy strategies during hegemonic wars, they assumed that domestic factors played the primary *causal* role in the outbreak of war. International factors were not seen as primary causes, but only as "triggers" of domestic political processes and distortions that were the primary cause of war.

The problem with this explanatory approach is that it ignores a simple fact. Each hegemonic war was preceded by an international dispute over an international problem. Is it possible that these international factors were the primary cause of the wars – or even the sole cause? My theory assumes the international factors were the sole cause of each war – the sole necessary and sufficient condition for war to occur – and then demonstrates that the international factors *explain* the domestic political patterns identified by the third generation theories. From this explanatory angle the domestic political factors were only intervening or epiphenomenal variables. They had no causal impact. Specifically, my theory derives each of these factors as a *consequence* of one state's need to take military action to stop the flow of negative spillover effects from another state's domestic activity, and that state's need to defend its domestic autonomy from such military action. These policy needs alone explain the motivations of all of the sub-state actors the third generation theories claimed to play independent causal roles: ruling circles, state bureaucracies, populist politicians, and societal coalition partners.

### **3.1 Domestic distortion 1: Leaders with a preference for power rather than security (e.g. revision rather than defense of the status quo)**

Conventional Realists have developed four different theories based on power-seeking motives. Each theory assumes that in each case of hegemonic war, at least one state had a preference for power rather than security, and as a result it became a "revisionist" state. Each theory includes an *ad hoc* auxiliary hypothesis that is scientifically degenerative for several reasons. It is refuted by the historical evidence. It is not needed to explain the evidence. And it is outperformed in explanatory power by the Contractual Realist theory presented in this study, which does not include that hypothesis.



### 3.1.1 *Ad hoc* auxiliary hypothesis A: There are two different kinds of states in the world

This hypothesis forms the basis of three theories in the Conventional Realist canon: structural, neoclassical and defensive realism. Each theory assumes that there are two types of states in the world. There is the "revisionist" type, which seeks to change the status quo distribution of power, wealth and authority in the international system. And there is the "status quo" type, which seeks to maintain the status quo distribution. The hypothesis is that a state's type is embedded in its fixed preference structure – the things it is assumed to want at all times – past, present and future. The revisionist type is simply assumed to want to change the status quo. The status quo type is simply assumed to want to maintain the status quo. By embedding such desires in a state's fixed preference structure, the theorist is claiming that there is no scientific need to ask where revisionist preferences come from – and no scientific value added in theorizing about where they come from. It is sufficient simply to *assume* that in each case of hegemonic war there was at least one revisionist state.

The hypothesis that there are two types of states in the world is scientifically degenerative for several reasons. First, it is refuted by the historical evidence on hegemonic wars. In each case the state that pursued revisionist goals in the war *did not have* those goals before the international spillover problem arose. This fact alone suggests that if the international spillover problem had not arisen, the state would not have pursued those revisionist goals. Therefore those goals were not a part of the state's fixed preference structure. Rather, they were a product of events in the world that caused an irreconcilable contract dispute between two states.

Second, hegemonic wars can be explained without assuming that any state has revisionist goals in its fixed preference structure. The Contractual Realist theory presented here explains hegemonic wars by assuming that each state has only one goal in its fixed preference structure: To maximize its national income. When an international dispute over income erupts between two states, and the dispute becomes irreconcilable, each of the opposing states gains an incentive to wage war for long enough that the other state does not receive a net income gain from splitting the difference in the dispute in the final peace settlement. The incentive is: To deter the opposing state from creating similar disputes for its own gain

in the future. In each case of hegemonic war, the international spillover problem caused such a dispute, and that is why each war lasted so long. Moreover, each spillover problem altered the productivity of military labor relative to civilian labor within each of the opposing states, thereby altering its incentive to transfer men from its civilian economy into its military. In states where the spillover problem *lowered* the productivity of military labor relative to civilian labor, thereby lowering the state's incentive to conscript manpower into the military, the state had an incentive to contract voluntary allies, or failing that, to conquer allies and establish hegemony over them. This logic explains the origin of revisionist preferences without resorting to the *ad hoc* auxiliary hypothesis that revisionist preferences are embedded in a state's fixed preference structure.

The clearest examples of this logic are in the four hegemonic wars caused by secession spillovers. In each case, one great power was suffering income losses from a domestic secession movement that was inspired by reformists abroad. The secession movement lowered the productivity of the great power's military labor relative to its civilian labor, thereby lowering the number of men it had an incentive to transfer from the civilian economy into the military. To combat the spillovers emitted by reformists abroad, therefore, it had to contract voluntary allies or conquer allies and establish hegemony over them. This logic explains: (1) the Spanish drive for hegemony in the Wars of Dutch Independence, (2) the Austrian drive for hegemony in the Thirty Years War, (3) the Spanish/French drive for hegemony in the War of the Spanish Succession, and (3) the Austrian/German drive for hegemony in World War I. As we will see in a moment, this logic also explains hegemonic wars caused by revolution spillovers. I use this logic in Parts II through V of this study to derive a range of observable implications for the fact patterns in hegemonic wars. In so far as all of the implications are corroborated by the evidence, this logic offers greater explanatory power than the hypothesis that there are two types of states in the world distinguished by a difference in their fixed preference structure.

### **3.1.2 *Ad hoc* auxiliary hypothesis B: Power shifted right before each war started**

This hypothesis forms the basis of the fourth theory in the Conventional Realist paradigm based on power-seeking motives: offensive realism. According to this theory, *every* state has

a preference for power embedded in its fixed preference structure. More precisely, every state always wants to achieve hegemony over its region of the world. Although it *wants* to achieve this goal, however, it does not always have the power to do so. When power conditions shift in its favor, though, its fixed preference for regional hegemony becomes an active policy and the consequence is a hegemonic war.<sup>25</sup>

The *ad hoc* auxiliary hypothesis in this theory is that relative power shifted in one state's favor just before each hegemonic war started. And this boon launched the state on a drive for hegemony. The hypothesis is scientifically degenerative for the same reasons the previous one was. It is refuted by the evidence. It is not needed to explain the evidence. And it is outperformed in explanatory power by Contractual Realism. We just saw that in each war caused by secession spillovers, the great power that pursued regional hegemony had suffered a *decrease* in its power (due to a decrease in the productivity of its military labor relative to its civilian labor, which reduced its optimal army size). In those cases it was the state *suffering* the income losses from the spillover problem that suffered a decrease in its relative power and thereby gained an incentive to pursue hegemony. In the wars caused by revolution spillovers, by contrast, it was the state *emitting* the spillovers that experienced this fate. It is worth considering these examples to see how the inner logic of Contractual Realism illuminates key similarities and differences between one hegemonic war and another.

When the Glorious Revolution broke out in England in 1688, it emitted ideological spillovers that threatened to provoke a similar revolution in France. King Louis XIV, his court and its domestic coalition partners would suffer large income losses in that event. This problem provoked an international contract dispute between France and England over the right of a noble parliament to impose legal restrictions on their monarchy's succession. At the moment the dispute became irreconcilable, France gained an incentive to wage offensive war against England to overturn its domestic revolution – simply to defend the incomes of the French court and its clients in society from the spillovers. This increased the productivity of France's military labor relative to its civilian labor. By comparison, England did not experience a similar increase in the productivity of its military labor, because it was not threatened by income losses in the same way that France was. So in comparative terms,

---

<sup>25</sup>Mearsheimer 2001.

England was placed at a relative power disadvantage at that moment in the dispute. To fend off the threat of a French invasion, therefore, England had to contract voluntary allies or conquer allies and establish hegemony over them. Since England was a *conservative* revolutionary state (and France had a history of minor territorial encroachments on its neighbors), England was able to find many voluntary allies to compensate for its power disadvantage.

The case of revolutionary France a century later tells the same story with the exception of that last point. When the French Revolution broke out in 1789, it emitted ideological spillovers that threatened to provoke a similar revolution in Austria. The Austrian monarchy and nobility would suffer large income losses in that event. This problem provoked an international contract dispute between Austria and France over the right of a popular parliament to abolish their country's monarchy and replace it with a republic. At the moment the dispute became irreconcilable, Austria gained an incentive to wage offensive war against France to overturn its domestic revolution – simply to defend Austrian noble incomes from the spillovers. This increased the productivity of Austria's military labor relative to its civilian labor. By comparison, France did not experience a similar increase in the productivity of its military labor, because it was not threatened by income losses in the same way that Austria was. So in comparative terms, France was placed at a relative power disadvantage at that moment in the dispute. To fend off the threat of an Austrian invasion, therefore, France had to contract voluntary allies or conquer allies and establish hegemony over them. Since France was a *radical* revolutionary state, it was unable to find voluntary allies, so it had no choice but to conquer allies and establish hegemony over them to compensate for its power disadvantage. This logic explains the large increase in France's army size in 1793. Thus, France's abrupt turn toward a strategy of regional hegemony was not driven by any *increase* in its underlying power potential – as offensive realism would suggest. On the contrary, France's underlying power base had been deteriorating for most of the 18th century, as the inefficiencies of tax farming and the excesses of the manorial regime drove many sectors of the economy into stagnation or even absolute decline. Rather, the French strategic pivot was driven by the international contract dispute over the spillover problem and the emergence of a French power *disadvantage* relative to Austria at the moment the dispute became irreconcilable.

## 3.2 Domestic distortion 2: The entrenchment of outdated beliefs about relative power

### 3.2.1 Offense-defense theory

In response to the anomalies in their other theories, Conventionalist Realists developed a theory that explains war initiation is a consequence of developments in military technology. According to this theory, some military inventions give an advantage to offensive military action while others give an advantage to defensive military action. The offensive technologies cause war by giving statesmen a belief that victory will be quick and easy. If this belief is accurate, then states will initiate attacks more often than they did prior to the technology's invention. If this belief is false, then states will make the *mistake* of initiating attacks more often than they did prior to the invention. Conversely, the defensive technologies cause war by giving statesmen a belief that they are unlikely to be attacked, because no state contemplating an attack would ever expect to win against a target having a defensive advantage. If this belief is accurate and all states subscribe to it, then wars will be less frequent than they were prior to the invention of the defensive technology. But if this belief is untrue and *some* states know that, then wars will be more frequent. For the states that know it is untrue will have an incentive to attack the other states that are still feeling secure in their (false) belief in a defensive advantage. And those states will make the *mistake* of failing to see the attack coming and failing to arm sufficiently to deter it.

Conventional Realists claim that this theory explains the initiation of World Wars I and II as well as the differences between the two cases. According to this explanation, World War I erupted because states believed they had offensive advantages that would produce quick and easy victories, so they were more willing to initiate war than they should have been, given the truth that the advantage actually lay with the defensive. This story seems to explain: (a) why the war started so quickly and (b) why it turned out to be so indecisive and long. Conversely, World War II erupted because France and England believed they had a defensive advantage, but Germany knew it wasn't true. So Germany had an incentive to attack them. And they made the mistake of failing to see the attack coming and failing to arm sufficiently to deter it. This story seems to explain: (a) why Germany started planning

an offensive war as early as 1933, (b) why the Western Powers still believed Germany had peaceful intentions as late as 1938, and (c) why the Western Powers failed to arm sufficiently to deter a German attack.

The evidence seems consistent with the theory. Yet offense-defense theory is scientifically degenerative for the same reasons that the previous theories were. Its auxiliary hypotheses are *ad hoc* in nature and not necessary to explain the evidence they seem needed to explain. It is refuted by the preponderance of the evidence. And it is outperformed in explanatory power by Contractual Realism.

First consider the evidence that refutes offense-defense theory. World War I was not the only hegemonic war caused by a secession problem. Three previous hegemonic wars between 1515 and 1750 were also caused by secession problems. Each of these wars started quickly as well. And each war turned out to be long and indecisive despite the need and hope of the initiating states to win quickly and decisively. Yet all three wars occurred long before the military technology inventions of the 19th and early 20th centuries that supposedly caused World War I by creating an offensive advantage – or false beliefs in an offensive advantage. This evidence refutes offense-defense theory.

World War II also provides evidence that refutes offense-defense theory. Britain did have a defensive advantage (as the theory would suggest), but the advantage was due to Britain's position as an island nation rather than any military technology. This defensive advantage deterred Germany from launching an invasion of Britain (as the theory would suggest), but it did not deter Germany from declaring war on Britain and waging war against Britain in other ways. Finally, some British analysts believed that Germany had pacific intentions well into the late 1930s (as the theory would suggest), but other British analysts discerned Germany's aggressive intentions in the early 1930s. This difference is not explained by offense-defense theory, but it is explained by Contractual Realism – along with a range of other evidence on hegemonic wars.

According to Contractual Realism, each of the seven hegemonic wars since 1550 was caused by an international spillover problem. Before the war erupted, one state developed a new domestic activity that increased its income but generated international spillover effects that decreased another state's income. (I term the first state the *emitting* state and the

second state the *harmed* state.) The spillover problem also affected each state's labor productivity in the military sector relative to the civilian sector, thereby altering its incentive to transfer men from the civilian economy into the military. The effect was different from one type of spillover problem to another, however. In the wars caused by *secession* spillovers (like World War I), the spillovers *increased military* labor productivity in the emitting state, but *decreased military* labor productivity in the harmed state. By contrast, in the wars caused by *depression* spillovers (like World War II), the spillovers *increased civilian* labor productivity in the emitting state, but *decreased civilian* labor productivity in the harmed state. These patterns are displayed in Table X along with each state's optimal strategy given the impact of the spillovers on its income and labor productivity. The easiest way to understand the table is to read the cells for a secession war first, starting with the harmed state (lower left cell) and moving to the emitting state (upper right cell). Then read the cells for a depression war, against starting with the harmed state (lower right cell) and moving to the emitting state (upper left cell).

This theory explains all of the patterns of evidence from World Wars I and II noted above *without* the need for auxiliary hypotheses about military technology and the advantages it is presumed to confer on the offensive or defensive. It explains why World War I started quickly while World War II took a long time to start (counting from the date the spillover problem first cropped up). It explains why the Entente balanced quickly in the run up to World War I, while Britain and France balanced slowly and insufficiently in the run up to World War II. And it explains why each war failed to produce a quick victory for either side: Each side had to continue waging war for long enough that the other side would not receive a net gain from splitting the difference in the dispute over the spillover problem in the final peace settlement (e.g. giving each of the opposing sides half of the right it had been fighting for).

This theory also explains key similarities and differences between the four hegemonic wars caused by secession spillovers, thereby proving the mechanism of causation behind World War I. In each of the four cases, the state *harmed* by the secession spillovers started the war quickly, sought a great power ally quickly to avoid losing the war in a one-on-one fight, offered the ally generous terms, and went on the territorial offensive in the hope of winning

the war in a quick rout. In each case, once the harmed state found an ally, the state *emitting* the spillovers sought allies quickly. But her potential allies only joined her coalition after their own interests were touched by the harmed state. The length of time that this took was different from case to case (and from ally to ally within each case, at least in two of the four cases). Consider the four cases in comparative perspective.

In the Wars of Dutch Independence (1568-1609), Spain's offensive into the Dutch Republic touched England's interests immediately and England supported the Dutch immediately. In the Thirty Years War (1618-48), Austria's offensives into the Kingdom of Bohemia between 1618 and 1620 did not touch Danish, Swedish or French interests, so they declined requests for an alliance. But when the Austrian coalition expanded its offensive into southern Germany in the early 1620s, it touched Danish interests and Denmark supported the Protestant cause immediately. When the Austrian coalition expanded its offensive into northern Germany in the late 1620s, it touched Swedish interests and Sweden supported the Protestant cause immediately. When the Austrian coalition expanded its offensive into western Germany in the early to mid 1630s, it touched French interests and France supported the Protestant cause immediately. In the War of the Spanish Succession (1701-14), the Spanish/French offensive into northern Italy did not touch Dutch or British interests, so they declined Austria's request for an alliance. But when France expanded the offensive into the southern Netherlands, it touched the interests of the Dutch Republic and the Dutch joined the Austrian coalition immediately. A few months later France acquired exclusive trading rights in the Spanish Empire at England's expense. This touched England's commercial interests and she joined the Austrian coalition immediately. Finally, in World War I (1914-18), the Austro-German threat to Serbia in the summer of 1914 touched Russian interests immediately, so she supported Serbia immediately. In turn, the Austro-German threat to Russia touched the interests of her ally, France, so France joined the Russian/Serbian coalition immediately. Finally, the Austro-German threat to France touched British interests, so Britain joined the Allied coalition immediately.

This comparative evidence shows that the cause of "chain-ganging" dynamics in the initiation of a single hegemonic war was *not* any offensive advantage created by military technology – or any offensive illusion created by the false belief in such an advantage. Rather,



chain-ganging occurred in the one case where potential allies' interests were touched quickly. In the three other cases where their interests were not touched quickly, chain-ganging did not occur. But even in these cases, the potential allies eventually joined the coalition against the state going on the offensive to stop the secession spillovers. This similarity between all four cases of hegemonic war caused by secession spillovers sheds new light on the true causes of World War I's spiral into the abyss. The spiral happened faster in this case than it did in the other three cases. But it was the same spiral nonetheless.

A critic might object that in the case of World War I, the *likelihood* of war was increased by the offensive advantage conferred by military technology (or the false belief in it). According to this objection, peace could have been maintained in the face of the secession spillover problem *if* the opposing sides had had more time to negotiate a diplomatic settlement of the problem. But the offensive advantage (or illusion) forced them to rush into war before they could reach such a settlement. This objection is refuted by three pieces of evidence. In two of the earlier cases between 1550 and 1715, there *was* more time to negotiate a settlement of the secession spillover problem. Yet the opposing sides still failed to reach a settlement and the result was war. These facts show that the cause of war in all of the secession cases was the spillover problem alone, not the military technology inventions of the 19th and early 20th centuries. In one of the earlier cases (the Thirty Years War), the elapsed time from the eruption of the spillover problem to the outbreak of war was exactly two months – the same length of time as in the case of World War I. This fact shows that the rush to war in both cases was due to the spillover problem, not the military inventions in the decades before World War I.

### **3.2.2 The vulnerability of empire**

## **3.3 Domestic distortion 3: The uneven distribution of war costs arising from shifts in relative power**

In response to the anomalies in Conventional Realist explanations of the two World Wars, one critic proposed a theory in which these wars were caused by the impact of Germany's

rise in relative power on its domestic system for distributing war costs.<sup>26</sup> According to this theory, Germany's rise to great power status through "late industrialization" created a set of domestic interest groups and a system for aggregating their interests that were distorted in several ways. First, some domestic groups stood to gain from shifting war costs onto other domestic groups. Second, the system was vulnerable to myopic vote trading agreements among the groups that sought to shift war costs onto others. Third, the vote trading groups promoted ideologies and "myths of empire" that claimed territorial expansion would yield net gains for the nation as a whole – and everyone had incentives to continue believing the myths even after they were disproven by catastrophic military losses. The core claim of the theory is that one or two of these distortions alone would not have caused the wars, but together the three distortions did. To assess this theory, it is necessary to consider its logic more closely.

According to the theory, some domestic groups in Germany stood to gain from shifting war costs onto other domestic groups. Each group of potential gainers had a preferred policy that would bring it net gains but impose net costs on the nation as a whole. The agricultural landowners stood to gain from trade tariffs even though it would provoke Russia. The heavy industrialists stood to gain from investment in a strong navy even though it would provoke a naval race with Britain. The military establishment stood to gain from doctrines of offensive military action and territorial expansion even though it would provoke other great powers into encircling Germany.<sup>27</sup> Each *one* of these groups was too weak in the domestic political system to enact its preferred policy, in part because of opposition from the other groups and the rest of the nation that would lose from the policy. So each group agreed to support the other groups' preferred policies in return for their support for its preferred policy. This vote trading arrangement, or "logrolling coalition", led to the adoption of a policy package of militarism, expansionist ideology, war initiation and domestic cost shifting.

"The power and interests of separate groups are necessary but not sufficient to explain why Wilhelmine and Nazi foreign policy got so out of hand. German expansion in those periods was more extreme than any elite interest group indi-

---

<sup>26</sup>Snyder 1991.

<sup>27</sup>Snyder 1991, pp. 98–9.

vidually desired.”<sup>108</sup>

This theory seems to explain Germany’s adherence to expansionist ideologies and strategies in peacetime as well as its decision to initiate the first world war. The developer of the theory admits, however, that it does not explain why the war was allowed to continue once it started to generate large net losses even for the logrolling groups. To explain continued support for the war after this point, the theory assumes that domestic actors across the political spectrum continued to hold on to the same self-defeating ideologies and ”myths of empire” that they believed in peacetime. This is where the domestic distortions were mutually reinforcing. The domestic logrolling distortion enabled the *expected* war costs of *future* expansionist wars to be distributed unevenly during peacetime. Then the domestic mythmaking distortion prevented those who lost from the logroll from seeing the problem in wartime. In short, the myths of empire became so firmly entrenched in everyone’s minds that they simply could not respond rationally to large war losses and bring an end to the war.

First consider how this theory explains the entrenchment process that allowed World War I to continue for so long despite large net losses. Then consider how it explains the entrenchment process that allowed World War II to continue for so long despite large net losses.

### **3.3.1 The domestic mythmaking and entrenchment theory of World War I**

According to this theory, in the decades before the war, two elite groups used ideologies and myths of empire to influence mass politics and gain electoral support for their preferred policies. As a result the idea that territorial expansion would produce net gains took hold among large segments of the population. Each elite group then had to maintain its variant of the idea to protect its interests in the domestic political process from electoral losses should its supporters became disillusioned and vote for the other group or the socialists. The agricultural land holders protected their interest in trade tariffs and economic empire building through militarist ”agrarian leagues”. The heavy industrialists protected their

---

<sup>108</sup>Snyder91

interest in naval investment through militarist "patriotic leagues" and a "Navy League". At this point elite and popular politics combined to *entrench* the myth that territorial expansion would produce net gains *so deeply* that it could not be abandoned even in the face of huge net losses in the war. In the words of the theorist:

"Recruiting mass groups complicated the elite logrolling process. [The involvement of a greater number of groups] made it more cumbersome for the conservative elites to terminate the vote trading arrangement if it became counterproductive..... The additional problem was that many participants in the logrolling had fully internalized the myths of empire and consequently were unable to recognize that the big stick policies were failing.... Though the elites helped shape mass preferences, they found they could not limit the mass passions they had unleashed, especially when newly emerging elite groups found they could use nationalist arguments to flog the more cautious old elites.... Increasingly, the sorcerer's apprentice felt overwhelmed by the monster he had helped create.... Thus the old elites came to depend on uncontrollable, true believing radical nationalists for protection against socialist electoral victories. This explains, then, why the elite coalition of iron and rye found it impossible to retreat from an imperialist, logrolled platform even when its continuation seemed to jeopardize the interests of the logrollers themselves."<sup>28</sup>

In short, logrolling succeeded *before* World War I because the coalition of groups that expected to gain from a war employed the "myths of empire" to convince other domestic groups that stood to lose that a war was in their interest nonetheless. Then those groups continued to believe it even *after* the first world war started generating huge net losses, simply because their "mass passions" had become "unleashed". This is how a war that brought huge net losses for the nation as a whole *and* its main elite *and* mass groups dragged on for years without any broad domestic campaign to stop it. Each of the logrolling groups that created the myths of empire *before* the war continued to receive its preferred policy *during* the war. And the rest of the nation onto whom the costs of those policies were shifted became so

---

<sup>28</sup>Snyder 1991, pp. 102-5.

enthralled by the myths that they could not be enrolled in an effort to stop the war after it turned into a net loss for the logrollers themselves.

The following table lists the key facts about Germany in World War I that are explained by this theory of domestic mythmaking and entrenchment.

#	Key Facts about Germany in World War I
	Before and during the war, each domestic interest group demanded the <i>current</i> policy benefit that it valued the most from the central government.
	Before and during the war, the central government conceded to each of the main domestic interest groups the <i>current</i> policy benefit that it valued the most in return for its <i>current</i> support for the war plan and then the war.
	As a consequence, control over specific policy areas became decentralized and distributed across the domestic interest groups.
	Each domestic group continued to support the war despite suffering large net losses from it.
	The German nation as a whole suffered large net losses from the war in the end.

### 3.3.2 The domestic mythmaking and entrenchment theory of World War II

According to this theory, the initiation and continuation of World War II are explained entirely by the domestic mythmaking and entrenchment process before and during World War I. Hitler and the Nazis imbibed the myths in the 1920s and once they gained absolute power in 1933, their beliefs alone drove the policy of territorial expansion. They neither needed nor received broader support for expansionism from other elites or the broader population.

”Hitler and the Nazis were true believing victims of the stew of strategic myths that the Wilhelmine and wartime radical right concocted and over time spiced up [between 1870 and 1918].... Hitler’s expansionist ideas helped him win the loyalty of the hard core of his nationalist supporters.... However, there is little support for the argument that Hitler’s expansionist ideas directly helped him win

broad electoral support....”<sup>29</sup>

On this view Hitler and the Nazis *alone* turned the myths of empire into an expansionist war. The mainstream voters who cast ballots for the National Socialist party were attracted by its *other* promises, *not* by a promise of territorial expansion.<sup>30</sup> And there was no logrolling coalition among elite groups that might have expected to gain from a war of conquest.

”Junkers, industrialists, and the military...did not in any way set the tone for Nazi foreign policy. Indeed, Hitler had to complete the subjugation of these groups before he could go on the warpath in the late 1930s.”<sup>31</sup>

The theoretical assumption seems to be that if mainstream voters and elite groups had known the National Socialist party would spearhead a policy of territorial expansion, they would have stopped it. And once they found out, it was too late to stop it.

In short, once the National Socialist regime was in power, it implemented the narrow policy preferences of its leader without constraint from any other elite groups or segments of society.

”The Nazi era was a period of rule by a single dictator, during which expansion and strategic mythology ran amuck.... [T]he political system present[ed] few constraints, for good or ill, on the whims of the dictator.”<sup>32</sup>

On this view, the ruler received his mythology about the expected value of territorial expansion from the period before and during World War I. Hence his expansionist policy is explained, first, by the myth’s entrenchment in that period, second, by its entrenchment in his mind in the 1920s, and third, by his skill at entrenching it in the minds of his core followers in the 1930s.

From this perspective, popular loyalty to Hitler’s rule expanded during the peaceful years of recovery from 1933 to 1939 for the same reason that loyalty to his party expanded in the

---

<sup>29</sup>Snyder 1991, p. 107.

<sup>30</sup>Snyder 1991, pp. 106–77.

<sup>31</sup>Snyder 1991, p. 105.

<sup>32</sup>Snyder 1991, p. 109.

depression years from 1929 to 1932. It was *not* because mainstream groups in German society believed that an expansionist war would pay dividends in the future, but because they believed Hitler's *other* promises that his other policies would pay dividends in the future (e.g. sustained economic growth, higher farm incomes, upward mobility for the middle classes, etc). As popular loyalty expanded, the regime simply *added* the myth of territorial expansion to the list of promises for which people were willing to grant the regime loyalty, compliance and sacrifices *now* in return for expected dividends *later*. This is how the ruler's program of territorial expansion became the people's program of territorial expansion.

This theory of mythmaking, entrenchment and expansion explains some key facts about Germany in World War II. These facts are listed in the following table along with the list of facts from World War I. Note that on facts 1-3, each case is the *opposite* of the other case, while on facts 4-5, each case is the *same* as the other case.

#	Key Facts about Germany in World War I	Key Facts about Germany in World War II
	Before and during the war, each domestic interest group demanded the <i>current</i> policy benefit that it valued the most from the central government.	Before and during the war, the central government demanded <i>current</i> loyalty and compliance from each domestic interest group in the area where it could add the most value to the national war effort as a whole.
	Before and during the war, the central government conceded to each of the main domestic interest groups the <i>current</i> policy benefit that it valued the most in return for its <i>current</i> support for the war plan and the war.	Before and during the war, each domestic interest group conceded <i>current</i> loyalty and compliance to the central government in return for its promise of <i>future</i> policy benefits once the war was over and the empire established.
	As a consequence, control over specific policy areas became decentralized and distributed across domestic interest groups.	As a consequence, control over specific policy areas became centralized in the state and the National Socialist party.
	Each domestic group continued to support the war despite suffering large net losses from it.	Same.
	The German nation as a whole suffered large net losses from the war in the end.	Same.

### 3.3.3 The mythmaking theory's conclusion: Germany's attachment to costly war strategies was caused by domestic distortions

According to this theory, Berlin's attachment to a costly war strategy in each case was caused entirely by its domestic distortions, *not* by any need to combat foreign threats. The extent of the distortions and the attachment can be measured by Germany's overexpansion in each case. In turn the degree of overexpansion is indicated by "the strength of the opposing



coalition” of great powers that formed to stop and reverse it.<sup>33</sup> By implication, the net war losses that Germany racked up in each case by *continuing* to wage war against such a large opposing coalition indicate the extent of all three dysfunctions: the overexpansion, the attachment to a losing war, and the domestic distortions that drove it all. The theory’s conclusion is that if the domestic distortions had not existed, then the wars would not have occurred for any other reason – or if they had occurred for some other reason, Germany would not have waged the wars for anywhere near as long as it did. In this sense, the domestic distortions were the sole cause of the wars that actually occurred.

### 3.3.4 Contractual Realism’s conclusion: Germany’s attachment to costly war strategies was caused by the need to combat foreign threats

Prior to each war, Germany was threatened by an international spillover problem that was going to reduce its income significantly if not solved in one way or another. Once the international dispute over the spillover problem became irreconcilable due to a legal incompatibility problem, Germany’s only option was to resolve the spillover problem by military force. The military solution was to stop the country that was emitting the spillovers from continuing the domestic activity that was emitting them. If Germany were to wage war against the country until its war costs were greater than the income gains it was reaping from the domestic activity, that would deter it from continuing to emit the spillovers. Germany itself would have to incur net war losses *during the war* to meet this deterrence condition. But this would be worth it to avoid the even greater income losses Germany would suffer from allowing the other country to continue emitting the spillovers forever. Thus Germany was fighting a war of *loss avoidance*: a choice between the lesser of two net losses. From the vantage of the German nation as a whole, net war losses were *rational* in each case because they were less than the income losses that would be *avoided* by solving the international spillover problem. That is, Germany’s net losses *during the war* were simply the price of establishing a peace with the other country *after the war* in which it no longer had the right to emit the spillovers.

From this perspective, the fact that Germany suffered net war losses during the war does

---

<sup>33</sup>Snyder 1991, pp. 67–8.

not necessarily indicate that its domestic institutions were distorted or its domestic actors were behaving irrationally. On the contrary, in a war of loss avoidance it is rational to incur net war losses. When facing such a situation, a war that yields net war losses is the *loss-minimizing* foreign policy. And just as Germany had a strong incentive to choose a loss-minimizing *foreign* policy, it had an equally strong incentive to choose a cost-minimizing domestic strategy to support that policy. By *minimizing the domestic resource cost* of producing military power for long enough to meet the deterrence condition mentioned above, Germany could minimize the net war losses it would need to incur to solve the spillover problem and establish a peace in which its income were not threatened by the other country. From this angle the developments in Germany's domestic politics and institutions were *rational* responses to the challenge of minimizing the domestic resource cost of producing, projecting and sustaining military power for long enough to deter the foreign threat posed by the spillover-emitting country.

This analysis explains a wide range of facts about German domestic politics and institutions in the World Wars. Moreover, it explains significant differences between the two cases on the domestic level. The key to the analysis is to recognize that the international spillovers affected Germany's domestic labor productivity *differently* in the two cases. In World War I, the spillovers lowered the productivity of Germany's *military* labor, thereby *reducing* its incentive to transfer men from its civilian economy into its military. In World War II, by contrast, the spillovers lowered the productivity of Germany's *civilian* labor, thereby *increasing* its incentive to transfer men from the civilian economy into the military. This difference explains why the domestic institutions and political alignments needed to sustain a war of loss avoidance were so different in the two cases.

**the effects of secession and depression spillovers on domestic politics** Figure 30 displays a state's marginal labor productivity in the military and civilian sectors. The *x-axis* represents the state's total population  $N$ , which is divided between a military labor pool  $A$  and a civilian labor pool  $L$ . The size of the military labor pool is counted from the left end of the *x-axis* up to the hash mark at  $A_0^*$ . The size of the civilian labor pool is counted from the right end of the *x-axis* up to the hash mark at  $A_0^*$ . Thus  $A + L = N$ . The *y-axis* measures the marginal productivity of labor in the two sectors. The marginal productivity of labor in

the military sector is represented by curve  $mpA$ . The marginal productivity of labor in the civilian sector is represented by curve  $mpL$ . The optimal allocation of the state's population between the two sectors is the point where the two marginal productivity curves cross. This point is represented on the  $x$ -axis by the hash mark at  $A_0^*$ . This is the allocation of labor that maximizes national income. If the state allocated its labor at any point to the *left* of  $A_0^*$ , then it could increase national income by moving rightward to  $A_0^*$ , because each man transferred from the civilian sector to the military sector would *add* a larger increment to national income through his military productivity than he would *subtract* from his foregone civilian productivity. Conversely, if the state allocated its labor at any point to the *right* of  $A_0^*$ , then it could increase national income by moving leftward to  $A_0^*$ , because each man transferred from the military sector to the civilian sector would add a larger increment to national income through his civilian productivity than he would subtract from his foregone military productivity.

**secession spillovers** In a war caused by secession spillovers, like World War I, the state harmed by the spillovers experiences a *decrease* in the productivity of its *military* labor. This decrease is shown in the upper panel of Figure 30. The marginal productivity curve for military labor shifts down from  $mpA_0$  to  $mpA_S$ . As a result the state's optimal army size *decreases* from  $A_0^*$  to  $A_S^*$ . The segment of the labor force between  $A_0^*$  and  $A_S^*$  is now more productive in the civilian economy, so it can earn a higher return on its labor effort in the civilian economy than in the military sector (upward arrows). Hence these domestic groups to be transferred from the military sector into the civilian economic sector. But the central government prefers to keep these groups in the military sector to wage war against the secessionists and their foreign supporters and reclaim the rebellious provinces. By doing so, however, the government is imposing a wage cut on these groups (from  $mpS2$  to  $mpS1$ ) as well as an income loss on the other groups in the civilian economy that pay taxes to fund the military sector. To maintain domestic support for a war at this army size, therefore, the government must *compensate* both sets of groups by guaranteeing them shares of the *future* benefits from the war to reclaim the rebellious provinces. The government can guarantee their shares by granting them *current* authority rights over policy making and institutional reforms in central government agencies and their own regions of the country.

They will see the government's willingness to compensate them as an opportunity to demand higher compensation in the form of a greater share of the future benefits from the war. Their demands will induce a bidding war among central government officials to maintain domestic support for the war by granting domestic groups more current authority rights. Domestic politics and institutions will be transformed by such transactions. Government authority will become more decentralized. The distribution of rights and obligations between government and society will shift in society's favor. There will be tensions and disputes between domestic groups over claims that certain groups are receiving more authority rights than they should be receiving. Some groups will scapegoat others to induce the government to grant them less authority rights (and the scapegoaters more).

This model explains the domestic transformations in the four hegemonic wars between 1550 and 1950 that were caused by secession spillovers. The close similarity between the four cases demonstrates that domestic developments in Germany in World War I were driven by structural forces, not by the idiosyncracies of Wilhelmine politics, ideology or logrolling. In Part IV of this study, I provide case studies of domestic politics in Austria during the Thirty Years War and World War I to demonstrate this point.

**depression spillovers** In a war caused by depression spillovers, like World War II, the state harmed by the spillovers experiences a *decrease* in the productivity of its *civilian* labor. This decrease is shown in the upper panel of Table 30. The marginal productivity curve for civilian labor shifts down from  $mpL_0$  to  $mpL_S$ . As a result the state's optimal army size *increases* from  $A_0^*$  to  $A_D^*$ . The segment of the labor force between  $A_0^*$  and  $A_D^*$  is now more productive in the military sector than in the civilian economy, so it can earn a higher return on its labor effort in the military sector (upward arrows). Hence these domestic groups – and the others in the civilian economy that also pay taxes to fund the military sector – prefer that they be transferred from the civilian economic sector into the military sector. If the government wants to maintain the army size at its pre-crisis optimum,  $A_0^*$ , then the domestic groups must compensate the government for increasing the army size to  $A_D^*$  by guaranteeing it a share of the *future* benefits from increasing it and then waging a war to solve the spillover problem by force. The domestic groups can guarantee this share by increasing the government's *current* authority rights over policy making and institutional

reforms in all central government agencies and around the country.

The incumbent government – and any challenger that seeks to replace it – will see the domestic groups' willingness to compensate it in this way as an opportunity to demand higher compensation in the form of a greater share of the future benefits from the war. Its demands will induce a bidding war among domestic groups to maintain the government's support for the war by granting the government more current authority rights. These transactions will transform domestic politics and institutions in several ways. Government authority will become more centralized. The distribution of rights and obligations between government and society will shift in the government's favor. There will be tensions and disputes between domestic groups over their claims that certain groups are giving less support, loyalty and obedience to the government than they should be. The government will scapegoat some groups to induce other groups to give more support, loyalty and obedience.

This model explains the domestic transformations in both Germany and Japan in World War II. The close similarity between the two cases demonstrates that German developments before and during World War II were driven by structural forces – the same ones that drove Japanese developments – not by the idiosyncracies of Hitler, Nazism or the myths and logrolls of the Wilhelmine era. In Part IV of this study, I provide a case study of Japanese domestic politics in World War II to demonstrate this point.

This model also explains the major differences between Germany's domestic politics in World War I and its domestic politics in World War II. The model's success in explaining the other cases of hegemonic war (mentioned above) demonstrate that the differences between the German cases were driven by structural forces, not by the idiosyncratic factors claimed by the theory of domestic mythmaking and entrenchment.

### **3.3.5 The domestic mythmaking theory *versus* the Contractual Realist theory: Testing their causal assumptions on the evidence**

The domestic mythmaking theory claims that Germany was not threatened significantly by international factors before either world war. Germany brought the military opposition of the great powers on itself by its own mythmaking and expansionism. This reading of the evidence leads naturally to the conclusion that these two hegemonic wars were caused by

domestic distortions within Germany.

Contractual Realist theory suggests the opposite reading of the evidence. Germany *was* threatened significantly by international factors before each world war. In the next two subsections I summarize the evidence on the foreign threat that Germany faced in each case. In the third subsection I consider the logic and evidence offered by the domestic mythmaking theory to try to refute the hypothesis that German strategy was responding to foreign threats.

**the foreign threat to Germany before World War I** In the summer of 1914, the Austro-Hungarian empire was threatened by a domestic secession movement among its South Slav populations. The movement received ideological and logistical support from fellow Slavs across the border in the Kingdom of Serbia. Other ethnic groups in the Austro-Hungarian empire took inspiration from the movement to seek greater autonomy from the imperial center in Vienna. These movements lowered Austria's military labor productivity both in domestic operations and in foreign operations against supporters of the domestic secession movement. In this way the spillovers coming from Serbia threatened Vienna with the disintegration of its multi-ethnic empire.

Germany was dependent on Austria as a military ally in the event of a war with Russia or France (who were allies). If Austria's empire were to disintegrate, it would leave Germany vulnerable to territorial losses in a great power war. Over the previous 175 years Germany had acquired territory by force from its neighbors in all four directions – including territories with mixed populations of Germans and Slavs. So German statesmen knew well that a war could produce territorial losses if the enemy had a power advantage. The prospect of losing territory was not simply a problem of losing military power relative to other states. For the goal of maintaining relative power was simply a means to an end – securing the incomes of Germany's workers, elites and rulers. Thus the prospect of losing territory was ultimately a problem of losing income.

There were two reasons why the threat of income losses was the most fundamental threat to Germany from the spillover problem. First, the German government was managed by statesmen and administrators whose salaries depended on tax revenue. A loss of territory

would mean a loss of tax revenue. Second, Germany had territories with mixed populations in which the Germans preferred to remain under its sovereignty, because they could earn higher incomes as first-class citizens in Germany than as second-class citizens in a neighboring realm. Statesmen in Berlin knew that if they lost such territories in a war, it could provoke fears of the same fate in other provinces with mixed populations of Germans and other nationalities. And this could lead to an increase in demands on the regime from any number of domestic provinces – which would cost tax revenue to placate or quell.

From this perspective, the main threat to Germany from a drop in Austrian military productivity was not simply that it would leave Germany at a power disadvantage in a war with Russia or France, but more fundamentally, that it would impose income losses on German populations, elites and rulers. This was the same threat that Austria faced in the event that its empire disintegrated. Thus Germany adopted Austria's secession crisis as its own because the stakes were ultimately the same for both states.

**the foreign threat to Germany before World War II** Between 1929 and 1932, German domestic incomes fell dramatically. One of the main causes was the tariff and credit restrictions enacted by France, England and the United States. In 1932 Hitler gave a speech to key business elites in which he asserted that it was too risky to reopen trade and finance with the Western Powers, because this would leave Germany vulnerable to another cycle of market closure and depression. The only other option, he told them, was to capture markets by force in the east. While the conservative elites in business, the state and the military did not share Hitler's more extreme views on nation and race, they did share an interest in sustainable economic recovery. While they tried to avoid giving him the top leadership post, they eventually did so in the belief that the shared interest in recovery would give them the policies they wanted – and the belief that they could control his extremism on other issues. In this sense the National Socialist regime was a least common denominator that served as a vehicle not only for Hitler's extremism but also for the *conservative* mainstream's interest in economic recovery.

As far as territorial expansion to capture markets, the conservative mainstream certainly did not draw the line on expansion as far and wide as Hitler did. But a substantial group of

conservatives drew the line *beyond* Germany's 1933 borders. Their preference was revealed later by their formation of a coalition to lobby the regime for a halt to the war and the opening of peace negotiations with the Allied Powers. Their coalition did *not* lobby for a return to Germany's 1933 borders, but rather for an international settlement that would give Germany a modest, sustainable empire in Central Europe – an empire much smaller than Hitler's ideal one but much larger than Germany in August 1939.

These facts demonstrate that the main threat perceived by German elites before the war was that of further income losses and another depression if Germany were to reopen trade and finance with the Western Powers. These facts also show that the elites were interested in modest territorial expansion to capture markets if it promised a more sustainable economic recovery than renewed economic relations with the West – and it was combined with the domestic economic policies they valued highly.

### 3.3.6 The domestic mythmaking theory fails to refute rival explanations

Rival Explanation #1: *Germany faced foreign threats to its core interests.*

Rival Explanation #1A: *Before World War I, Germany faced military threats from Russia and France.*

Rival Explanation #1B: *Before World War II, Germany was threatened by the restriction of its export markets and debt markets in France, England and the United States.*

Rival Explanation #2: *Even if Germany was not currently threatened by any of the other great powers, Germany had to protect her core interests from possible future opportunism or predation by the other great powers.*

Rival Explanation #3: *"German statesmen valued expansion rather than security."*

Rival Explanation #4: *By 1932, key elite groups in German agriculture, industry and the military preferred economic autarky through territorial expansion over renewed open trade with the Western Powers (with its risk of a repeat episode of trade restrictions and depression). Without these elites' acquiescence, support and participation – which the National Socialist party purchased by securing their domestic and international interests – the National Socialists never could have gained or maintained enough power to implement the strategy of*



*expansion and autarky.*

Rival Explanation #5: *Germany's foreign policies were optimal responses to the external threats coming from the international environment. Germany's domestic coalitions were optimal means of carrying out its optimal foreign policies.*<sup>34</sup>

### **3.3.7 Diversionary motives**

Examples include:

1. the narrow-minded political ambitions of ruling circles;<sup>35</sup>
2. domestic coalition formation and logrolling;<sup>36</sup>
3. the bureaucratic inertia of beliefs about relative power, strategic doctrines, military plans;<sup>37</sup> and
4. the diversionary motives of domestic politicians.<sup>38</sup>

## **4. The fourth generation (1995-2006): Game theoretic accounts of costly conflict due to relative power factors**

In the fourth generation scholars criticised the power shift theories of the second generation from a different angle. They accepted the core assumptions of the second generation:

1. Each state is a "unitary rational actor" whose policy decisions are optimal from the point of view of the state as a whole (e.g. there are no domestic political dysfunctions); and
2. Wars are caused by relative power factors.

But they identified a logical gap in the second generation theories. Those theories failed to explain why two states that act rationally would *ever* wage a war that imposes net losses

---

<sup>34</sup>Snyder 1991, "The international environment dictated both an optimal foreign policy and an optimal domestic coalition to carry it out."

<sup>35</sup>Kaiser 1990.

<sup>36</sup>Snyder 1991.

<sup>37</sup>Posen 1984; Van Evera 1984; Miller 1985; Levy 1986; C. Kupchan 1994.

<sup>38</sup>Morgan and Bickers 1992; Richards et al. 1993; Downs and Rocke 1994.

on *both* of them. Thus the puzzle had changed again. It was no longer "Why do shifts in relative power cause war?" The new puzzle was "Why do states go to war when they could settle their disputes peacefully and save the costs of war?" To put the puzzle another way: Why can't the two states split the difference in the dispute peacefully? If they could split the difference peacefully, then they could save the costs of war. The savings can always be shared between them in a way that would leave each state better off than it would be by going to war. So why would the two states ever go to war?

By defining the analytical problem in this way, the fourth generation theorists specified the key condition that a "rationalist" theory of war must meet. It must provide a rational explanation of why the two opposing states were unable to settle their dispute peacefully and thereby save the costs of war. This condition is particularly appropriate to use in assessing theories of hegemonic war, because these wars were the costliest wars in human history. In each case the historical evidence shows that decision-makers on both sides expected war to be a very costly strategy, so they had a strong rational incentive to avoid war. Their failure to avoid it requires an explanation. While the fourth generation theorists changed the puzzle, however, they retained the underlying *causal* assumption of the first and second generation theories: Wars are caused by relative power factors. Thus the new puzzle became: "How do *relative power factors* prevent states from settling their disputes peacefully, thus forcing them to go to war?"

Scholars of the fourth generation offered two main theories to resolve this puzzle. Each theory demonstrated how a decision to initiate war might be explained rationally. However, each theory suffers from both logical and empirical flaws that render it incapable of explaining hegemonic wars. First consider the logic of each theory and then consider its flaws that prevent it from explaining hegemonic wars.

#### **4.1 Wars due to unusually large power shifts**

This theory assumes that war becomes unavoidable once a power shift becomes *larger* than a critical threshold. The logic seems straightforward. As long as a power shift is *smaller* than the critical threshold, there is always a negotiated settlement that would leave both states better off than they would be fighting a war. But if the power shift is *larger* than

the critical threshold, then there is no negotiated settlement that would leave both states better off than they would be fighting a war. For after the rising state's power increases by *that much*, it can expropriate *so much* from the declining state that the rising state would need to receive a very large concession from the declining state to be persuaded to accept a settlement rather than pursue expropriation. But the declining state would be better off initiating a war to prevent the power shift from occurring in the first place than giving the rising state such a large concession.

While the theory seems logical, there are four problems with it from a Lakatosian perspective. Its main observable implications are refuted by the historical evidence. Its underlying observational theory is also refuted by the historical evidence. It includes a *ceteris paribus* condition that amounts to ruling out alternative rival explanations by assumption rather than disproof. And it includes an auxiliary hypothesis that might simply be a "conventionalist strategem for harmonizing theory and facts".<sup>39</sup> Consider each problem in turn.

#### 4.1.1 The anomalies

The theory generates two main observable implications:

1. The war will be initiated by the declining state: a preventive war intended to stop the increase in the rising state's power.
2. The declining state will initiate the war early in the power shift and win the war: the rising state will be prevented from using its growing power to expropriate the declining state.

Both implications are refuted by the evidence on hegemonic wars in Europe since 1550. This evidence is summarized in the following table, which lists each war and the two main opposing states in the war. The state that was rising in relative power prior to the outbreak of war is denoted by (r). The state that was declining in relative power prior to the outbreak

---

<sup>39</sup> As Lakatos writes, "[T]heories and factual propositions can always be harmonized with the help of auxiliary hypotheses.... [S]aving a theory with the help of auxiliary hypotheses which do not [satisfy well-defined conditions] represents [scientific] degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, 'conventionalist stratagems'." (Lakatos70)

of war is denoted by (d).<sup>40</sup> The state that initiated the war is indicated in boldface type. In three of the seven cases, it was the rising state that initiated the war, not the declining state. These cases refute the first implication (that the declining state initiates the war), so they are anomalies for the theory. In the other four cases, the declining state initiated the war, as predicted by the theory. But in two of these cases, the evidence shows that the declining state initiated the war late in the power shift and lost the war: The rising state succeeded in using its growing power to expropriate the declining state. These cases refute the theory's second implication (that the declining state initiates the war early in the power shift and wins the war, preventing the expropriation gambit of the rising state).<sup>41</sup>

#	Hegemonic War	State Emitting Spillovers	State Harmed by Spillovers	Impl 1	Impl 2
1	Eighty Years War	Dutch Netherlands (r)	<b>Spain</b> (d)		anomaly
2	Thirty Years War	Bohemia (r)	<b>Austria</b> (d)		anomaly
3	Nine Years War	England (d)	<b>France</b> (r)	anomaly	
4	War of the Spanish Succession	Austria (r)	<b>Spain</b> (d)		
5	French Revolutionary Wars	France (d)	<b>Austria</b> (r)	anomaly	
6	World War I	Serbia (r)	<b>Austria</b> (d)		
7	World War II	France (d) England (d)	<b>Germany</b> (r)	anomaly	

Note that in all seven cases, it was the state harmed by the spillover effects (e.g. the state suffering the income losses from the spillover effects) that initiated military hostilities first. This fact constitutes *prima facie* evidence that it was the *income losses* generated by the

---

<sup>40</sup>I code the *rising* state as the state whose military labor productivity was rising relative to its civilian labor productivity, and the *declining* state as the state whose military labor productivity was declining relative to its civilian labor productivity. These codings are derived from Parts II through IV of this study. There I present a theory in which the international spillover problem in each historical case affected the productivity of military labor relative to civilian labor in each state. This changed each state's *power*, defined as the number of men it had an incentive to transfer from its domestic economy into its military. This type of power shift drove all of the other types of power shift that occurred before and during each war.

<sup>41</sup>In the Eighty Years War, the Dutch Netherlands declared autonomy from the Spanish Monarchy and expropriated their large tax base from Madrid. In the Thirty Years War, the Kingdom of Bohemia declared autonomy from the Austrian Monarchy and expropriated its large tax base from Vienna.

spillover problem that caused the war, not any power shift arising from the spillover problem.

It is worth looking at these empirical anomalies from another angle. The theory of war due to large power shifts predicts victories in situations where defeats actually occurred; and it predicts defeats in situations where victories actually occurred.

1. The theory predicts that a potentially *declining* state will enter a war quickly and win in situations where it actually stayed out and lost.
2. The theory predicts that a potentially *declining* state will enter quickly and win in situations where it actually entered quickly and lost.
3. The theory predicts that a potentially *declining* state will fight and win alone in situations where it actually capitulated alone and could only reenter the war by gaining numerous allies.
4. The theory predicts that a potentially *rising* state will fight early and lose against a single opponent in situations where the rising state actually fought early and won against numerous opponents.
5. The theory predicts that a potentially *rising* state that fights late in its rise will win, whereas rising states that actually fought late in their rises lost.

In short, this theory "leads to strained or even bizarre historical readings"<sup>42</sup> of actual hegemonic wars.

#### **4.1.2 The underlying observational theory is refuted by the historical evidence**

The theory of war due to large power shifts is based on an underlying observational theory for turning the historical evidence into "observations". This observational theory specifies how to sift through the raw historical evidence, select the pieces of evidence that are necessary to build and test the substantive theory, and code those pieces into "observations". In particular, the observational theory specifies how to measure and code the "power shifts" that supposedly cause war and the phenomenon of "war" that they supposedly cause. In the seven cases of hegemonic war, however, this observational theory is refuted by the evidence.

---

<sup>42</sup>Powell 2006, p. 173.

1. *Should power shifts be coded as exogenous or endogenous?*

The observational theory specifies that power shifts are assumed to be exogenous (e.g. beyond anyone's control). This assumption is refuted by the evidence on hegemonic wars. In four of the seven cases, it was only *after* negotiations to solve the spillover problem failed due to the legal incompatibility problem that one or both of the main opposing states made the discretionary decision to increase its power. This fact suggests that the resulting power shifts were not exogenous. They were endogenous to the problem of legal incompatibilities.

The problem of legal incompatibilities caused power shifts through two channels. First, the legal incompatibility problem rendered the international spillover problem unresolvable through negotiations. Each of the main opposing states perceived that if it were to agree to a negotiated settlement that allowed the other state to gain on net from the matter, it would give the other state an incentive to commit further opportunism of a similar kind. Hence each state perceived that its only option was to wage war against the other state to insure it suffered a net loss from the matter – and that option required each state to increase its military power through some combination of domestic mobilization, alliance formation and territorial hegemony-seeking.

Second, once the legal incompatibility problem rendered the dispute over the international spillover problem intractable, the spillovers caused changes in labor productivity that shifted relative power. Each of the main opposing states experienced either a rise or a decline in its military labor productivity relative to its civilian labor productivity.<sup>43</sup> These shifts in relative power shaped state strategies and international outcomes in many ways (as shown in Parts II through V of this study). In particular, the state that experienced a decline in its military labor productivity was forced to find natural allies in the dispute or to expand territorially and impose a coercive hegemony on neighbors to secure allies.

All of these power shifts were endogenous to the problem of legal incompatibilities. By assuming that power shifts are endogenous to this deeper cause, the theory of Contractual Realism is able to explain a variety of types and causes of power shifts *without* assuming that

---

<sup>43</sup>Secession spillovers *increased* the military productivity of the emitting state and *decreased* the military productivity of the harmed state. Depresssion spillovers *decreased* the military productivity of the emitting state and *increased* the military productivity of the harmed state. Revolution spillovers had neither effect, but when the harmed state gained military options to solve the spillover problem by military force, this *increased* its military productivity and *decreased* the emitting state's military productivity by comparison.

power shifts are the reason why war becomes inevitable. In this sense, Contractual Realism refutes the Conventional Realist claim that power shifts are the root cause of war by beating that school at its own game: Understanding the true relationship between power shifts and war.

One might object that there is one type of power shift that is truly exogenous (e.g. beyond anyone's control). That is power shifts due to uneven economic growth over the long run. Although that type of power shift is truly exogenous, it is the exception that proves the rule. For the Contractual Realist theory demonstrates that that type played no role in causing the hegemonic wars in Europe between 1550 and 1950 (or in shaping the state strategies and international outcomes that emerged during and after each war). The theory specifies the necessary and sufficient conditions for hegemonic war to occur; and none of these conditions include changes in relative power due to uneven long-run growth. In other words, the theory assumes that there were no shifts in relative power due to uneven long-run growth – and yet each war erupted anyway. The theory generates a large number of observable implications for state strategies and international outcomes before, during and after each war; and all of the implications are corroborated by the historical evidence. This combination of causal assumptions and empirical corroboration proves that none of these wars was caused by uneven long-run growth. Thus, the one type of power shift that is truly exogenous played no role in causing these wars or shaping strategies and outcomes during and after these wars. Thus it is safe to say that all of the different types of power shifts that *did* matter in these wars were endogenous. And the deeper cause driving all of these power shifts was the problem of legal incompatibilities. Since these are the main wars that Conventional Realists claim were caused by power shifts, this conclusion constitutes a resounding refutation of the Conventional Realist focus on power shifts as a cause of war.

*2. One way to define the "war" to be explained is: the initiation of military hostilities. Another way to define the "war" to be explained is: the initiation and prolongation of military hostilities. Which definition should be used to test the theory of war due to large power shifts?*

That theory adopts the first definition. The "war" to be explained is: the initiation of military hostilities. This approach appears to be validated by some of the evidence on hege-

monic wars. In some cases the declining state aborted the negotiations over the international spillover problem and initiated a war in an attempt to solve the spillover problem by military force. Even in these cases, however, the state that lost the first round of the war did not stay out of the war (as the theory predicts). Rather it secured allies and reentered the war. This pattern of evidence calls into question the assumption that the "war" to be explained is simply: the initiation of military hostilities. It is worth spelling out this critique more clearly.

The theory of war due to large power shifts assumes that once the declining state initiates the war, there are only two possible outcomes – and neither outcome leads to a long and costly war. On one hand, if the declining state waits too long to initiate the war, then the rising state's power will have grown by so much by that point that it can commit expropriation for a net gain despite the declining state's military opposition. In this case the declining state will stop fighting quickly, because it foresees that it cannot avoid being expropriated no matter how long it prolongs the war. To be precise, the declining state recognizes that it waited too long to initiate the war, so the power shift has already progressed so far that it cannot reverse the power shift no matter how long it prolongs the war. This is precisely the outcome that the declining state would wish to avoid by adopting a strategy of starting the war *sooner*.

On the other hand, if the declining state initiates the war early enough in the power shift, then it will succeed in stopping the power shift before the rising state becomes powerful enough to commit expropriation for a net gain. In this case the rising state will stop fighting quickly, because it foresees that it will not gain enough power to commit expropriation and hence it will not reap a net gain from the war no matter how long it prolongs the fighting. To be precise, the rising state recognizes that the war began too early in the power shift, so the power shift cannot be completed no matter how long it prolongs the war.

In either case, therefore, the theory predicts a short war that is cheap for at least one side. In the first case the war is cheap for the rising state, because it succeeds in committing expropriation for a net gain. In the second case the war is cheap for the declining state, because it suffers fewer net losses from the war than it would suffer by allowing the power shift to proceed and winding up expropriated.



These are exactly the patterns that occurred in the *first round* of each hegemonic war (in which the first round was a one-on-one fight<sup>44</sup>). In three of these four cases, the declining state waited too long to initiate the war and ended up being expropriated. In the other case, the declining state initiated the war early enough and stopped the power shift before the rising state could become strong enough to commit expropriation.

- **The Eighty Years War.** The declining state, Spain, waited too long to initiate war against the rising state, the Dutch Netherlands. The Dutch succeeded in expropriating their tax base from the composite Spanish monarchy. Spain quickly dropped out of the fighting in the Dutch provinces and stayed out until it could secure allied support (in South America by coercion and in Genoa by compensation).
- **The Thirty Years War.** The declining state, the Kingdom of Austria, waited too long to initiate war against the rising state, the Kingdom of Bohemia. The Bohemians succeeded in expropriating their tax base from the composite Austrian monarchy. Austria dropped out of the fighting quickly and stayed out until it could secure allies (in Spain by common interest and in Bavaria by compensation).
- **The French Revolutionary Wars.** The declining state, France, initiated war quickly enough to prevent the rising state, Austria, from invading and restoring the old regime of feudal expropriation. Austria quickly dropped out of the fighting and stayed out until it could secure allies.
- **World War II.** The declining states – Poland, France and Britain – waited too long to initiate war against the rising state, Germany. The Germans succeeded in expropriating Poland and France (which dropped out of the fighting quickly). England also dropped out of the fighting quickly and stayed out until it could secure allies.

In all four cases, the loser of the first round dropped out of the fighting quickly rather than fighting on without allies. According to the theory *that should have been it* – the war should have been over. But instead, the loser sought allies and reentered the war; and it became a

---

<sup>44</sup>In the other cases, one or both of the main opposing states found an ally before the fighting began, thus making it a two-on-one fight or a two-on-two fight.

long and costly war for both sides. This evidence suggests that the theory can only explain the *initiation* of military hostilities by the declining state. It cannot explain the *prolongation* of the fighting by both states to the point that it becomes a long and costly war.

This evidence also refutes the theory in another way. The theory assumes that each state's main aim in the war is to knock out the other state *so that it stays out* by stabilizing the distribution of relative power *permanently*. The declining state's aim is to knock out the rising state before its power grows by too much, so that it *never* gains enough power to expect a net gain from war again. In turn, the rising state's aim is to delay the start of the war or accelerate the rise in its power, so that its power becomes great enough to expropriate the declining state *for good*. Thus the theory predicts that, whichever state winds up losing the war in a one-on-one fight, the loser stays out for good. The theory simply cannot explain why the loser would go searching for allies to be able to reenter the war again. In fact, there were multiple rounds of exit and reentry in each of the first three wars listed above. In each round, one of the main opposing states lost, dropped out of the fighting, secured allies, and then reentered the fighting to start up the next round. The theory of war due to large power shifts simply cannot explain this evidence. For it assumes that the winner only entered the war in the first place because it expected to be able to stabilize the distribution of relative power permanently. If it had succeeded in this goal in the course of winning the first round, then how did the loser ever expect to succeed in shifting the distribution of relative power back in its favor again? Or: If the loser's expectation of success was accurate, then how could the winner ever have expected to prevent such reversals permanently?

Thus, the power shift theory might be able to explain the *timing of the start* of each hegemonic war. (That is, in cases where the state harmed by the spillover problem was *declining* in relative power, the war started *quickly*. But in cases where the state harmed by the spillover problem was *rising* in relative power, the war started slowly.) But that theory cannot explain why the two main opposing states prolonged the fighting to the point where it became a long and costly war. To explain this observation, one must explain: (1) why the loser of each round perceived a need to reboost its power and reenter the war, and (2) how the loser ever expected to be able to shift the distribution of relative power in its favor again after losing that round.

In sum, the theory of war due to large power shifts is simple in its basic logic. But both the substantive theory and the underlying observational theory suffer from many anomalies. This suggests that some amendment or addition is needed to save the theory. Let us examine the measures taken by the theory's main developer to save it. Do these measures constitute progressive or degenerative problemshifts?

#### 4.1.3 The *ceteris paribus* condition

The theory assumes that the declining state can stop the rising state's increase in power through a preventive war. At the same time, however, the theory also assumes that the *rising* state *cannot* stop the increase in its own power by voluntary agreement (to save itself the cost of such a war). These assumptions are contradictory. If one assumes the declining state has the capacity to stop the power shift through preventive war, then one should also assume the rising state has the capacity to stop the power shift by voluntary agreement if this would leave it better off than it would be after such a war. By assuming that such an agreement is impossible, the theorist is effectively adding a *ceteris paribus* condition to the theory. The amended theory is as follows.

Once a power shift becomes larger than a critical threshold, war becomes unavoidable – as long as no other causes are operating to avoid war. But if another cause is operating, such as the capacity of the rising state to stop the increase in its own power by negotiated agreement, then the theory does not apply. War is avoidable.

In this amended version of the theory, *two* conditions are necessary for war to occur: (1) the power shift is larger than a certain threshold, and (2) there is some *obstacle* to a negotiated settlement by which the rising state agrees to stop the increase in its power. The first condition alone is not sufficient to cause a war. If only the first condition is present, then peace can be maintained even in the face of a large power shift.

Now consider the same logic from the opposite point of view: the necessary and sufficient conditions for peace to be maintained (according to the amended version of the theory). There are two conditions for peace to be maintained, and either condition alone is sufficient:

(1) the power shift is smaller than the critical threshold, or (2) the opposing states can negotiate a settlement in which the rising state agrees to stop the increase in its power. The methodological implication is obvious. If a researcher wants to prove that a war was caused by this causal mechanism (an unusually large power shift), then he or she must explain why *both* conditions failed. That is, the researcher must prove not only that the power shift was larger than the critical threshold, but also that there was some obstacle that prevented the two states from negotiating a settlement by which the rising state would agree to stop the increase in its power.

It is true that in the run-up to each of the seven hegemonic wars, the two main opposing states tried to negotiate a settlement that would avoid a war, but they failed. How does this theory account for such negotiation failures?

#### 4.1.4 The auxiliary hypothesis

The theory's main developer seems to acknowledge that a large power shift alone may not be sufficient to cause a war. Causation may also involve some kind of "informational problem" that causes a negotiation failure.

"The basic idea behind this [theory's] approach is to study war and the inefficiency puzzle in the context of complete information games where there are no informational problems. This approach, it is important to emphasize, should not be seen as discounting the role of informational accounts in explaining key aspects of war. As just noted, informational arguments have made fundamental contributions. Rather a complete-information approach simply lets one abstract away from informational problems to focus more directly on other possible solutions to the inefficiency puzzle.... [This complete-information approach focuses on] commitment problems [and] may help to overcome the limitations of informational accounts, *either* as a complement to an underlying informational problem *or* as the primary cause of conflict."<sup>45</sup>

Thus the author is suggesting two alternative interpretations of his theory. In the "primary

---

<sup>45</sup>Italics added. Powell 2006, pp. 170–1.

cause” interpretation, a large power shift alone is sufficient to cause a war. In the ”complement” interpretation, a large power shift alone is *not* sufficient to cause a war. But the two problems together – a large power shift and an informational problem – *are* sufficient to cause a war. The author thus acknowledges that his theory might need to include an auxiliary hypothesis: There is an informational problem that ”complements” the power shift in making war unavoidable. Is this auxiliary hypothesis a progressive problemshift that is needed to save a valid theory? Or is it merely a ”conventionalist strategem needed to harmonize theory and facts”? This depends on what exactly the ”informational problem” is and what kind of negotiated settlement it prevents. The theorist does not specify.

In the case of the hegemonic wars in Europe, one can imagine two very different kinds of negotiated settlement that might have been blocked by an informational problem of some kind:

1. An agreement by which the rising state would commit to stop the increase in its power.
2. An agreement to settle the dispute over the international spillover problem through some combination of mutual concessions:
  - the emitting state would commit to reduce some of the spillovers in order to reduce the income losses suffered by the harmed state, and in return
  - the harmed state would commit to accept a lower level of income losses due to the remaining spillovers.

The historical evidence shows that in each case of hegemonic war, the main opposing states tried to reach the second type of settlement but failed. Why did they fail?

The evidence shows that they failed due to a problem of *legal incompatibility*. The existing international laws gave *both* states the right to the *same* asset. The emitting state had the right to continue its domestic activity that was generating the international spillover effects. But the harmed state had a right to be free of the income losses imposed by those spillovers, and hence a right to the *cessation* of that domestic activity. The historical evidence shows that the demands and counter-demands issued by the two opposing states in the diplomatic crisis prior to the outbreak of war pertained to these overlapping rights. Yet none of these rights related to power or power shifts.

So why did the legal incompatibility cause negotiations to fail? The evidence shows that the legal incompatibility caused an informational problem, but it was a problem that had nothing to do with power or power shifts. Each state lacked the information it needed to know whether the opposing state was committing contractual opportunism in the dispute over the spillover problem. The state harmed by the spillovers lacked the information it needed to know whether the state emitting the spillovers was emitting them intentionally or unintentionally. In turn, the emitting state lacked the information it needed to know whether the harmed state *truly* didn't know it was emitting the spillovers unintentionally, or was ignoring evidence the spillovers were unintentional in order to *claim* it didn't know that and make demands for its own gain.

This complex informational problem was rendered intractable by the legal incompatibility between the states' respective rights in the dispute. On one hand, the emitting state had a right to engage its new domestic activity – which was only generating the spillovers as an *unintended secondary effect*. It held firmly to this right and showed every intention of continuing the domestic activity. The harmed state perceived such conviction as evidence that the spillovers were being emitted intentionally, however. On the other hand, the harmed state had a right to be free of the income losses imposed on it by the spillovers, and hence a right to the cessation of the emitting state's domestic activity. The harmed state demanded the cessation of that activity and held firmly to this demand based simply on its right to be free of the income losses. But the emitting state perceived such conviction as a sign that the harmed state was ignoring evidence the spillovers were unintentional to be able to interfere in its domestic activities.

Thus the legal incompatibility problem gave each state the conviction that it was in the right. But each state's conviction prevented the opposing state from obtaining the information it needed to know whether the dispute over the spillover problem had arisen intentionally or unintentionally on the other state's part. In this way the legal incompatibility problem created an intractable informational problem. Neither problem had anything to do with power or power shifts, however.

This explanation of negotiation failures raises a critical question. Why would a lack of information *about anything* ever require a state to go to war? The conventional wisdom

holds that it is a lack of information *about relative power* that requires states to go to war. When one state lacks information about another state's power level, it can bluff that it is more powerful than it really is in a ploy to extract concessions it could not extract otherwise. To deter such bluffing, the uninformed state must wage war in response to at least some of the informed state's demands. This informational cause of war is the "complement" that may be required to prove the theory that unusually large power shifts make war unavoidable (according to the theory's developer). Notice that both components of this theory are based on relative power concerns. The informational problem is driven by uncertainty about relative power, while the commitment problem is driven by the notion that unusually large shifts in relative power necessitate preventive war.

In a moment I will examine the conventional wisdom that uncertainty about relative power is the main informational cause of war (in section 6). For now the key point is that this is not the only informational cause of war that one might imagine. The alternative theory developed in this study – Contractual Realism – is based on a different informational problem that has nothing to do with relative power. Here it is a lack of information about intentionality in the emergence of contract disputes that requires states to go to war. When a legal incompatibility generates a contract dispute between two states, each state lacks information about whether the dispute arose intentionally or unintentionally on the other state's part. Hence each state anticipates that if it agrees to a compromise settlement in which the other state receives a net gain from the matter, then it will give the other state an incentive to create further disputes intentionally for its own gain. So each state calculates that it must wage war over the current dispute to impose enough costs on the other state to insure it suffers a net loss from the dispute – simply to deter it from creating further disputes intentionally for its own gain.<sup>46</sup>

This cause of war has nothing to do with relative power. Even if each state has full information on the other state's power level and there are no power shifts, a legal incompatibility problem still causes war in this way. A main goal of this study, therefore, is to specify this theory of hegemonic war, derive as many observable implications as possible from it, and

---

<sup>46</sup>The original source of this approach to explaining conflict is Rowe (1989, 1990). While this source inspired my Contractual Realist theory of conflict, my theory is novel in the sense that it makes additional assumptions that were not made in this source – and my theory shows why these assumptions are necessary.

present the evidence that corroborates these implications. In this way the study will show that the hegemonic wars in Europe were caused by legal incompatibility problems, not power shifts.



## 4.2 Theories of war due to uncertainty about relative power

This theory assumes that war is caused by an informational problem related to relative power. When one state is uncertain about another state's power, it can bluff that its power has increased even when it has not. Then it can demand concessions from the first state, backed by a threat that its power has increased by enough to extract the concessions by force with a net gain – even when this isn't true. The only way to stop it from bluffing in this way is to impose a cost on it by rejecting the demand and going to war over the demand (at least some of the time).<sup>47</sup> As long as the uncertain state goes to war over a high enough proportion of such crises, the demanding state will suffer a net loss from issuing demands based on bluffs that its power has increased, and it will be deterred from such bluffing. But when its power really has increased by enough to issue demands and extract concessions by force with a net gain, the uncertain state does not know this fact (because power shifts are assumed to be unobservable in this theory). So it must go to war over the matter simply to deter the demanding state from bluffing about its power. The consequence is war.<sup>48</sup>

This theory has been termed the "bargaining theory of war", because its aim is to explain why diplomatic bargaining crises sometimes break down in war. The theory has generated a large literature that builds on this basic logic.<sup>49</sup> While the theory seems logical, it has several problems from a Lakatosian perspective. First, its main observable implications are refuted by the evidence on hegemonic wars. Second, although it may explain why diplomatic bargaining crises break down in war, it does not explain why such crises arise in the first place. Any theory of war must explain this too. The only way the bargaining theory of war can explain it, however, is by adding an auxiliary hypothesis that is scientifically degenerative. Third, the bargaining theory of war is based on an observational theory for selecting and coding evidence that generates observable implications of its own. But these implications are refuted by the historical evidence on hegemonic wars as well as other

---

<sup>47</sup> Fearon 1995; Powell 2002.

<sup>48</sup> Fearon (1995) and Powell (2002). Fearon assumes that the root cause of the problem is the risk-reward tradeoff that arises in principal-agent relationships due to risk aversion. However, technical game theorists have shown that the principal-agent problem does not arise from risk aversion. It arises simply from the information asymmetry between the principal and the agent, even when both actors are risk neutral.

<sup>49</sup> Powell 2002; **CITE**.

evidence on international relations. Consider each problem in turn.

#### 4.2.1 The anomalies

The bargaining theory of war generates two main observable implications.

1. When a bargaining crisis arises, only *one* of the two opposing states will issue a demand: the state claiming that its power has increased by enough that it could secure its demand by force if necessary. This state's demand will be the only demand in dispute in the crisis. In response, the other state will simply reject the demand and initiate war over it.
2. The ensuing war will be short, because the targeted state's uncertainty about relative power will be resolved quickly by the fighting itself.<sup>50</sup>

The first implication is refuted by the evidence on hegemonic wars. In the diplomatic crisis prior to the outbreak of each war, one state did issue a demand. But the other state did not simply reject the demand and initiate war over it. Rather, the other state issued a counter-demand *pertaining to a different right that was also in dispute*. In each case the sequence was as follows. First, the state harmed by the spillover effects claimed that according to one law the emitting state had no right to emit the spillovers, and demanded that it stop its new domestic activity that was emitting them. In response, the emitting state did not simply reject the demand and initiate war over it. The emitting state claimed that according to *another* law the harmed state had no right to interfere in its domestic activities, and demanded that the harmed state stop doing so. These observations show that the demand and counter-demand were driven by the problem of overlapping and inconsistent rights – what I have called the problem of legal incompatibility – not by any incentive to bluff about relative power. The second implication (that the war will be short) is also refuted by the evidence on hegemonic wars. These conflicts were seven of the ten longest and costliest wars in human history.<sup>51</sup>

---

<sup>50</sup>Powell 2006.

<sup>51</sup>CITE.

### 4.2.2 The auxiliary hypotheses

The bargaining theory of war may explain why some diplomatic bargaining crises end in war. However, it does not explain why such crises arise in the first place. Why do states that are at peace *ever* make demands on other states? And why do the targeted states *ever* stand firm in response, thereby creating bargaining crises? There are three possible answers to this question. The first two answers are consistent with the bargaining theory of war, but scientifically degenerative in one way or another. The third answer is inconsistent with the bargaining theory of war, but consistent with the rival theory developed in this study: Contractual Realism.

#### 4.2.2.1 Why bargaining crises arise (I): We don't need to know

The first answer is simply to say that it does not matter why bargaining crises arise. They may arise for any number of reasons, but the reasons are irrelevant to the question of why some of the crises end in a negotiated settlement while others end in war. This first answer says, in effect, that when two states get into a diplomatic bargaining crisis and the crisis ends in war, a scientific explanation of this outcome does not need to explain why the bargaining crisis *arose* in the first place. It simply needs to explain why the crisis *ended* in war.

This answer is scientifically degenerative for two reasons. First, it ignores a necessary condition for such wars to occur, namely, that the crises arose in the first place. By failing to explain how this condition comes about, it fails to provide the full set of necessary conditions for the causation of war. Second, it assumes that bargaining crises may arise for many different reasons that have little or nothing in common. So, if an explanation of their occurrence must be provided, each of these myriad reasons becomes another *ad hoc* auxiliary hypothesis that must be included for the theory to prove causation. To put the point another way, this answer ignores the possibility that there is a single *general* reason why bargaining crises arise in the first place. Hence it ignores an opportunity for theoretical generalization and assumes that scientific inquiry for that purpose is irrelevant.

#### 4.2.2.2 Why bargaining crises arise (II): New opportunities to bluff about relative power

The second answer is that bargaining crises arise because states are constantly looking for new opportunities to bluff about their power. If they can bluff that their power has increased when it really hasn't, then they can make demands backed by threats to secure the demands by force if necessary – and get away with it even if their threats are not credible. The only way to deter such bluffing is to reject demands backed by threats that may not be credible and wage war over them. The assumptions behind this understanding of international relations must be identified clearly, because the first two assumptions constitute hidden auxiliary hypotheses in the bargaining theory of war. So it is critical to ask whether they are scientifically degenerative. Accordingly, I will identify these assumptions, show that they lie at the root of this understanding of international relations, and show that their main observable implication is refuted by the historical evidence on hegemonic wars. Then I will provide an alternative understanding of international relations based on the *opposite* assumptions and show that its main observable implication is corroborated by the evidence. Finally, I will identify the observational theories that underlie these rival understandings of international relations, derive *their* main observable implications, and show that the first one is refuted by the evidence while the second one is corroborated by the evidence.

Here are the assumptions behind the view that bargaining crises arise because states are constantly looking for new opportunities to bluff about their power.

1. States divide valuable assets according to relative power at all times. Even in peacetime, the distribution of assets shifts between states whenever the distribution of military power shifts between states.
  - (a) Whenever relative power shifts by enough that the favored state could reap a net gain by seizing some of the disfavored state's assets by military force, the favored state issues a demand for assets backed by a threat to seize them if the demand is not met;
  - (b) Whenever the state targeted by such a demand is *certain* that relative power has shifted against it by that much, it complies with the demand peacefully to avoid

the costs of a war over the issue. (It would lose the war anyway, so its war costs would be a pure waste of its resources.)

- (c) As a result of demands and concessions like these, the distribution of assets between states shifts whenever the distribution of military power shifts.

2. Each state has four main foreign policy goals:

- (a) To respect the existing distribution of assets between states as long as the distribution of relative power remains unchanged;
- (b) To look for situations where its true power level has increased by enough that it can expropriate assets from another state by making demands and credible threats;
- (c) To look for situations where another state is *uncertain* about its true power level, so that it can *bluff* it is more powerful than it really is, in order to make demands and secure assets that it could not secure were its power level known; and
- (d) To estimate other states' power levels in order to determine whether their claims about their power are true – and respond accordingly.

3. When the targeted state is *certain* about the demanding state's power, the targeted state can respond differently in the two cases:

- (a) If the demanding state's power *has* increased by enough that it could reap a net gain from a war over the matter, then the targeted state acquiesces to the demand; but
- (b) If the demanding state's power has *not* increased by that much, then the targeted state rejects the demand.

4. When the targeted state is *uncertain* about the demanding state's power, the targeted state cannot respond differently in the two cases. Its optimal response is simply to reject the demand. By rejecting the demand it creates a bargaining crisis. Some of these crises end in war simply because the targeted state must deter the demanding state from bluffing that its power has increased by enough to prevail in a war when it really hasn't.

5. In this type of conflict, the targeted state wages war for long enough that a demanding state that was only *bluffing* that its power had increased will suffer a net loss in the

war. Observing that the war lasts for this long, *all* states will expect to suffer a net loss if they run the same gamble – bluffing their power has increased when it really hasn't. In equilibrium, all bluffing about power will be deterred: No state will bluff about its power. While states may constantly *look* for new opportunities to bluff about their power, and they may *see* some opportunities, no state will ever *seize* such opportunities.

6. Wars of this kind will be short and low in cost, because one of the two opposing states will drop out of the fighting quickly.<sup>52</sup> In either case the targeted state wages war for the same length of time: Long enough that the demanding state would suffer a net loss *if* it was only bluffing that its power had increased. The targeted state does not know whether the demanding state was bluffing or not, however, until this length of time is reached.
  - (a) If the demanding state was only bluffing that its power had increased, then *it* will drop out of the fighting quickly once the war has lasted long enough for it to suffer net losses. For it would only incur greater net losses by continuing to fight on in the absence of any new power advantage. (Wars of this kind will not occur in equilibrium, since bluffing is deterred completely in equilibrium.)
  - (b) If the demanding state's power really had increased, then the *targeted state* will drop out of the fighting quickly once it has fought for long enough to impose a net loss on a demanding state that was only bluffing its power had increased. For the targeted state would only incur greater net losses by continuing to fight on against an opponent with such a new power advantage.

This understanding of international relations is driven by the first two assumptions. States divide assets according to relative power at all times, not only in war but in peace as well. And states are constantly looking for opportunities to bluff about their power. If either assumption were *untrue*, this perspective would *not* provide a good general explanation of bargaining crises and war. If states did not divide assets according to relative power in peace time, then states would have nothing to gain by bluffing about their power. So uncertainty

---

<sup>52</sup>As Powell (2006) expressed it, wars due to uncertainty about relative power will end quickly because the targeted state's uncertainty about the demanding state's power will be resolved quickly by the fighting itself.

about relative power and bluffing about power could not explain the eruption of bargaining crises and war. And even if states did divide assets according to relative power in peace time, this model would still fail to explain crises and war if no state ever searched for opportunities to bluff about its power or seized the opportunities it found.

The main observable implication of this understanding of international relations is that wars will be short and low in cost. This implication is refuted by the historical evidence on the hegemonic wars in Europe since 1550 (all of which were long and costly). These anomalies cast doubt on this understanding of international relations. So consider an alternative understanding of international relations, based on the opposite assumptions, which does explain long and costly wars.

#### **4.2.2.3 Why bargaining crises arise (III): New opportunities to bluff about something else besides relative power**

According to this approach, states do not divide valuable assets according to relative power in peace time. They divide valuable assets according to contracts whose terms are *invariant* to shifts in relative power. The terms are tied to anchors in the material world that do not move even when relative power shifts among states (hence the label *anchor term contracts*). In this understanding of international relations, the main cause of bargaining crises is not bluffing about power. The main cause is contract disputes that arise from ambiguities in the contract terms. The main cause of war is the need to deter actors from creating such disputes intentionally for their own gain – by moving real world events into the gaps between the contract terms and then bluffing that the resulting contractual ambiguity arose unintentionally on their part. This cause of war has nothing to do with relative power. Even if there were no shifts in relative power and no uncertainties about relative power, this problem would still cause war.

It is important to identify the assumptions behind this understanding of international relations as well, because they form the *empirical basis* of Contractual Realism (in the Lakatosian sense of the term). Note that each of these assumptions is the opposite of its counterpart in the list provided above.

1. In war time, states divide assets according to relative power. In peace time, states

divide assets according to contracts whose terms are *invariant* to changes in relative power. The terms are written to refer to anchors in the real world that do not move when relative power shifts. Since the world is always changing, the anchors sometimes move randomly due to forces beyond any state's control. The contracts specify how the distribution of assets between states will be allowed to shift when the anchors move due to random forces beyond any state's control.

- (a) When an anchor moves in one state's favor due to random forces beyond its control, it requests a greater share of the asset governed by that contract from the other state (e.g. the state disfavored by the anchor's movement);
- (b) As long as the disfavored state is certain that the anchor moved due to random forces beyond the favored state's control, it permits the favored state to receive a greater share of the asset. That share is the favored state's *random variable benefit*.

2. Each state has four main foreign policy goals:

- (a) To respect the current distribution of assets specified by the existing contracts as long as the real world anchors remain in the same place – regardless of shifts in relative power between states;
- (b) To look for situations where an anchor has moved in its favor due to forces beyond its control, and request a greater share of the relevant asset as specified in the contract;
- (c) To look for situations where other states are *uncertain* whether the anchors are moving in its favor due to random forces beyond its control or its intentional actions, so it can move an anchor in its favor intentionally *but bluff the anchor moved in its favor unintentionally* – and request a greater share of the asset as specified in the contract; and
- (d) To judge whether an anchor has moved in another state's favor due to random forces beyond its control or its intentional actions – and respond accordingly.

3. When the disfavored state is *certain* about whether the anchor moved in the favored state's direction due to intentional or unintentional factors, the disfavored state can



respond differently in the two cases:

- (a) If the anchor moved due to random forces beyond the favored state's control, then the optimal response is to grant the favored state a greater share of the asset as specified in the contract; but
- (b) If the favored state moved the anchor in its own favor intentionally, then the optimal response is to reject its request for a greater share share of the asset; and if it has already taken that share, then the optimal response is to demand that it relinquish its illicit gains, and if it does not, then to impose the penalty of *random benefit withdrawal* (e.g. withholding the random variable benefits it would normally receive at other times when the anchor moved in its favor due to random forces beyond its control).

- 4. When the disfavored state is *uncertain* about whether the anchor moved in the favored state's direction due to intentional or unintentional factors, the disfavored state cannot respond differently in the two cases. Its optimal response is simply to reject the favored state's request for a greater share of the asset even though it has a right to a greater share in that situation according to the contract. By rejecting the request, the disfavored state creates a bargaining crisis. Some of these crises end in war simply because the disfavored state must deter the favored state from moving anchors in its own favor intentionally but bluffing they moved unintentionally.
- 5. In this type of conflict, the disfavored state wages war for long enough that a favored state that moved the anchor intentionally and bluffed it moved unintentionally will suffer a net loss in the war. Observing that the war lasts for this long, *all* states will expect to suffer a net loss if they run the same gamble – moving an anchor intentionally and bluffing it moved unintentionally. In equilibrium, all such bluffing will be deterred: No state will move an anchor in its own favor intentionally but bluff it moved unintentionally. While states may constantly *look* for new opportunities to engage in this kind of bluffing, and they may *see* some opportunities, no state will ever *seize* such opportunities.
- 6. The war lasts long enough that *each* state incurs enough net war losses to be deterred from committing opportunism in the future. On one hand, the disfavored state wages

war for long enough to deter the favored state from moving an anchor intentionally but bluffing it moved unintentionally. On the other hand, the favored state wages war for long enough to deter the disfavored state from bluffing it can't tell an anchor moved unintentionally when actually it can tell, e.g. from bluffing uncertainty about this when it has certainty. (For if it could bluff uncertainty, then it could deny a greater asset share to the favored state even in situations where it has a contractual right to it because the anchor moved in its favor due to unintentional factors beyond its control). Thus all of this kind of bluffing will be deterred too: No state will bluff *uncertainty* about why an anchor moved in another state's favor when it has *certainty* about that. While states may constantly *look* for new opportunities to engage in this kind of bluffing, and they may *see* some opportunities, no state will ever *seize* such opportunities.

If the anchor moved in a way that was codified in the contract, then the opposing states may be able to shorten or avoid the war by agreeing to establish a monitoring and adjudication system to distinguish between intentional and unintentional movements of the anchors. Such a system can prove that the anchor moved in one state's favor due to random forces beyond its intentional control – in situations where this is the case – and thereby enable the two states to shorten or avoid a war over the matter. Both states have a strong incentive to agree on such a system: to reduce or avoid their war costs.

Such an agreement is impossible, however, if the anchor moved in a way that was *not* codified in the contract and the resulting contractual ambiguity creates a problem of legal incompatibility. (Recall that this is a situation where one contract term gives one state the right to the disputed asset, but another contract term gives the other state the right to the same asset.) In such cases it is *impossible* to prove that the anchor moved in the favored state's direction due to random forces beyond its control. For the favored state cannot possibly claim that it violated the other state's contractual right *unknowingly or unintentionally*. For example, in the early years of the French Revolution, the revolutionaries invoked a natural law of popular sovereignty and claimed that it gave them the right to abolish the French monarchy. When Austria objected that this would violate the pan-European law of the divine right of kings, the revolutionaries could not possibly claim to be

violating the law unknowingly or unintentionally.

Thus wars caused by legal incompatibilities cannot be shortened or avoided by agreeing to create a monitoring and adjudication system. In such conflicts, each state's only strategy to deter bluffing is to wage war for long enough that the other state receives a net loss from standing firm in the dispute and eventually receiving half of the disputed asset in the post-war settlement. The main observable implication follow naturally:

The more valuable is the disputed asset, the longer and more costly the war will be. If each state places a very high value on the disputed asset, then the war will be long and costly.

This understanding of international relations explains why each hegemonic war was so long and costly. In each case the disputed asset was highly valued by the opposing states.

In this understanding, the main cause of bargaining crises and war is *random* changes in the world (due to forces beyond anyone's control) that move the anchors in a way that was not codified in the contracts. This causes a dispute over the meaning of the contract terms. If either state allows the other state to receive a net gain from the dispute, then that state will have an incentive to create further disputes for its own gain in the future, *because each is always looking for opportunities to move anchors intentionally and bluff they moved unintentionally*. So even though the current dispute arose unintentionally on either state's part, each state must wage war over the dispute for long enough that the other state suffers a net loss from it. Only this will deter the other state from *seizing* opportunities to create disputes for its own gain by moving anchors intentionally but bluffing they moved unintentionally. Such wars are not caused by shifts in relative power or uncertainty about relative power. They are caused by the problem of *unintended* changes in the real world that cause contractual ambiguities and legal incompatibilities.

This understanding of international relations is driven by its first two assumptions. If either of these assumptions were *untrue*, this perspective would *not* provide a good general explanation of bargaining crises and war. If states did not divide assets according to anchor term contracts, then states would have nothing to gain by moving anchors intentionally and bluffing they moved unintentionally. So uncertainty about why an anchor moved would

not explain the eruption of bargaining crises and war. And even if states did divide assets according to such contracts, this model would still fail to explain crises and war if no state ever searched for opportunities to bluff in this way or seized the opportunities it found.

In sum, there is a sharp difference between the second and third answers to the question of why bargaining crises arise. The second answer was that states are constantly looking for new opportunities to bluff about their power. The third answer is that states are constantly looking for new opportunities to bluff about something else besides power. Thus scholars face a choice between two different understandings of international relations: the relative power approach (represented by the second answer) and the contractual approach (represented by the third answer). Which approach offers more explanatory power – more “excess corroborated content” to use Lakatos’ phrase? To assess this issue it must be recognized that each approach consists of a substantive theory *and* an observational theory. Each approach is a “set-up” in Lakatos’ terminology.

#### **4.2.3 Competing set-ups for analyzing the conditions for peace and the causes of war**

Each set-up offers a different *substantive* theory of the conditions for peace and the causes of war. Hence each approach requires a different *observational* theory for measuring and coding the evidence on the conditions for peace and the causes of war. Consider the substantive theory and then the observational theory for each approach.

According to the relative power approach, states divide assets according to relative power at all times, not only in wartime but in peacetime as well. So the main conditions for peace are:

1. The distribution of assets corresponds to the distribution of relative power, and
2. this correspondence is not perturbed by any large shifts in relative power or uncertainties about relative power.

If a war erupts, therefore, it must have been because relative power shifted or some uncertainty about relative power emerged. This view of war thus requires an observational theory for measuring power shifts and the level of information that statesmen have about power (e.g. full or limited information).

According to the contractual approach, by contrast, states divide assets according to contracts whose terms are invariant to changes in relative power. So the main conditions for peace are:

1. The contract terms are written to refer to anchors in the real world that do not move when relative power shifts;
2. the contracts specify how the distribution of assets will be allowed to change when the real world anchors move randomly due to forces beyond anyone's control;
3. the anchors move randomly *only* in the ways specified in the contracts; and
4. there is an adjudication system to distinguish between random movements of the anchors due to forces beyond anyone's control and movements caused by someone's intentional actions.

If a war erupts, therefore, it must have been because an anchor moved randomly in a way that was *not* specified in the contracts, so the existing adjudication system could not distinguish whether the movement was caused by intentional actions or unintentional factors forces beyond anyone's control. This view of war thus requires an observational theory for measuring the assets that states were dividing, the contract terms on which they had reached consensus, the real world anchors to which those terms referred, and the emergence of novel situations that were not codified in the existing terms.

According to Lakatos, the main question is which of these two approaches offers greater explanatory power.

"The problem [of scientific inquiry] is then *shifted* from the old problem of replacing a [substantive] theory refuted by 'facts' to the new problem of how to resolve inconsistencies between closely associated theories [a substantive theory and an observational theory]. Which of the mutually inconsistent theories should be eliminated? The sophisticated falsificationist can answer that question easily: one had to try to replace first one, then the other, then possibly both, and opt for that new set-up which provides the biggest increase in corroborated content, which provides the most progressive problemshift."<sup>53</sup>

---

<sup>53</sup>Lakatos70.

We have already seen that the main observable implications of the relative power approach are refuted by the historical evidence on hegemonic wars. As we will see in a moment, there are also reasons to doubt its observational theory for measuring the conditions for peace and the causes of war. Therefore, a main goal of this study is to specify an alternative observational theory for measuring the conditions for peace and the causes of war – one based on the contractual view of international relations – and use it to test the substantive theory I call Contractual Realism.

#### **4.2.4 The connection between the substantive theory and the observational theory in Contractual Realism**

The main substantive claim of Contractual Realism is simple. Once a legal incompatibility emerges, giving two states the right to the same asset, each state faces a dilemma. If it allows the other state to receive a net gain from the dispute in the end, then the other state will have an incentive to create more legal incompatibilities intentionally in the future in order to reap more net gains. Hence each state must wage war over the current dispute for long enough that in the end, the other state has borne more costs from the war than it gains from receiving the right to *half* of the disputed asset in the final peace settlement. The main observable implication of Contractual Realism follows naturally:

The higher is the value of the asset at stake in the dispute, the longer the war is predicted to last.

This implication seems to explain why each of the hegemonic wars in Europe since 1550 lasted so long. To prove this point, however, it is necessary to obtain evidence on how highly the opposing states in each war valued the asset in dispute. This is where the observational theory enters the picture.

The observational theory specifies coding rules for identifying and measuring the evidence required to test the substantive theory. Evidence on four main causal variables is required:

1. the political, social and economic assets that states had previously divided according to the contractual method,
2. the contract terms on which statesmen had previously reached consensus prior to the

war,

3. the real world anchors to which those terms referred, and
4. the emergence of novel situations that moved these anchors in a way that was not codified in the existing terms, thereby creating overlapping rights to the same asset.

Evidence on these variables directs observational attention to the key documented facts in each historical case of hegemonic war:

1. A problem of legal incompatibility arose;
2. it caused an international dispute;
3. statesmen valued the asset at stake in the dispute very highly; and
4. they expressed the view that it would be worthwhile to wage a long and costly war over the asset if necessary.

Together these facts enable the main causal inference. Each hegemonic war was so long and costly because it arose from a problem of legal incompatibility: This problem required statesmen to wage war over the dispute for long enough that neither side would receive a net gain from receiving half of the disputed asset in the final settlement.

In Parts II through V of this study, I employ this substantive theory to derive a large number of observable implications for a wide range of facts about hegemonic war. All of the implications are derived from the assumptions that (1) statesmen valued the asset at stake in the dispute very highly and (2) the legal incompatibility problem required them to wage war over the dispute for long enough that neither side would receive a net gain from receiving half of the asset in the final settlement. In so far as all of the observable implications are supported by the empirical evidence, it proves that hegemonic wars in Europe were caused by this mechanism. The proof of this claim, in turn, establishes a foundation for proving the other substantive claims of Contractual Realism that are made and tested in the remainder of the study.

#### **4.2.5 Reasons to doubt that states divide assets according to relative power in peacetime**

The relative power view assumes that states divide assets according to relative power at all times, not only in war but in peace as well. But it is doubtful that states divide assets

according to relative power in peacetime, because this is a very costly method of dividing assets. There are two reasons why it is so costly. First, it provokes competitions in raising relative power. These competitions require states to make large investments in power and to jockey for power continuously to avoid being caught at a positional disadvantage. The root driver of these costly endeavors is simply the fact that states are dividing assets according to relative power in peacetime.

Second, each state has an incentive to bluff that its power has increased even when it has not simply to gain more assets for itself. There are only two strategies that other states can use to avoid losing assets to a bluffer state – and each strategy is very costly. On one hand, they can accept a degree of uncertainty about the bluffer state’s power – allowing it some opportunities to bluff. But then they will need to fight costly wars to deter it from such bluffing (as shown by the theory of war due to uncertainty about relative power, discussed above). On the other hand, they can attempt to gather full information on its power. Then they can call its bluffs every time it bluffs about its power, but capitulate every time it is telling the truth, thus avoiding war in both cases. However, it is very costly to gather full information on relative power. Indeed if *all* assets in the international system were divided *strictly* according to relative power in peacetime – and redivided every time relative power changed – then each state would have to invest valuable resources in power assets and jockey for power on so many dimensions of activity that the cost of gathering *full* information about relative power would be astronomical. This is why a bluffer state that tries to maximize its power on *all* available margins is quickly branded as a rogue state and excluded from key forms of international cooperation. It is simply too costly to gather full information on the rogue’s power level on all of those margins and invest in the power assets needed to jockey for power on all of those margins. And it is too confounding to optimize foreign policy toward the rogue state on so many margins of relative power simultaneously. Thus, with either strategy – accepting some uncertainty about relative power or gathering full information about relative power – states will bear a high cost to avoid losing assets to a bluffer state. And the root driver of these costs is simply the fact that states are dividing assets according to relative power in peacetime.

For all of these reasons, dividing assets by relative power is very costly. To avoid these



costs, states have a strong incentive to divide assets by another method: contracts whose terms are invariant to changes in relative power. These facts of international life have implications for scholars who are trying to explain war.

#### **4.2.6 If states divide assets by contracts rather than relative power, then theorists should attribute wars to contract disputes rather than relative power factors**

Theories of war based on relative power have one thing going for them. They are easy to conceptualize, because hypothetically speaking, there is always a relative power quotient between any two states. So a theory based on relative power can be austere and elegant – hallmarks of a good scientific theory. Yet there is an inherent tension between *theories* of war based on relative power and *empirical testing* of the theories that requires scholars to measure relative power in the real world. For just as statesmen find it difficult and costly to estimate relative power quotients accurately, so do theory testers. In order to construct a database containing a single variable labeled "relative power", a theory tester must sift through the myriad factors that constitute state power and the many dimensions on which statesmen perceive power – and select the ones that can be assumed to drive decisions to initiate war. This measurement problem points to a deeper analytical problem. Just as statesmen find it very costly to conduct foreign policy based mainly on relative power quotients, so do scholars find it confounding to discern and model the exact calculus that would drive a foreign policy based mainly on relative power quotients.

One might object that statesmen somehow manage to estimate relative power quotients well enough. So there is no reason why theory builders and testers cannot estimate these quotients too. At some point this logic starts to become circular, however. Statesmen are *assumed* to make decisions to initiate war or remain at peace on the basis of relative power. So scholarly estimates of relative power are *assumed* to provide a valid means of testing theories of war. And the underlying observational theories for measuring power are *assumed* to be good enough for testing the theories. But what if statesman do *not* make decisions to initiate war or remain at peace on the basis of relative power? What if their main foreign policy goals are the ones mentioned earlier in connection with the contractual

view of international relations?

1. To respect the current distribution of assets specified by the existing contracts as long as the real world anchors remain in the same place – regardless of shifts in relative power between states;
2. To look for situations where an anchor has moved in its favor due to forces beyond its control, and request a greater share of the relevant asset as specified in the contract;
3. To look for situations where other states are *uncertain* whether the anchors are moving in its favor due to random forces beyond its control or its intentional actions, so it can move an anchor in its favor intentionally *but bluff the anchor moved in its favor unintentionally* – and request a greater share of the asset as specified in the contract; and
4. To judge whether an anchor has moved in another state's favor due to random forces beyond its control or its intentional actions – and respond accordingly.

If this is true, then theorists who are trying to explain war can shift the focus of attention away from relative power for the same reason that statesmen do! Just as it is less costly and confounding for statesmen to base war and peace policy on contracts whose terms are invariant to shifts in relative power, it is less confounding for social scientists to build and test theories of war and peace based on such contracts. If a contractual theory receives more empirical support than the theories based on relative power – as the contractual theory presented in this study does – then the shift in analytical focus is warranted.

#### **4.3 Net assessment of the fourth generation: These theories still dominate research, yet they are refutable both logically and empirically**

In the field of international relations theory, the current view seems to be that most conflicts are not caused by shifts in relative power, but by asymmetric information or commitment problems.

”Realists argue that balancing occurs in response to changes to the balance of power. Recent informational approaches have focused primarily on informational asymmetries or commitment problems.” (Savic and Shirkey 2009)

Yet it turns out that these authors mean: asymmetric information *about relative power* and commitment problems *related to relative power*. They support their point by citing Fearon's model of conflict due to asymmetric information *about relative power*, and Powell's model of conflicts arising from commitment problems *related to relative power*.<sup>54</sup> Thus the conventional wisdom remains the same as it was in earlier generations: War is caused by relative power factors.

According to the conventional wisdom, these fourth generation theories provide *general* explanations of war. They are general explanations in two senses. First, each explanation has been characterized game theoretically in a way that identifies the necessary and sufficient conditions for war to occur. Because these conditions are very general in nature, they are assumed to explain a wide range of conflicts and wars. Second, each explanation has been embedded in regression studies that produced statistically significant coefficients. Since each study included a large number of cases, the studies together are assumed to provide proof that these factors – the incentive to bluff about relative power and commitment problems related to relative power – caused war in a wide range of cases. Given the presumed generality of these explanations, in turn, they have become the default hypotheses for researchers who study many kinds of conflict using regression analysis. For example, studies of local land disputes and violence between neighboring farmers have defaulted to the hypothesis that conflict is caused by asymmetric information about relative power.<sup>55</sup>

Thus, the current view is that war is caused by relative power factors – either bluffing about relative power or commitment problems related to relative power. If this claim were true, then it would indeed constitute a generalization about the rational causes of conflict. The previous two subsections showed, however, that this claim suffers from severe logical

---

<sup>54</sup>Fearon 1995; Powell 2006.

<sup>55</sup>Blattman et al. (2015b) attribute conflict to asymmetric information about relative power, citing Fearon's model in which this is the cause of conflict:

"Broadly, conflict is a breakdown of peaceful bargaining, and is more likely with information asymmetries and absence of commitment mechanisms. Strong, well-functioning social and political institutions at the town level reduce information asymmetries and improve coordination and enforcement (Fearon, 1995; Blattman et al., 2014)."

Similarly, Blattman et al. (2014) explain conflict between neighboring farmers over land disputes as a result of bargaining breakdowns due to asymmetric information about relative power, citing Fearon's model in which relative power is defined as one side's cost of bargaining delays relative to the other side's (Fearon 1998).

and empirical problems. On an empirical level the observable implications of these theories are refuted by the historical evidence on hegemonic wars. On a logical level these theories ignore other factors beside relative power that prevented states from settling their disputes peacefully.

## CHAPTER 17

### The Contractual Realist explanation of hegemonic wars

There were seven hegemonic wars in Europe between 1550 and 1950. Each war was caused by an international spillover problem. One state developed a new domestic activity that was causing harmful spillover effects abroad in another state. Three different types of spillover problem caused these wars:

1. Secession problem. One state was reforming its domestic communal governance in a way that inspired similar reformists in a neighboring state to create a provincial secession movement there.
2. Depression problem. One state was setting domestic prices on imported goods in a way that spread its economic depression to another state.
3. Revolution problem. One state was conducting a domestic political revolution in a way that inspired political agitation for a revolution in a neighboring state.

The conventional wisdom holds that in each case, the diplomatic crisis over the spillover problem was only a "trigger" of the war. The real "cause" of the war was a concern about relative power. Contractual realist theory says the opposite. The international spillover problem was the sole cause of the war in each case, regardless of any concerns about relative power.

If this theory of hegemonic war is correct, then it implies the need to distinguish between three questions.

**Question 1.** *Why did each state decide to go to war over the dispute, rather than settling the dispute through peaceful negotiation?*

**Question 2.** *Once each state decided to go to war, what were its ultimate war aims?*

**Question 3.** *What means and strategies did each state need to employ to achieve its ultimate war aims?*

Contractual realist theory agrees with the conventional wisdom in answering the third ques-

tion. Each state's strategy for achieving its ultimate war aims was to maximize its power and minimize the opponent's power. In fact this is the classic definition of a war: a battle for power. The conventional wisdom assumes, however, that this is therefore the answer to the first two questions as well. Thus the conventional wisdom assumes that each state decided to go to war either to take advantage of an increase in its own power or to minimize the threat posed by an increase in the opponent's power.<sup>1</sup> Proceeding logically from this assumption, the conventional wisdom then makes two assumptions regarding ultimate war aims. It assumes that in each war the ultimate war aim of one state – the "challenger" or "revisionist" state – was to maximize its power in the world; and the ultimate war aim of the other states was to minimize that state's power in the world – simply to maximize their own security. Contractual realist theory holds, by contrast, that the answers to the first two questions have nothing to do with relative power. Consider each question in turn.

## **1. Why did each state decide to go to war rather than settle the dispute peacefully? (Question 1)**

The international spillover problem was the sole cause of the war in each case, regardless of any shifts in relative power, uncertainty about relative power or commitment problems related to power. The spillover problem caused an international dispute because each state claimed that the other state was responsible for bearing the cost of solving the spillover problem, yet each state responded that it was not responsible for bearing the cost of solving it. The state harmed by the spillover effects claimed that it had the right to be free of their harmful effects, so the state producing them was obligated to stop its domestic activity that was producing them. In turn, the state producing them claimed that it had a right to continue that activity free of foreign interference, so the state harmed by the spillovers was obligated to accept them. Each state could cite an international law to back its claim, however. So the spillover problem provoked a deeper problem of "legal incompatibility":

---

<sup>1</sup>See the review of the conventional theories in Section 2 of this chapter. The first theory outlined there assumes that one state went to war because its power had increased by enough that it expected to reap net gains from a war. The second theory assumes that one state went to war to prevent another state from taking advantage of an impending increase in its power. The third theory assumes that one state went to war to prevent another state from bluffing that its power had increased when it had not.

One law accorded one state the right that was in dispute but another law accorded another state the right that was in dispute.

When each state stood firm in its claim to be in the right, this created mutual uncertainty about their intentions to violate each other's rights. Neither state could tell whether the legal incompatibility had arisen unintentionally on the part of the other state, or it had created the legal incompatibility intentionally but was only *bluffing* that it had arisen unintentionally. For all the harmed state knew, the emitting state was emitting the spillovers intentionally and only bluffing that it was doing it unintentionally to privilege its right in the dispute (e.g. the right to continue its domestic activity). Yet for all the emitting state knew, the harmed state understood the spillovers were an unintended side effect of its legitimate domestic activity, and was only bluffing ignorance of this fact to privilege its right in the dispute (e.g. its right to be free of the spillover effects).

Each of the spillover problems that caused these wars was new – it had never arisen before in European history. So there was no adjudication system to validate the evidence proving that the first state was producing the spillovers unintentionally and insure that the second state acknowledged that evidence. Hence each state was left with only one way to defend its rights from encroachment and bluffing by the other state. It had to go to war over the dispute and impose enough war costs on the other state to deter it from continuing to encroach and bluff in the way that it appeared to be doing. The harmed state had to impose such high war costs on the emitting state that it would be better off stopping its domestic activity that was producing the spillover effects than fighting on. In turn the emitting state had to impose such high war costs on the harmed state that it would be better off acknowledging the spillover problem had arisen unintentionally than fighting on.

At the moment the dispute became irreconcilable and each side decided to go to war over it rather than settle it peacefully, each state was thrown from the world of law – in which enforcement is based on justice bargains – into the world of power – where enforcement is based on power. From this perspective, the decision to go to war was simply a matter of choosing between two alternative modes of contract enforcement: a peaceful mode based on justice bargains and a conflictual mode based on power. The function of the war in each state's calculus was to secure its income from the threat posed by the other state due to the

international spillover problem. Once each state decided to go to war, therefore, its *ultimate war aim was to defend its income*, not to maximize its power or minimize the opponent's power.

In each of the seven crises, the dispute over the spillover problem became irreconcilable and erupted in war due to the legal incompatibility problem alone, regardless of relative power factors. This fact is demonstrated by the demands and counter-demands issued in the diplomatic crisis that led up to the outbreak of each war. They show that it was the legal incompatibility problem that drove the dispute over the states' respective rights and eventually rendered the dispute irreconcilable. In some cases the spillover problem shifted relative power in favor of the state *emitting* the spillover effects. In other cases the spillover problem shifted relative power in favor of the state *harmed* by the spillover effects. In still other cases the spillover problem did not shift relative power at all. Yet in all seven cases the opposing states went to war over the spillover problem regardless of such relative power factors.

To give clarity to these ideas, consider some examples.

### **1.1 The 30 Years War**

In the 1500s and early 1600s, the Protestants in southern Germany were practicing their religion in a way that inspired some Catholics across the border in Austria to convert to Protestantism, thereby threatening the Catholic led government there with popular disaffection and rebellion. The German Protestants denied they were responsible for the problem on the grounds that the conversions in Austria were unintentional on their part. The key question for the Austrian government, therefore, was whether the German Protestants were *knowingly* practicing their religion in a way that was intended to inspire conversions in Austria. If they were doing it knowingly, and the Austrian government permitted them to continue doing it, this would give them an incentive to look for *other* ways to practice their religion so as to inspire *more* conversions in Austria.

To deter this kind of opportunism, the Austrian government would have to demand that the German Protestants change the way they were practicing their religion, and threaten to go to war over the issue if the demand were not met. If the Germans rejected the



demand, Austria would have to fight over the issue for long enough to deter the Germans from continuing to practice their religion in the way they were. That is, Austria would have to impose such high war costs on the Germans that they would be better off changing their religious practice than continuing to fight on in defense of it.

This was a new type of spillover problem. There was no adjudication system to distinguish whether the German Protestants were causing it knowingly or unknowingly. For all Austria knew, therefore, the Germans were doing it knowingly and only bluffing that they were doing it unknowingly. Austria could not permit them to continue doing this (for the reason explained above). Hence the Austrian government demanded that the German Protestants change the way they were practicing their religion, and threatened to go to war over the issue if the demand was not met.

In turn, the key question for the German Protestants was whether the Austrian government truly could not tell they were causing the problem unknowingly, or it was only bluffing it could not tell when it really could. Since there was no adjudication system, there was no way for the German Protestants to know what evidence the Austrian government had about the Germans' intention or lack of intention in causing the problem. So there was no way to know whether the Austrian government was ignoring any evidence it had. For all the German Protestants knew, therefore,

1. The Austrian government had evidence that the German Protestants were simply mind-ing their own religious business and not intentionally conducting it in a way so as to inspire political agitation in Austria; but
2. the Austrian government was ignoring the evidence and bluffing that for all it knew, the German Protestants were intentionally inspiring conversions in Austria; so that
3. the Austrian government could demand moderation in the German Reformation for its own gain, even though the demand was not necessary to deter opportunism by the Germans.

So if the German Protestants were to acquiesce to the demand, it would give Austria an incentive to do the same thing on other issues in order to make more demands opportunistically (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew the opposite). So the German Protestants rejected the Austrian demand and went to war over the issue.

Once war erupted, Austria had to fight over the issue for long enough to deter the German

Protestants from continuing to practice their religion in a way that was intended to inspire religious conversions in Austria. In turn, the German Protestants had to fight over the issue for long enough to deter the Austrian government from claiming that, for all it knew, the Germans were inspiring conversions in Austria intentionally.

## **1.2 The French Revolutionary Wars**

In the early 1790s, the revolutionaries in France were restructuring their domestic government in a way that encouraged activists in Austria to advocate for a similar restructuring there, thus threatening the noble led government in Austria with being overthrown. The French revolutionaries denied they were responsible for the problem on the grounds that the political agitation in Austria was unintentional on their part. The key question for the Austrian government, therefore, was whether the French revolutionaries were knowingly conducting their domestic revolution in a way that was intended to inspire political agitation in Austria. If they were doing it knowingly, and the Austrian government permitted them to continue doing it, this would give them an incentive to look for more ways to conduct their domestic revolution so as to inspire more political agitation in Austria. Again this was a new type of spillover problem. There was no adjudication system to distinguish whether the French revolutionaries were causing it knowingly or unknowingly. For all Austria knew, therefore, the French were doing it knowingly and only bluffing that they were doing it unknowingly. Austria could not permit them to continue doing that. Hence the Austrian government demanded that the French revolutionaries moderate their domestic revolution – and threatened to go to war over the issue if the demand was not met.

In turn, the key question for the French revolutionaries was whether the Austrian government truly could not tell they were causing the problem unknowingly, or it was only bluffing it could not tell when it really could tell. Since there was no adjudication system, there was no way for the French to know what evidence the Austrian government had about French intention or lack of intention in causing the problem. So there was no way to know whether the Austrian government was ignoring any evidence it had. For all the French knew, therefore,

1. The Austrian government had evidence that the French revolutionaries were simply minding their own domestic political business, and not intentionally conducting it in a way to inspire political agitation in Austria; but
2. the Austrian government was ignoring the evidence and bluffing that, for all it knew, the French revolutionaries were inspiring the political agitation in Austria intentionally and knowingly; so that
3. the Austrian government could demand moderation in the French Revolution for its own gain, even though the demand was not necessary to deter opportunism by the French.

So if the French were to acquiesce to the demand, it would give Austria an incentive to do the same thing on other issues in order to make more demands opportunistically (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew the opposite). So the French rejected the Austrian demand and went to war over the issue.

Once war erupted, Austria had to fight over the issue for long enough to deter the French from continuing to conduct their domestic revolution in a way that would inspire political agitation in Austria. In turn, the French had to fight over the issue for long enough to deter the Austrian government from claiming that, for all it knew, the French were inspiring political agitation in Austria intentionally.

### **1.3 World War II**

In the 1930s, France, England and the United States raised the prices that their domestic consumers were paying for imported goods, thereby lowering the export earnings of German firms.<sup>2</sup> France, England and the United States denied they were responsible for the income loss in Germany on the grounds that it was unintentional on their part. The key question for the German government, therefore, was whether the Western powers were knowingly setting their economic policies in a way that was intended to increase their incomes at Germany's expense. If they were doing it knowingly, and the German government permitted them to continue doing it, this would give them an incentive to look for other ways to set their economic policies so as to increase their incomes at Germany's expense again. Once again this was a new type of spillover problem. There was no adjudication system to distinguish

---

<sup>2</sup>The financial markets in France, England and the United States also increased the interest rates that investors were paid for providing capital, thereby raising the capital costs of German firms and lowering the wages of German workers. The same kind of analysis applies to this part of the spillover problem.

whether the Western powers were causing it knowingly or unknowingly. For all Germany knew, therefore, the Western powers were doing it knowingly and only bluffing that they were doing it unknowingly. Germany could not permit them to continue doing this. Hence the German government demanded that the Western powers moderate their economic policies, or failing that, accept Germany's right to conquer other states and establish direct control over their economic policies.

In turn, the key question for the Western powers was whether the German government truly could not tell they were causing the spillover problem unknowingly, or it was only bluffing it could not tell when it really could. Since there was no adjudication system, there was no way for the Western powers to know what evidence the German government had about Western intention or lack of intention in causing the problem. So there was no way to know whether the German government was ignoring any evidence it had. For all the Western powers knew, therefore,

1. The German government had evidence that the Western powers were simply minding their own economic business, and not intentionally setting their economic policies so as to impose income losses on Germany; but
2. the German government was ignoring the evidence and bluffing that, for all it knew, the Western powers were setting their economic policies at Germany's expense intentionally and knowingly; so that
3. the German government could demand moderation in their economic policies for its own gain – or acceptance of its right to conquest – even though these demands were not necessary to deter opportunism by the Western powers.

So if the Western powers were to acquiesce to these demands, it would give Germany an incentive to repeat the same opportunistic behavior on other issues (e.g. bluff that for all it knew it was being harmed intentionally, when it really knew the opposite). So the Western powers rejected the demands and went to war over the issue.

Once war erupted, Germany had to fight over the issue for long enough to conquer other states and establish direct control over their economic policies – and to deter the Western powers from adopting economic policies designed to increase their incomes at Germany's expense again. In turn, the Western powers had to continue fighting for long enough to deter Germany from engaging in such conquests – and to deter Germany from claiming that, for all it knew, they were setting their economic policies to gain income at its expense

intentionally.

## 2. What were each state's ultimate war aims? (Question 2)

Once each state decided to go to war, its ultimate war aim was to defend its income rights from the other state. The ultimate war aim of the state harmed by the international spillover effects was to defend its income from the spillover effects – by forcing the state emitting them to change the way it performed its domestic activity to stop emitting them. In turn, the ultimate war aim of the state producing the international spillover effects was to defend its right to perform its domestic activity in the way it wished, because that activity generated income. These income gains would be lost if the opponent achieved its aim of forcing changes in the way the domestic activity was performed. So the ultimate war aim was to stop the opponent from forcing such changes, and thereby secure the right to perform the domestic activity without foreign interference. Consider the same three examples:

### 1. The Thirty Years War.

State harmed by the spillovers. The income and wealth of Austria's rulers and administrators was threatened by the prospect that their subjects would convert to Protestantism and then throw them out of office. So the Austrian government's ultimate war aim was to induce the German Protestants to change the way they practiced their religion, to stop inspiring Catholics in Austria to convert to Protestantism.

State producing the spillovers. The German Protestants gained income and wealth from their reformed religion, because it reduced the capacity of their churches to exact taxes and tithes for religious infractions, and it also increased the incomes of rulers and administrators who shared the reformed religion of their subjects.

### 2. The French Revolutionary Wars.

State harmed by the spillovers. The income and wealth of Austria's nobles was threatened by the prospect that their subjects would stage a political revolution and expropriate their land and privileges. So the Austrian government's ultimate war aim was to induce the French revolutionaries to moderate their domestic revolution, to stop inspiring political agitation in Austria.

State producing the spillovers. The lower and middle classes in France gained income and wealth by staging a political revolution that would stop the nobles from exploiting them.

### 3. World War II.

State harmed by the spillovers. The Western powers (thought they) could gain income and wealth by setting higher tariffs on imported goods.

State producing the spillovers. The income and wealth income of Germany's firms and

workers was threatened by the prospect that the Western powers would continue to set their economic policies intentionally to gain income at Germany's expense. So the German government's ultimate war aim was to induce the Western powers to change the way they set their economic policies or, failing that, to conquer other states whose economic policies it could control directly.

Contractual realist theory implies that each state needed to pursue its ultimate war aim regardless of relative power considerations, simply to secure its income. And each state needed to *continue* to pursue its ultimate war aim regardless of *shifts* in relative power, either before or during the war, again simply to secure its income. Even if the state's relative power fell to the point where it needed to drop out of the war temporarily, it would continue to harbor its ultimate war aim of securing its income while it figured out a way to increase its relative power again and reenter the war. Ultimately the function of the war for each state was to enforce a contract delineating its rights to income. Power was simply a means to that end. Shifts in power only induced changes in the means and strategies employed to achieve that end, not a change in the end itself.

From this perspective, the means and strategies that each state used to achieve its ultimate war aim become understandable in a completely new way. The first step toward this new understanding is to define each state's switch from an enforcement strategy based on law and justice bargains to an enforcement strategy based on power and conflict.

### **3. From the world of law to the world of power**

In each crisis the decision to go to war marked a fundamental switch in each state's understanding of its international environment – a switch from seeing this environment as a world of law to seeing it as a world power. Once this switch occurred, each state stopped playing the game of international life by the rules of law and started playing the game by the rules of power. To see the significance of the switch consider each set of rules in turn.

#### **3.1 The world of law**

In the world of law, the rights to valuable assets are distributed among states by means of anchor term contracts – agreements whose terms do not shift when there are shifts in the

relative power of states. These include international laws, rules and norms. For example, in the international disputes that caused hegemonic wars in Europe, the key rights in dispute were: (1) the right of people in one state to be free of harms caused by the activities of people in another state, and (2) the right of people in that state to be free of foreign interference in their domestic activities.

In the world of law, there is value to sticking to contract terms *as originally written* – simply because that is what the terms say. The value is that states save the costs of competitions in raising relative power. States pay a price for sticking to contract terms as originally written, however. They cannot maximize their income perfectly by changing *any* contract term *any* time their relative power changes. Sometimes they have to stick to the contract terms as originally written, even though those terms are *not* the income maximizing terms given relative power at the moment. In most cases this price is worth paying to obtain the greater value of avoiding wasteful competitions in raising relative power.<sup>3</sup>

Thus states maximize their incomes in two ways. First, each state calculates its income maximizing action based on its current relative power. And it continuously adjusts its actions to maximize its income on this margin. Second, each state gains additional value by committing *not* to maximize its income on this margin, and instead to stick to contract terms as originally written – regardless of shifts in relative power – simply because that is what the terms say. The additional value comes from avoiding competitions in raising relative power. Often (though not always), the value of maximizing income in this second way exceeds the value of maximizing income in the first way *by enough* to give states an incentive to do it the second way: by committing to anchor term contracts.

I call this the world of law. In this world there is a transcendent value to sticking to contract terms as originally written – simply because that is what the terms say. Sticking to contract terms for this reason constitutes a "logic of appropriateness". The right action to take in a particular situation is simply the action that is "appropriate" given what the contract terms say, not the action that maximizes income on the margin of relative power. At the same time, this logic of appropriateness is grounded in a "logic of consequences" at

---

<sup>3</sup>The key exception is the setting of prices in perfectly competitive markets with no asset-specific investments. This is the exception that proves the rule. In markets with asset-specific investments, fixed price contracts are the norm. Cite review articles on fixed price contracts.

the deeper level: Sticking to contract terms avoids wasteful competitions in raising relative power and thereby saves resources that can be invested in productive economic activity.

When a dispute arises over the distribution of a valuable asset, the dispute is settled by evidence on whether it arose intentionally or unintentionally – as long as there is an adjudication system to assess the evidence and insure that both parties acknowledge it. When this condition is met states live in a world where international disputes are settled by evidence. Absent such an adjudication system, however, the opposing states in a dispute have no other option but to resort to power-based methods to enforce their contractual rights. Thus, when a new kind of dispute arises and the existing adjudication system fails, states are thrown from the world of law and evidence-based enforcement into the world of power and power-based enforcement.

### **3.2 The world of power**

In the world of power, the rights to valuable assets are distributed among states according to their relative power rather than according to law and evidence. Hence each state has an incentive to increase its relative power to gain a greater share of the asset in contention. In the event of a dispute over the distribution of an asset, each state must increase its relative power to prevent the other state from shifting the distribution of power its way.

Rule number one in the world of power is that if a state's military power falls by too much, then it will be better off dropping out of a war than continuing to fight. Or if the war has not begun yet, it will be better off not starting the war in the first place. Strategy number one, therefore, is for each state to maximize the number of men deployed in military combat, simply to avoid being forced to drop out of the war. This strategy is subject to only one constraint: the state must also maximize the amount of goods and services produced by the civilian economy in order to keep the civilians and the soldiers supplied with income (food, materiel, wages). Each state's competitive strategy is constrained tightly by this income maximization problem. The less income the soldiers receive, the less powerful the army will be in combat (due to low morale or mutinies). The less income the civilian workers receive, the less goods and services they will produce to supply the army. If a state's strategy strays too far from the income maximizing one, then it will lose the income it needs to fund the



army and maintain civilian morale sufficiently to stay in the war. Thus each state's number one strategy is to maximize the number of men deployed in military combat – subject to the constraint that it be the income maximizing number.

From this perspective the world of power is a world of pure security competition in which each state is an income maximizing "firm" that produces two outputs: (1) manpower deployed in military combat to impose costs on the enemy and (2) civilian economic goods and services to supply the soldiers and the civilian workers. Each state is constrained by the forces of international security competition to choose the income-maximizing mix of these two outputs. It is also constrained by these forces to choose the income-maximizing strategy on any other trade-offs that it faces. For example, it has two options for maximizing the military manpower on its side in a war: domestic mobilization and international alliances. In choosing which option to rely on more, it is constrained to make the income-maximizing choice.

#### **4. What means and strategies did each state need to employ to achieve its ultimate war aims? (Question 3)**

Each state's overall strategy was to maximize its military power relative to the opponent. Each state's military power was determined first and foremost by the number of soldiers it deployed in military combat. It could not deploy its entire population in combat, of course, because some men and women were needed in the civilian economy to produce the food and materiel required by the soldiers. Therefore, each state's strategy was to maximize the number of soldiers it deployed in military combat subject to the constraint that this be the *income*-maximizing number. Each state needed to deploy the income maximizing number of soldiers because if it were to deploy any other number, it would have less income to spend on food, materiel and wages for the soldiers and less income to distribute to the civilians at home to maintain their morale in the war. Then that state would risk losing the war, either by being unable to fund the army sufficiently to avoid desertions or by being unable to maintain civilian morale at home. Each state could avoid this outcome by deploying the income maximizing number of soldiers. That is to say, each state could avoid the mistake

of losing a war that it might have won by doing a better job of maintaining its population's loyalty.<sup>4</sup> To analyze military power in this way, it is necessary to distinguish between the main elements of a state's military power:

1. Productivity power = the productivity of the state's labor force in the military sector relative to the productivity of its labor force in the civilian economic sector.
2. Manpower = the number of soldiers the state deploys in military combat.
3. Alliance power = the number of soldiers deployed by all states within a military alliance.

#### 4.1 The impact of shifts in productivity power on state strategy

According to contractual realist theory, productivity power is the most important element because it constrains the state's choice of the other two elements. In particular, each state's income maximizing number of soldiers is determined by the productivity of its labor force in the military sector relative to the productivity of its labor force in the civilian economic sector. If the military productivity was *high* relative to the civilian productivity, the income maximizing number of soldiers was *high*. Conversely, if the military productivity was *low* relative to the civilian productivity, the income maximizing number of soldiers was *low*. In each of the hegemonic wars since 1550, these relative productivity levels were determined primarily by the nature of the spillover problem. Three different types of spillover problem caused these wars, and each type had a *different* effect on the relative productivity of military labor and civilian labor in each state.

1. **Secession problem.** One state was reforming its domestic communal governance in a way that inspired similar reformists in a neighboring state to create a provincial secession movement.

The spillover problem increased military productivity in the state emitting the spillover effects and decreased military productivity in the state harmed by the spillover effects.

2. **Depression problem.** One state was setting domestic prices on imported goods in a way that spread its economic depression to another state.

---

<sup>4</sup>The point is not to assume that each state deployed the *exact* number of soldiers that maximized its national income. Rather, the point is to assume that if any state *strayed too far* from that number in its deployment decisions, it would risk losing the war unnecessarily. So it had to get as close to that number as humanly possible.

The spillover problem decreased military productivity in the state emitting the spillover effects and increased military productivity in the state harmed by the spillover effects.

3. **Revolution problem.** One state was conducting a domestic political revolution in a way that inspired political agitation in a neighboring state.

The spillover problem itself did not affect military productivity relative to civilian productivity in either state. But the state harmed by the spillovers experienced an increase in its military productivity when it gained a military option to overturn the emitting state's domestic revolution by force.

These shifts in productivity power shaped each state's strategies for achieving its ultimate war aims decisively. The impact of productivity power on state strategies and international outcomes differed significantly from one type of spillover problem to another.

## 4.2 Explaining differences in grand strategy across states and wars as a consequence of shifts in productivity power

### 4.2.1 Wars caused by secession crises

The state *harmed* by the spillover effects was *declining* in military productivity. So once it decided to go to war over the dispute rather than settle it peacefully, it had to go to war quickly before its military productivity declined even further. Even if its military productivity had declined by so much (by the time it decided to go to war rather than settle) that it could not win the war fighting alone, it still had to go to war anyway, simply to defend its income in the dispute over the spillover problem. Because of its decline in military productivity, it was severely constrained in the number of its own men that it could deploy in combat. So the only way it could deploy enough men to stay in the war was to contract allies. In three of the four wars caused by secession crises, the state had no natural allies in its region. So it had to conquer allies by force and establish a coercive hegemony over them. Yet its hegemony was simply a means to fund the strategy of maximizing the number of soldiers deployed on its side in the war, in a situation where its own military productivity had decreased by so much as to limit severely the number of its own men it could deploy from its own population. By contrast, the state *emitting* the spillover effects was *rising* in military productivity as a result of the spillover problem. So it did not need to rush into

war, and it did not need to contract allies to be able to stay in the war.

These patterns of war initiation and grand strategy were manifested clearly in The War of Dutch Independence (1568-1609), The 30 Years War (1618-48), The War of the Spanish Succession (1701-1713), and World War I (1914-18).

#### **4.2.2 Wars caused by depression crises**

The state *harmed* by the spillover effects was *increasing* in military productivity. So it did not need to rush into war. On the contrary it had an incentive to obfuscate its intention eventually to go to war, since its military productivity was increasing over time (and it would need time to produce the military hardware necessary to take maximum advantage of it). Its increase in military productivity also meant that it did not need to contract allies to be able to stay in the war. By contrast, the states *emitting* the spillover effects were *declining* in military productivity as a result of the spillover problem. So they needed to go to war quickly once any threat became clear to them. The state harmed by the spillovers was obfuscating its intention to go to war, however, so the threat remained unclear to them. Since their military productivity had decreased, moreover, each of them experienced a decrease in its income maximizing number of soldiers. So none of them could stay in the war fighting alone. Each could stay in the war only if the others entered and stayed in the war as well. Since they were natural allies, none of them needed to establish a *coercive* hegemony over the others. Yet they did need to establish a unified military command over a combined army simply to ensure that each of them stayed in the war.

These patterns of war initiation and grand strategy were manifested clearly in World War II in both the Asian theater (1931-45) and the European theater (1939-45).

#### **4.2.3 Wars caused by revolution crises**

The spillover problem itself did not affect military productivity in either state. So neither state needed to rush into war. Once the spillover problem became serious enough, though, the state harmed by the spillovers gained an incentive to overturn the emitting state's domestic revolution by force. So the harmed state experienced an increase in its military productivity,

while the emitting state experienced a decrease in relative terms. At that point it gained an incentive to rush into war. And it could not stay in the war fighting alone. So it had to contract allies to be able to stay in the war. In one case it had no natural allies because it was a radical revolutionary state. So it had to conquer allies by force and establish a coercive hegemony over them. In the other case it was a conservative revolutionary state so it had a number of natural allies. It did not need to conquer any of them or establish a coercive hegemony. Yet together, the allies did need to establish a unified military command over a combined army to ensure that each of them stayed in the war.

These patterns of war initiation and grand strategy were manifested clearly in The Nine Years War (1688-97) and the French Revolutionary Wars (1792-1815).

#### **4.3 Implications of the role of productivity power for the explanation of hegemonic wars**

The central role of productivity power carries several implications. First, each state decided to go to war over the dispute rather than settle it peacefully regardless of whether productivity power was shifting in its favor or against it. This fact provides support for the hypothesis that each state's decision to go to war was not driven by relative power concerns (shifts in relative power, uncertainty about relative power or commitment problems related to power) , but rather by the threat of income losses due to the international spillover problem.

Second, the key turning point in each diplomatic crisis was the moment at which the dispute over the spillover problem became irreconcilable. At this moment each state decided to go to war over the dispute rather than settle it peacefully. Relative power was irrelevant up to and including this moment. After this momentous decision, however, relative power drove everything – with one exception, the choice of ultimate war aims. Each state's ultimate war aim was to defend its income from the threat posed by the spillover problem, regardless of relative power considerations.

Third, once each state decided to go to war with this aim, its choice of strategies to achieve the aim was driven by relative power more than any other factor.

Fourth, these strategic choices were driven by shifts in productivity power caused by

international spillover problems, rather than by any other kind of shift in power. In particular, shifts in productivity power drove its strategies for when to enter military hostilities and whether to seek hegemony.

Fifth, shifts in its productivity power were the main driving factor behind shifts in the other two dimensions of its power – military manpower and alliance power. In this sense, productivity power was the most fundamental determinant of state military power.

Shifts in productivity power also explain a wide range of other strategies and outcomes in hegemonic wars (see Parts II, III and IV of this study). This suggests that the answers to questions one and two given above are correct. Specifically, the decision to go to war rather than settle the dispute peacefully was driven by the spillover problem rather than by relative power considerations. And each state's ultimate war aim was to defend its income and wealth from the threat posed by the spillover problem, regardless of relative power considerations.

## **5. War duration and settlement**

The evidence also shows that the opposing states prolonged each war for long enough to satisfy the deterrence conditions mentioned above – regardless of how relative power may have shifted one way or another during the war. If one side fell in relative power to the point where it was better off dropping out of the war than continuing to fight on, it dropped out only temporarily. As soon as it could boost its power to the critical threshold again, it reentered the war – again and again if necessary – and kept fighting until those deterrence conditions were satisfied. Once that point was reached, the opposing states could settle for peace.

In the three cases where the peacemakers were able to build an effective international adjudication system to govern the spillover problem that had originally caused the war, the ensuing peace was stable and long. The key terms in each of these settlements constituted an anchor term contract and a justice bargain. The emitting state conceded an obligation not to perform its domestic activity in a way that would emit international spillover effects *intentionally*, and agreed to an international adjudication system that would judge disputes over its compliance with this obligation. In turn, the harmed state agreed to an international adjudication system that would assess evidence that the emitting state was producing spillover

effects *unintentionally*, and conceded an obligation to acknowledge such evidence. In the other four cases, however, the peacemakers were not able to build an effective international adjudication system of this kind, so the ensuing peace was unstable and short.

## CHAPTER 18

### **International security competition and the motive of income loss avoidance**

The purpose of this theory is to explain what caused the seven hegemonic wars in Europe between 1550 and 1950. In each of these cases the main antagonists experienced a dispute and tried to resolve it peacefully through diplomatic bargaining. Hence the main fact that needs to be explained is why the two states failed to resolve the dispute peacefully and chose to go to war instead.

The conventional wisdom is that these "bargaining breakdowns" were caused either by domestic political dysfunctions or by relative power factors, such as shifts in relative power, uncertainty about relative power or commitment problems related to relative power (see the literature review in the previous chapter).

I provide an alternative theory that attributes the wars to a different cause. In each case one state developed a new domestic activity that increased its income but lowered another state's income through an international spillover effect. There were multiple international laws relating to the first state's new domestic activity. Yet the laws were logically incompatible with each other. Hence it was not clear whether the first state had the right to continue its new domestic activity (and continue to reap the income gains from it) or the second state had the right to be free of the income losses caused by that activity. This legal incompatibility caused a dispute over income between the two states. Each state expected that if it capitulated in the legal dispute it would lose income, and the other state's income gain would give it an incentive to create further disputes and impose further income losses on it in the future. Each state calculated that it needed to wage war on the other state to impose enough costs on it to deter it from such opportunism. In short, each state's decision to wage war rather than settle the dispute peacefully was driven by the need to deter the other state from imposing more income losses on it, not by relative power factors.



While it is true that the strategies of the main antagonists were influenced by power factors, their decisions to wage war rather than settle the dispute peacefully were not driven by power factors. At the same time, however, relative power factors influenced both states' strategies in other respects. A main aim of my theory is therefore to explain the role of relative power in hegemonic wars while simultaneously showing that these wars were not *caused* by relative power factors.

To accomplish these explanatory goals, I develop a theory that answers six questions about international security competition in general:

1. What are the most fundamental factors that drive state strategies and international outcomes?
2. What are the most important elements of state power?
3. What are the causal relationships between these elements in the short run and the long run?
4. Why do states form peace agreements?
5. Why do states initiate wars?
6. What caused the seven hegemonic wars in Europe between 1550 and 1950?

## 1. Basic assumptions of the theory

In international security competition, each state's strategy is constrained by six fundamental factors. While other factors may also constrain a state's strategy depending on the circumstance, these six factors *always* constrain a state's strategy. The statesman either finds that these factors are impossible to ignore or discovers that by ignoring them his state pays such a high cost that he must stop ignoring them. Either way history shows that state strategies always heed these six constraints. The theory therefore assumes that in equilibrium, each state's strategy is determined by these six constraints.<sup>1</sup> At the outset, let us assume for simplicity that the international system contains only two states. This assumption will be

---

<sup>1</sup>In technical terms, each assumption is essential in the sense that if it were removed from the theory, then one could not prove the existence and uniqueness of the equilibrium on which the theory is based.

relaxed later.<sup>2</sup>

1. **The territory constraint.** The most basic fact of international life is that the international system contains a limited amount of territory. The more territory one state holds, the less territory is available for the other state to hold. This is the "territory constraint" on state strategy.
2. **The relative power constraint.** If the two states divide the territory by war, the resulting distribution of territory is determined by relative army size. This is the "relative power constraint" on state strategy.
3. **The population constraint.** Each state's population is limited in the short run. The more workers a state devotes to military and government functions to capture and secure territory, the less workers it has left in the civilian economy to perform productive economic activity. This is the "population constraint" on state strategy.
4. **The productivity constraint.** Each state requires at least two inputs to produce goods and services in the domestic economy: territory and civilian labor. Without either input, the economy will not produce any goods or services at all and the population's income will be zero. Hence the state must devote at least *some* of its population to each function: securing territory by military force and performing productive activity in the civilian economy. The marginal productivity of labor in each function is assumed to be declining. This means that the more people are added to the military sector to capture and secure territory, the smaller is the increment to national income from adding another person to the military sector. And the more people are added to the civilian sector to produce goods and services in the domestic economy, the smaller is the increment to national income from adding another person to the civilian sector. This is the "productivity constraint" on state strategy.
5. **The domestic political constraint.** If the state devotes too many people to the

---

<sup>2</sup>Assumptions 1-4 and 6 are derived from economic models of conflict (Michelle R. Garfinkel 1990; Powell 1993; Hirshleifer 1989; Hirshleifer 1991; Hirshleifer 1995; Stergios Skaperdas 1992; Stergios Skaperdas 1996; Findlay 1996). I have modified and extended these assumptions as necessary to achieve the explanatory goals of this study. Assumption 5 is new in this study as far as I am aware. The motivation it provides for assumption 6 is therefore new in this study. I provide a formal characterization of my theory in the technical appendix to this Part of the study. My model also follows in the tradition of an earlier generation of economists who pioneered the study of coercion, conflict and settlement using basic price theory (Hirschman 1945; Lane 1942; Lane 1950; Lane 1958; Knorr 1947; Knorr 1957; Demsetz 1967).

military sector, then there will not be enough civilian workers in the domestic economy to produce goods, services and income for the population. The domestic economy will be starved of civilian workers and national income will be lower than it could have been if the state had put more workers into the domestic economy. As a result the soldiers will receive a lower wage rate than they could have received if the state had put more workers into the domestic economy and fewer into the army. If the soldiers' wage rate falls too low as a result of this misallocation of labor, they may even mutiny and leave the state's territory undefended. In this case the state's leadership would be at risk of being removed from office and replaced for failing to defend the land. Conversely, if the state allocates too many people to the domestic economy and not enough people to the military sector, then the domestic economy will be starved of territory as an input to civilian production and national income will be lower than it could have been with more people in the military sector. Then the civilian workers will receive a lower wage rate than they could have received if the state had put more people into the military sector to capture territory (and fewer into the domestic economy). If the civilian workers' wage rate falls too low due to this misallocation of labor, they may stage a labor strike that would leave the economy unproductive. This would reduce the tax base needed to fund the army and, again, leave the state's territory undefended. Again the state's leadership would be at risk of being removed from office and replaced for failing to defend the land. This is the "domestic political constraint" on state strategy.

6. **The income maximization constraint.** When the productivity constraint and the domestic political constraint operate simultaneously, it generates an income maximization constraint. Each state's leadership is forced to allocate its population between the military sector and the civilian sector in a way that maximizes its national income – or at least in a way that does not stray too far from the income maximizing allocation – or else it will be removed from office for failing to defend the land. If the *current* allocation of labor is too far from the *optimal* allocation of labor (that maximizes national income), then the leadership is constrained to shift the allocation in the direction of the optimal allocation. It may not arrive at the optimal allocation, but it is constrained to move in that direction. In this sense the leadership prefers an allocation with a higher national income to an allocation with a lower national income, and so one can say

that the leadership is constrained to maximize national income. This is the "income maximization constraint" on each state's army size and the other state policies that depend on its army size.

In effect, each state's leadership is constrained to avoid policy errors that would lose too much national income (compared to the maximum level of income that is feasible given its current technology and labor productivity). The greater is the income loss from a policy error (compared to the income-maximizing policy), the stronger is the leadership's incentive to avoid that error. To avoid such policy errors, the leadership must choose an army size that is somewhere in the middle between "too few" soldiers and "too many" soldiers. In other words, the leadership is forced to adopt an army size that moves the state toward the maximum feasible national income. Although it may never achieve the maximum, it is forced in this direction by international security competition. In turn, all state security policies that depend on army size are constrained to operate within these upper and lower limits on the army size. This is the "system constraint" that international security competition imposes on each state's security policies.<sup>3</sup>

**International security competition imposes a "system constraint" on each state's leadership that forces it to choose the army size that maximizes national income, or at the least, not to stray too far from this army size.**

### **1.1 Choosing the army size that maximizes national income: a graphical view**

By translating these assumptions into graphical form, one can see each state's incentive to maximize national income, the costs of straying from that objective, and the security policies produced by that objective. The assumptions are displayed in figure 30. The  $x$ -axis represents the total population of state 1 (e.g. the population constraint). Any point on the  $x$ -axis represents an allocation of the state's total population between the military sector and the civilian sector. The military labor pool is counted from the left origin up to the allocation point. The civilian labor pool is counted from the right origin up to the allocation

---

<sup>3</sup>For general discussion of the "system constraint" that arises whenever two or more actors interact on the basis of individual utility maximization, see Langlois (1986b).

point. The  $y$ -axis represents the marginal *increment* to national income from adding another worker to each sector (e.g. the marginal product of labor in each sector). Each labor pool is subject to diminishing marginal productivity. The marginal product of military labor declines as more men are added to the military (e.g. moving the allocation point from left to right). In other words, the more soldiers are added to the military sector, the lower is the increment to national income from adding another soldier to the military sector. The marginal product of civilian labor declines as more men are added to the civilian economy (e.g. moving the allocation point from right to left). In other words, the more workers are added to the civilian sector, the lower is the increment to national income from adding another worker to the civilian sector. This is the productivity constraint.

Under these assumptions, each state faces the trade-off mentioned above. The more people it devotes to the army to secure territory, the less people it will have left to produce goods, services and income in the civilian economy. But the less people it devotes to the army to secure territory, the less territory it will have as an input to civilian economic production. The optimal allocation of the population between the two sectors is the point where the marginal products are equal in the two sectors. This point occurs at the intersection of the two marginal product lines. The optimal division of labor is labeled  $A_1^*$ . I call this point the *optimal army size*. At this division of labor, the state is maximizing its national income. If the state were at any other allocation point, then it could increase national income by reallocating people to this point, because the marginal gain from moving toward this point would exceed the marginal loss.<sup>4</sup>

**A state's income-maximizing army size is the size at which its marginal productivity of labor in the military sector is equal to its marginal productivity of labor in the civilian sector.**

This conclusion is evident in figure 30. It displays state 1's total national income as a function of the allocation of its population between the military and civilian sectors. Again the  $x$ -axis represents the total population of state 1, and any point on the  $x$ -axis represents an allocation of the total population between the military and civilian sectors. The  $y$ -axis now

---

<sup>4</sup>This conclusion assumes that state 2 is already at its income-maximizing division of labor,  $A_2^*$ . A full analysis using best response functions is provided in the appendix.

represents total national income (rather than the marginal increment to national income). The horizontal tangent line indicates the maximum feasible income level given the state's technology and labor productivity:  $q_{1w}^*$ . This maximum income level is achieved at the optimal army size,  $A_1^*$ .

International security competition forces each state's leadership to adopt the army size that maximizes national income – or at least to move toward this army size. Therefore any change in the marginal productivity of labor in the military or civilian sector will force the leadership to change the state's army size accordingly.

**Any shifts in a state's labor productivity will induce a change in its optimal army size, its actual army size and all policies that depend on its army size.**

## **1.2 The main elements of state power and the causal relationships between them**

This theory of international security competition identifies the main elements of state power and explains the causal relationships between them. The main elements are:

1. population size
2. labor productivity
3. military force size
4. amount of territory
5. economic output (GDP)

In long-run equilibrium, the most fundamental foundation of state power is population size. The larger a state's population is, the larger its army is, other things being equal. The next important foundation of power is the productivity of the state's labor in military and domestic governance functions. The higher a state's military productivity is, the larger its army is (other things equal). In turn, the size of a state's army determines how much territory it holds and how high its national income is in the long run (other things equal).

**The higher is the productivity of a state's military labor relative to the productivity of its civilian labor, the larger is its army size and its**

share of the international system's territory (other things equal).

The lower is the productivity of a state's military labor relative to the productivity of its civilian labor, the smaller is its army size and its share of the international system's territory (other things equal).

## **2. Incentives to form peace agreements**

So far the theory assumes that the two states are at war all the time. They divide the total amount of territory in the international system according to relative army size through continuous war. Each state maintains its optimal army size at all times. Its soldiers are never available to work in the civilian economy. Consequently each state's income is lower than it would be if it could transfer some soldiers back into the civilian economy. The lost income can be considered "*war costs*": the costs of devoting human resources and capital to war rather than productive activity in the civilian economy. Leaders have a strong incentive to avoid war costs, because war costs, like any costs that lower national income below its maximum feasible level, render leaders vulnerable to being removed from office by domestic constituents (as outlined above). To avoid war costs, states form peace agreements that allow them to transfer soldiers back into the civilian economy. To form a peace agreement, the two states agree to stop fighting and maintain the same distribution of territory that was produced by the war. As long as each state expects the peace to be sustainable, it can transfer some or all of its soldiers back into the civilian economy and thereby receive a "peace dividend": an increase in its national income.

### **2.1 Two types of peace agreement, two methods of enforcement, and two levels of enforcement costs**

The primary incentive to form peace agreements is to avoid war costs. This is not the only incentive, however, because war costs are not the only costs that states bear as a result of international security competition. They bear four other kinds of costs as well. If it were only the war costs that states wanted to avoid, it would be sufficient to form *any* type of peace agreement, even one whose terms shift with changes in states' relative power. To avoid the

other four kinds of costs, however, states must form a *special* type of peace agreement, one whose terms do not shift with changes in states' relative power. I call this type of agreement an *anchor term contract*.<sup>5</sup>

Thus there are two types of peace agreement: the type whose terms shift with changes in the relative power of states, and the type whose terms do not shift with changes in the relative power of states. These two types of peace agreement differ significantly in their methods of enforcement and costs of enforcement. Agreements whose terms shift with relative power are enforced by power-based methods such as *deterrent war*; and the enforcement costs are comparatively high as a result. Agreements whose terms do not shift with relative power are enforced by anchor-based methods such as *justice bargains* (to be explained below); and the enforcement costs are comparatively low. States therefore have a strong incentive to form the second type of peace agreement rather than the first type.

The two types of peace agreement are associated with two distinct causes of war. With agreements whose terms shift with relative power, the main cause of war is relative power factors, such as shifts in relative power, uncertainty about relative power, and commitment problems related to relative power. With agreements whose terms do not shift with relative power, the main cause of war is incompatibilities among the clauses of the agreement that cannot be resolved by existing institutions for dispute resolution. We saw in the previous chapter that the hegemonic wars in Europe between 1550 and 1950 were not caused by relative power factors. The theories based on those hypotheses are refuted by the evidence. The alternative hypothesis is that these wars were caused by incompatibilities among the clauses of existing international laws binding the major states of Europe. The remainder of this section explains these points in detail.

---

<sup>5</sup>Rowe (1989b). I first cited and applied Rowe's ideas in my earlier studies of hegemonic wars and peace settlements [Koppel (1997, 1998a, 1998b, 1999d, 2000b, 2002b)]. See Section 6.1 and 6.2 of the previous chapter for the relevant quotations from these earlier studies. I then extended and modified Rowe's ideas in my study of Britain's imperial wars that attributed them to breakdowns in common knowledge about non-power factors [Koppel (2000c, 2006a, 2006b)]. See Section 6.1 of the previous chapter for the relevant quotations from that study. I then combined these two strands of research to characterize a formal model of costly conflict due to breakdowns in common knowledge about non-power factors (Koppel 2007; Koppel 2008). Finally, I extended these developments of Rowe's ideas to explain domestic conflicts and settlements [Koppel (2009b, 2010c, 2011)]. I present all of this previous work in the present study.



### 2.1.1 Framing the question of how peace is enforced and war is avoided

Consider the most basic question of enforcement:

What makes a peace agreement sustainable?

What prevents one state from seizing territory or another valuable asset from the other state unilaterally?

What prevents one state from *demanding* territory or another valuable asset and threatening to seize it unilaterally if the demand is not met?

This type of opportunism is called a "hold up" in economic theory. So the question of enforcement becomes: What prevents this type of "holdup"?<sup>6</sup>

The most basic answer is: The targeted state must impose more costs on the opportunistic state than it would gain from committing such a hold up. As long as each state maintains a reputation for imposing this punishment on the other state as necessary, then neither state will expect to reap net gains from committing holdups and no holdups will occur in equilibrium. Then the peace agreement will be sustainable. The key question is: How can the targeted state impose costs on the opportunistic state? As we will see, there are two main ways to impose costs on the opportunist depending on the type of peace agreement that the states formed.

### 2.1.2 Peace agreements whose terms shift with relative power

Suppose the two states form a peace agreement whose terms shift in response to changes in the relative power of the two states. This type of agreement is enforced by the method of *deterrent war*. If one state seizes a valuable asset from the other state unilaterally, the second state can transfer some workers from the civilian economy into the military sector and wage a deterrent war that imposes more costs on the first state than it gained by seizing the asset. Then the first state will be deterred from seizing assets unilaterally in the future and the

---

<sup>6</sup>For the economic literature on holdups, see Williamson (1975), Klein, Crawford, and Alchian (1978), Williamson (1979), Williamson (1985), and Klein (1996) and Malcolmson. The political science literature contains a few studies that employ transaction cost economics to analyze holdups (Weber 2000; Cooley and Spruyt 2009; Carnegie 2014 and Dixit). These studies adopt different interpretations of "transaction costs" and "holdups" than the my study adopts. For my interpretations of these concepts, see the theory chapters in Parts II, VI and VII of this study.

peace will be sustainable. By extension, suppose the first state demands an asset from the second state and threatens to seize the asset unilaterally if the demand is not met. As long as the second state threatens to wage a deterrent war in response and its threat is credible (because of a history of waging deterrent war in the past), the peace will be sustainable.<sup>7</sup>

This conclusion depends on a key assumption: It is costless to transform civilian workers into soldiers and it is costless to transform physical capital into military hardware. If "resource transformation" is costless in this sense, then each state will always have an incentive to wage deterrent war no matter how small is the asset threatened by the hold up, that is, no matter how low its value. In this case each state will be free to transfer *all* of its soldiers and military hardware into the civilian economy in peacetime, and thereby receive the maximum feasible peace dividend (e.g. the maximum increase in national income from forming a peace agreement). For even with all of its resources in the civilian economy, it will always have an incentive to transform the resources into military form to wage a deterrent war as necessary *because resource transformation is costless*.

By contrast, if it is costly to transform civilian workers and capital into military forces, then each state will *only* have an incentive to wage a deterrent war under this condition: The value of the asset threatened by the holdup is *higher* than the transformation cost of rearming to defend it. But if the value is less than this threshold, then the state does not have an incentive to wage a deterrent war to defend it. For such "small" assets, threats to wage deterrent war are not credible and opportunistic states will have an incentive to commit hold ups. Moreover, such holdups will succeed in reaping gains for the opportunistic state *without provoking a war*: The harmed state will not have an incentive to fight back because its transformation costs of doing so would be greater than the value of the asset to be defended.

Such holdups are costly to the state that was held up, of course, because it loses the asset seized by the opportunistic state. These hold up costs cannot be treated as war costs, however, because there is no war in these cases. Rather, the hold up costs are due to a deeper problem: The two states are dividing a valuable asset according to relative power *in peacetime*. This point is worth explaining. Whenever one state takes advantage of the other

---

<sup>7</sup>This assumes the holdup is not too large. See below.

state's *transformation costs of defense* to seize "small" assets as the opportunity arises – here and there – the first state is taking advantage of the second state's *lack of power* to defend its assets. In other words the first state is taking advantage of its relative power position. When both states take advantage of their relative power positions in this way, control over territory and other valuable assets shifts back and forth between the two states according to their relative power positions. Wars never occur in this kind of world because, as explained, such holdups do not provoke wars. In effect, the two states are dividing valuable assets between themselves according to relative power *in peacetime*. Thus the costs that each state bears from losing assets to the other state in holdups are not war costs. They are a cost of dividing assets according to relative power in peacetime.

To avoid such hold up costs, each state has an incentive to transform some of its resources from civilian form into military form in advance, before any holdups occur, so that in the event of a hold up its *remaining* transformation costs will be lower. In other words, each state has an incentive to "frontload" its transformation costs to the time before any hold ups have been attempted. This restores the credibility of its deterrent threats and thereby enables it to deter the other state from committing holdups in the first place. How much transformation costs is it worth frontloading in order to deter holdups? Each state will incur transformation costs up to the point that the marginal benefit from avoiding holdups is just equal to the marginal cost of transforming resources into military form in advance to deter holdups.<sup>8</sup> Why do states choose to pay transformation costs in advance during peacetime to avoid holdup costs? The choice cannot be attributed to war since no wars occur either before or after states adopt this frontloading strategy. Rather, the choice to incur transformation costs in peacetime is due, again, to the fact that states are dividing valuable assets according to relative power in peacetime.

The same conclusion pertains to two additional costs that states incur in international security competition. First, when states maintain armed forces during peace time to deter holdups, they incur *stockpile costs*. These are simply the costs of keeping human and physical

---

<sup>8</sup>Usually, it is not optimal to transform so many resources that all holdups are deterred. So at least some holdups occur in equilibrium. In the technical appendix to this Part of the study, I provide a formal model of this trade-off, including necessary and sufficient conditions for the equilibrium and comparative statics results.

capital in military forms that cannot be employed in the civilian economy. The need to incur these costs cannot be attributed to war, since the stockpiling strategy deters holdups and renders deterrent war unnecessary, so no wars occur in this equilibrium. Rather, the need to incur stockpiling costs is due simply to the fact that states are dividing a valuable asset according to relative power in peacetime. Second, once states stockpile military assets, each state has an incentive to maneuver its military assets into positions where it might commit another kind of hold up: seizing non-military assets from the other state in places where its valuation of those assets is less than its cost of maneuvering military assets into position to protect those assets. To deter such positional seizures, states constantly jockey for position by moving their military assets from one place to another to fill emergent gaps in relative military power. The costs of all this maneuvering and jockeying for position can be termed *jockeying costs*. Again, the need to incur such costs cannot be attributed to war, since no wars occur in this equilibrium. The need is simply due to the fact that states are dividing a valuable asset according to relative power in peacetime.

In sum, when states divide territory and other valuable assets according to relative power in peacetime as well as in wartime, they incur five distinct types of cost. Only the first type – war costs – is avoided by forming peace agreements whose terms shift with changes in relative power. The second type – hold up costs – can be avoided to some extent by maintaining military assets in peacetime. But the remaining three types – transformation costs, stockpile costs, and jockeying costs – continue to be incurred in peacetime as long as the terms of the peace agreements shift with relative military power.

These conclusions are remarkable for several reasons. First, even if states avoid wars by forming peace agreements and deter hold ups by maintaining armed forces in peacetime, they must still incur the other three types of cost, simply because they live in a world where resource transformation is costly and they are dividing the world's assets according to relative army size in peacetime. Second, analysts must distinguish between costly conflicts (such as wars) and costly competitions in raising relative power. Even if states avoid costly conflicts, they will still engage in costly competitions in raising relative power as long as the production of military forces is costly and they continue to divide valuable assets according

to relative power in peacetime.

**When states form a peace agreement whose terms change with relative power, they avoid the costs of conflict, but they do not avoid the costs of competitions in raising relative power.**

**When states form a peace agreement whose terms do *not* change with relative power, they avoid the costs of conflict *and* the costs of competitions in raising relative power.**

### **2.1.3 Peace agreements whose terms do not shift with relative power**

To avoid the three types of cost that arise from costly competitions in raising relative power, states form peace agreements whose terms do not shift with relative power. The terms are based on anchors in the real world that do not shift when the relative power of the states changes. Since the contract terms do not shift when relative power shifts, neither state can move the terms in its favor intentionally by increasing its relative power. So neither state has anything to gain by increasing its relative power intentionally, and therefore neither state has an incentive to invest resources in raising its relative power.<sup>9</sup> Wasteful competitions in raising relative power are avoided and resources are liberated for productive investment in the economy.<sup>10</sup> I call such an agreement a *anchor term contract*.

Because the world is always changing, the anchors in the world move periodically too. An anchor term contract specifies how the distribution of rights between the two states will change when the anchors move due to random forces beyond either state's control. As long

---

<sup>9</sup>For the economic literature on fixed price contracts, see Goldberg 1976; Joskow 1986, 1988; Klein and Murphy 1988; Malcolmson 1990s. As far as I am aware, Rowe (1989b) was the first economist or social scientist to claim that arms races and other types of wasteful competitions in raising relative power are avoided mainly by reaching agreements whose terms do not shift with the signatories' relative power. [I welcome the provision of citations to any earlier studies making this claim.]

<sup>10</sup>For the economic literature on arms races, see Michael D. Intriligator (1975), M. D. Intriligator and Brito (1976), Leidy and R. W. Staiger (1985), Anderton (1989), and Michelle R. Garfinkel (1990). For analysis of the question of whether arms races can cause war, see M. D. Intriligator and Brito (1984) and M. D. Intriligator and Brito (1986). For the assumption that conflict itself is a wasteful competition in raising relative power, see Rowe (1989b). In the present study I assume that war itself is a wasteful competition in raising relative power. I am not aware of any other study that makes this assumption. [I welcome the provision of citations to any other studies making this claim.]

as each state knows that the anchor in the world moved due to forces beyond the other state's control, it can allow the distribution of contractual rights to shift in the other state's favor when this is what the contract specifies.<sup>11</sup> The state that is favored by the anchor's movement thus receives a *variable benefit*. The benefit is variable because the state receives the benefit only periodically – at those times when the anchor moves in its favor due to forces beyond its control.

Because of the variable benefit feature, anchor term contracts can be enforced without recourse to power-based methods. Each state can be deterred from reneging on the anchor term contract simply by the threat that it will be denied its variable benefit if it reneges. However, if this enforcement method is to work, the contracting parties must strike a special kind of bargain that I call a *justice bargain*. The first state commits that whenever it is found guilty of moving the anchor in its favor intentionally for its own gain, it will admit guilt and concede defeat voluntarily – without resistance. In return, the second state commits that whenever the anchor moves in the first state's favor due to forces beyond its control, the second state will grant the first state its variable benefit voluntarily – without resistance. As long as the first state gains more from receiving these variable benefits periodically than it loses from its concessions in the cases where it moved the anchor intentionally, it is better off conceding defeat voluntarily when found guilty of such intentional opportunism than resisting. In turn, as long as the second state loses less from granting away the variable benefits periodically than it would lose from having to switch to a power-based method of enforcement, it is better off adhering to its half of the justice bargain as well.

This kind of bargain can be enforced without any state having to raise its relative power to make its enforcement threats credible. If either state reneges on its half of the bargain, the other state can retaliate simply by reneging on its half of the bargain in return. Credible threats to retaliate in this way are sufficient to enforce the bargain. So neither state needs to maintain or deploy power assets for enforcement purposes. Hence neither state needs to invest in raising its relative power to render its retaliation threat credible. In particular, the first state can be deterred from moving the anchor in its favor intentionally by the threat that if it is found guilty of doing so, it will be denied its variable benefit when the anchor

---

<sup>11</sup>Rowe 1989b.

moves in its favor due to factors beyond its control (e.g. the threat that the second state will renege on its half of the justice bargain in retaliation). This form of retaliation is cost-free for the second state. In fact the second state actually *saves* resources by retaliating in this way. It does not have to grant the first state the variable benefit any more!

The variable benefit feature thus enables the natural variability of the world to be harnessed for enforcement purposes. However, certain conditions must be met if natural variability is to be leveraged in this way. First, the terms of the agreements between states must be tied to anchors in the real world that remain stationary at least some of the time. Second, states must be in consensus on the rules for determining whether violations of these terms were intentional or unintentional. As long as there is consensus on such rules, agreements between states can be enforced by threats to deny variable benefits. When a new kind of dispute arises, however, it may provoke dissensus on the rules for determining intentionality – and dissensus on the verdicts that result in particular cases. If so, then this method of enforcement will not work. The state accused of violating an agreement will deny having done anything wrong – and it will regard the denial of its variable benefit as an infraction itself. That state will be impelled to retaliate against the accusing state, simply to deter further “untrue” accusations and “wrongful” denials of the variable benefit. In turn, the accusing state will regard this response as an insistence on the right to continue violating the agreement with impunity. At this point the only enforcement method that can work is the first method discussed earlier – the power-based method: Each state wages deterrent war on the other to impose more costs on it than it would gain from continuing its opportunism.

## **2.2 Two types of peace agreement and two causes of war**

When states form peace agreements whose terms shift in response to shifts in relative power, the main cause of war between them is relative power factors (e.g. shifts in relative power, uncertainty about relative power, and commitment problems related to relative power). The logic behind this conclusion is straightforward. Each state is at risk of losing valuable assets if the other state’s power increases. Hence each state has an incentive to collect information on the other state’s power, so that it can call the other state’s bluffs about its power. If it lacks such information, each state has an incentive to punish the other

state through a deterrent war whenever it claims that its power has increased, simply to deter it from bluffing about its power. Thus the main cause of wars in this kind of world is uncertainty about relative power in a world of shifting power. The logic is equally straightforward for wars caused by commitment problems related to relative power <sup>12</sup> While these problems may explain some wars, however, they cannot explain the hegemonic wars that occurred in Europe between 1550 and 1950 (as explained in the previous chapter).

When states form peace agreements whose terms do *not* shift in response to shifts in relative power (anchor term contracts), the main cause of war is disputes caused by incompatibilities among the contract terms that cannot be resolved by the existing rules for dispute resolution. The logic here is equally straightforward. Each state is at risk of losing valuable assets if the other state moves the anchors of an agreement in its favor intentionally. Hence each state has an incentive to collect information on what is causing the movements of the anchors, so that it can know whether they were intentional or unintentional on the other state's part. With such information, each state can call the other state's bluff if it moves an anchor intentionally but claims the anchor moved unintentionally. If a new kind of dispute emerges, however, states may find it impossible to distinguish intentional from unintentional movements of the anchors. Then each state has an incentive to punish the other state through a deterrent war whenever it claims that an anchor moved due to random forces beyond its control, simply to deter it from bluffing that such movements were unintentional when they were intentional on its part. Thus, the main cause of war in this kind of world is the emergence of new kinds of disputes over contractual rights to valuable assets.<sup>13</sup>

In the remainder of this chapter and the next chapter, I derive the observable implications

---

<sup>12</sup>See Powell (2006). Summarize logic [tbd].

<sup>13</sup>I first analyzed legal incompatibilities as a cause of war in my earliest studies of this dissertation topic between 1997 and 2006. I provide quotations from these studies in Section 6 of the previous chapter. A legal incompatibility is a situation where one contract term gives one actor the legal right to an asset, but another contract term gives another actor the legal right to the *same* asset. The incompatibility causes a breakdown in common knowledge about the legal meaning of the contract terms. This is not a breakdown in common knowledge about relative power. It is a breakdown in common knowledge about a non-power factor: the legal meaning of contract terms. This breakdown can occur even when none of the four rationalist causes of war in the existing literature is present (e.g. even when there are no indivisibilities, no commitment problems relating to relative power, no bluffing about power, and no lapse in common knowledge about relative power).



of this hypothesis. In the following chapters I test these implications on the historical evidence from the hegemonic wars in Europe between 1550 and 1950.

### **2.2.1 The theory's most basic implication**

When a war erupts due to this cause, the opposing states will switch from dividing their assets by the contractual method to dividing their assets according to relative power. This unleashes a competition in raising relative power that imposes all of the costs described above – not only the loss in economic output from having to keep workers in the military sector during the war, but also the transformation costs of creating an optimally sized army, the jockeying costs of maneuvering the army to deter the opponent from committing positional holdups, and the hold up costs of losing assets to the opponent in locations where it is too expensive to defend them. These costs are borne not only when each side is firing its weapons but also during lulls in the fighting when *neither* side is firing its weapons. For even at these times the two sides are still in a *state* of war. What makes it a state of war is *not* that the two sides are firing their weapons at every point in time. They are not. What makes it a state of war is that the two sides are experiencing an irreconcilable dispute about the division of valuable assets between them, so their only recourse is to divide the assets according to relative power. As long as the state of war lasts, each state must create military power assets, maintain them, and maneuver them for a relative power advantage at every moment – even when neither side is firing its weapons. This is what makes the state of war so costly. As long as each side must continue to create, maintain and maneuver power assets, it must bear all of the costs it previously avoided by dividing the world's assets according to the contractual method: the transformation costs, stockpile costs, jockeying costs and hold up costs. In short, what makes the state of war so costly is that dividing assets according to relative power is much more costly than dividing assets according to the contractual method.

By implication, the historical evidence should show that when the main states of Europe became embroiled in each hegemonic war, they would switch from a peacetime international system of laws and norms that created stable asset distributions to a wartime system based on relative power that created unstable distributions. Each state's costs of securing its interests in this international security competition would rise exponentially. And each state would have to transform its social, economic and political systems in fundamental ways to

meet such costs.

## 2.2.2 Why hegemonic war is a paradigmatic type of war

From this perspective it is worth reconsidering the question of why hegemonic war has long been considered a paradigmatic type of war. The conventional wisdom claims:

1. The paradigmatic feature is that each hegemonic war was caused by a large shift in relative power, uncertainty about a large shift in relative power, or a commitment problem related to a large shift in relative power.
2. Each war was so long and costly because the power shift that caused it was so large.
3. Because power is the most important factor in war and the defining feature of international relations, it is likely that other wars were caused by relative power factors too – even if those wars were not as long or costly.
4. So research agendas on war in general should be based on the assumption that they were caused by relative power factors.

These conventional wisdoms, rooted in the body of theory claiming that hegemonic wars were caused by relative power factors, have driven research for decades. Yet we saw in the previous chapter that this body of theory fails to explain actual hegemonic wars. Its failure calls into question the conventional wisdom about what makes hegemonic war paradigmatic and suggests that a different view should be taken.

According to my theory, the paradigmatic feature of hegemonic war is that each war was caused by a dispute over contractual rights to valuable assets. Once the dispute became irreconcilable, the contractual method of dividing assets between states ceased to function. So states had to switch to the main alternative method: Dividing assets according to relative power. While the war was not *caused* by relative power factors, it was *conducted* under the pressures of the relative power constraint on state strategies (discussed above). This is why at least one state in each war dropped out of the fighting for a while after its relative power fell too low. Because the *state* of war continued as long as the income dispute remained unresolved, however, even these states wanted to resume fighting and eventually reentered the war after bolstering their power. Thus each war was so long and costly because the amount of income at stake in each dispute was so large. Because income is essential for survival and progress, be it in wartime or peacetime, agreements to divide income generating assets are the defining feature of international relations. And disputes over income are the most fundamental cause of war, not only hegemonic wars but likely other wars as well – even

wars that were not as long or costly.

In the next section I develop an analytical framework based on the concept of "the size of the income stakes in a dispute over contractual rights to income". The purpose of this framework is to enable the derivation of observable implications to be tested on the historical evidence on hegemonic wars.

### 3. The most fundamental cause of war: Disputes over income

Suppose that two states have reached an agreement to divide an income-generating asset between them. What happens if two clauses of the agreement turn out later to be incompatible with each other – in the sense that one clause gives one state the right to the asset but the other clause gives the other state the right to the same asset? I call this problem an *incompatibility between two contractual rights*, or in the case of laws, a *legal incompatibility*. Since both states value the asset because of the income it generates, each state has an incentive to interpret the agreement or the laws in its own favor. The incompatible interpretations give rise to a dispute which is, ultimately, a dispute over income. Consider some examples from the history of hegemonic wars in Europe.

**Example 1.** In 1618 political leaders in the Kingdom of Bohemia claimed the right to unseat government office holders who were violating the Kingdom's constitution and the right to nullify the laws they had passed. When the Bohemian leaders exercised these rights it inspired provincial leaders in the neighboring Kingdom of Austria to resist the central government in Austria in similar ways. In response the King of Austria (who was also the King of Bohemia) claimed the right to reseat government officials of his choosing in Bohemia and restore the laws passed by his previous appointees though they violated the Bohemian Constitution.

**Example 2.** Between 1789 and 1791, the French revolutionaries abolished the feudal system and claimed the right to impose legal restraints on the monarchy or abolish it entirely to prevent a resurgence of the nobility and restoration of the feudal system. Their moves inspired the peasants and the middle class in Austria to oppose feudalism violently and to mobilize for legal restraints on the Austrian monarchy. When the Austrian Emperor and his counselors perceive these developments as a threat to disempower or abolish the monarchy, they claimed the right to intervene in the French Revolution to protect the French monarchy from being disempowered or abolished.

### 3.1 Wars caused by disputes over income: a graphical framework

Suppose that state 1 takes an action that increases its income at state 2's expense. Once state 1 starts taking the action, its income is higher by  $\Delta$  in each period of time, say a month, and state 2's income is lower by  $\Delta$  in each period of time. These income changes are shown in figure 30. In each panel of the figure, the x-axis represents time, measured in periods of, say, a month. The y-axis represents one state's income in each period of time. The upper panel represents state 1's income and the lower panel represents state 2's income.

Before time zero, each state's income is assumed to be  $q_{1p}^* = q_{2p}^* = 50$ . State 1 starts taking the action in question at time zero. So its income is  $q_{1p}^* + \Delta = 55$  for the first time period, and state 2's income is  $q_{2p}^* - \Delta = 45$  for the first time period. Now suppose that state 2 claims that state 1 does not have the right to take the action and demands that state 1 stop taking the action. Suppose, however, that state 1 rejects the demand and insists that it has the right to continue taking the action. If these claims and counterclaims are due to a *legal incompatibility*, then they will cause a dispute over whether state 1 can continue to take the action. In this dispute, state 1 claims that it has the right to receive an income of  $q_{1p}^* + \Delta = 55$  per period. State 2 claims that it has the right to receive an income of  $q_{2p}^* = 50$  per period. Thus it is a dispute over income.

The stakes in this dispute are  $\Delta$  per period for each state. Either way the dispute is settled, one of the two states will lose income *compared to the income level that it claims it has the right to receive*. If state 1 backs down in the dispute and stops taking the action,

then its income will return to the status quo level before the dispute arose,  $q_{1p}^* = 50$ . In this case state 1 would lose the return of  $\Delta$  per period that it was gaining from the action. This outcome is represented in the upper panel of figure 30 by the thin solid line. Alternatively, if state 2 backs down in the dispute and allows state 1 to persist in taking the action, then state 2's income per period will remain  $q_{2p}^* - \Delta = 45$ . So state 2 would continue to lose  $\Delta$  per period. This outcome is represented in the lower panel of figure 30 by the thin dotted line. However the dispute is settled, one of the two states will lose income compared to the income it claims it has the right to receive. Thus it is ultimately a dispute over income.

If the two states cannot reach a peaceful compromise settlement of the dispute, then each state has only two options. The first option is to back down in the dispute and accept the income loss just described. If it chooses this option, then the other state will reap a net gain from the dispute, and that will give it an incentive to initiate further disputes of a similar kind in the future. The second option is to go to war over the dispute in order to force the other state to back down. If it chooses this option, it can deter the other state from initiating further disputes of a similar kind in the future that would impose further income losses on it. To achieve this deterrent effect, a state must initiate a war and sustain it for long enough to impose such high war costs on the opponent that, in retrospect, it would have been better off backing down in the current dispute than standing firm and fighting over the dispute. By fighting for this long, the first state insures that the second state is deterred from initiating similar disputes in the future. After fighting for this long, the first state can agree to a peace settlement in which it receives part of what it sought in the current dispute and concedes the other part to the second state.

If both states choose the second option of going to war over the dispute, then the outcome is a deterrent war. This type of war is illustrated in figure 30. Again, in each panel the x-axis represents time and the y-axis represents one state's income in each period of time. The upper panel represents state 1's income and the lower panel represents state 2's income. The thick solid lines show the progress of the dispute and the deterrent war over time. At time zero, state 1 starts taking the action that increases its income but reduces state 2's income. At time 1, state 2 initiates a deterrent war and sustains it until time  $T_\Delta = 18$ . Then state 2 agrees to a peace settlement that gives state 1 the right to take the action in question

a fraction of the time, but the obligation to refrain from the action the remaining fraction of the time. The same process of action and deterrent counteraction also occurs in reverse with a one period delay. At time 1, when state 2 initiates the deterrent war, it also takes an action intended to reverse state 1's initial action that increased its income. At time 2, state 1 joins the military hostilities and sustains the war until time  $T_\Delta = 18$ . Then state 1 agrees to a peace settlement on the same terms: State 1 has the right to take the action in question a fraction of the time, but the obligation to refrain from the action the remaining fraction of the time.

Each state sustains the war for time  $T_\Delta$ , that is, for long enough to increase the other state's war costs above its single period gain from standing firm in the current dispute (plus its gain from the compromise settlement after the war). Then each state suffers a net loss from having stood firm in the dispute and it is deterred from initiating further disputes of a similar kind in the future.<sup>14</sup>

**The minimum duration of the war that is necessary to achieve this mutual deterrent effect depends on the size of the stakes in the dispute.**

The higher the stakes are, the longer the war must be sustained to maintain mutual deterrence. This conclusion is demonstrated in figure 31. Again the x-axis represents time and the y-axis represents one state's income in each period of time. The size of the stakes in the dispute is represented by the parameter  $\Delta$ . In panel A, the size of the stakes in the dispute is relatively small ( $\Delta = 5$ ), so each state only needs to sustain the war for a short period of time to raise the other state's war costs above its single period gain from standing firm in the dispute. In panel B, the size of the stakes in the dispute is larger ( $\Delta = 10$ ), so each state needs to sustain the war for longer to raise the other state's war costs above its single period gain from standing firm in the dispute. If the size of the stakes in the dispute becomes large enough, then eventually the stakes cross a threshold where even a war that lasts "forever" is not sufficient to raise the other state's war costs above its single period gain from standing

---

<sup>14</sup>There is a unique value of the settlement fraction parameter, such that  $T_\Delta^* = T_1^* = T_2^*$ . A conflict of this duration minimizes the sum of the two states' conflict costs, subject to the condition that the conflict reestablishes bilateral deterrence of future activity of a dispute-initiating kind. There is no conflict of any shorter duration that reestablishes deterrence bilaterally. See appendix for proofs of the existence and uniqueness of this equilibrium.

firm in the dispute. So if one state fails to initiate a deterrent war before the size of the stakes rises above this critical threshold, then it loses the capacity to deter the other state from initiating similar disputes in the future by waging war over the current dispute.<sup>15</sup>

### 3.2 The upper threshold for the size of the stakes in the dispute that is consistent with mutual deterrence

The higher the stakes in the dispute are, the longer the war needs to last to restore mutual deterrence. If the stakes in the dispute are too high, however, then even a war that lasts forever would not enable either state to impose more war costs on the other state than it gained from standing firm in the dispute. In this case, *neither* state can deter the other state from initiating further disputes of a similar kind in the future through a deterrent war over the current dispute. This problem is illustrated in panel D of figure 31. Here,  $\Delta$  is so large that even if each state sustains its deterrent war over the current dispute forever, this would not ever raise the other state's war costs above its single period gain from having stood firm in the dispute. The implication is obvious.

**Each state must initiate a deterrent war before  $\Delta$  reaches this critical threshold, or else the state cannot maintain deterrence and it will be vulnerable to further income losses from the other state initiating similar disputes in the future.**

**By extension, each state must also take the actions preparatory to initiating a war before  $\Delta$  reaches the critical threshold, such as assessing the income threat from the other state's action, mobilizing the army, contracting allies, etc.**

Actually, there is a different threshold for each state. What determines these upper limits on the size of the stakes in the dispute that are consistent with mutual deterrence? As we will see in a moment, the magnitude of each state's threshold depends on the difference between the *other* state's income from standing firm and going to war over the dispute and its income from backing down in the dispute. Its war income may vary from one type of

---

<sup>15</sup>In this figure the discount factor is assumed to be 0.99, so the number of time periods from time period "zero" to time period "infinity" is approximately 450 periods.

dispute to another. And such variations generate observable implications about the speed with which each state takes preparatory actions and initiates war, depending on the type of dispute it is. The following subsections are devoted to explaining the logic behind these conclusions.

[Suppose that at time 0 state 1 begins to take an action that increases its income by  $\Delta$  at state 2's expense. Assume that the magnitude of the income change,  $\Delta$ , starts at zero and increases over time as state 1 takes the action more intensively. And assume that it causes a dispute that results in a deterrent war of the kind of just described. Mutual deterrence is restorable through a war of finite duration as long as  $\Delta$  is not too large, as we saw. But if  $\Delta$  is too large then mutual deterrence cannot be restored even by a war that lasts forever.]

### 3.2.1 State 2's threshold for the maximum size of the stakes in the dispute

The factors that determine state 2's critical threshold are shown in figure 30. The x-axis shows the size of the stakes in the dispute ( $\Delta$ ). The y-axis shows state 1's single period incomes from different scenarios. The thin solid line shows its income before it starts taking the action in question ( $q_{1p}^*$ ). The thick solid line shows its income after it starts taking the action that gives it an income gain of  $\Delta$  in each period ( $q_{1p}^* + \Delta$ ). The thick dashed line shows its income from standing firm and going to war over the dispute ( $q_{1p}^* + \Delta - w_1$ ). Here, the parameter  $w_1$  represents state 1's *war cost* per period. The thin solid line shows its income from backing down in the dispute and returning to the *status quo ante* (again  $q_{1p}^*$ ).

State 2's critical threshold is at  $\bar{\Delta}_2$ . Once the size of the stakes in the dispute rises above this threshold, state 1's income from standing firm and going to war over the dispute is greater than its income from backing down in the dispute:

$$q_{1p}^* + \Delta - w_1 > q_{1p}^* \quad (18.1)$$

So state 1 is better off standing firm and fighting forever than backing down. In this case state 2 would not be able to restore deterrence even if it were to fight forever in the current dispute. Hence, if it were to wage a deterrent war over the current dispute anyway, its war costs would be incurred for naught – it would be wasting its resources for nothing. So state



2 is better off accepting the loss of  $\Delta$  per period than initiating a deterrent war over the dispute.<sup>16</sup> The implication is that state 2 must initiate a deterrent war *before*  $\Delta$  reaches this critical threshold. This conclusion is worth examining graphically.

As long as the stakes in the dispute are below the critical threshold  $\bar{\Delta}_2$ , state 2 can maintain deterrence by waging a deterrent war of finite duration (as explained above). By fighting for finite time  $T_\Delta$ , state 2 will impose more war costs on state 1 than it gained from initiating the dispute. State 1 will incur a net loss from the current dispute and thus be deterred from initiating more disputes of a similar kind in the future. In this case state 2's total income over all periods – its "infinite horizon" income – will be higher by standing firm and waging a deterrent war over the current dispute than by backing down and accepting the income loss of  $\Delta$  per period in the current dispute. This income comparison is shown in figure 30. The thick dotted line shows state 2's infinite horizon income from standing firm and waging a deterrent war over the current dispute. The thick solid line shows state 2's infinite horizon income from backing down in the current dispute.

**As long as the stakes in the dispute are lower than the critical threshold,  $\bar{\Delta}_2$ , state 2 receives a higher income from standing firm and waging a deterrent war than from backing down. So it is better off standing firm.**

Once the stakes rise above the threshold, however, this is no longer true. State 2 cannot impose more war costs on state 1 than it gained from initiating the dispute, even if state 2 sustains a deterrent war forever. So state 2 cannot deter state 1 from initiating more disputes of a similar kind in the future even by fighting forever in the current dispute. Hence if state 2 were to wage a deterrent war over the current dispute it would be wasting its resources for nothing. In this case state 2's infinite horizon income will be higher by backing down and accepting the income loss of  $\Delta$  per period in the current dispute than standing firm and waging a deterrent war over the dispute. This income comparison is shown in figure 30.

---

<sup>16</sup>Hypothetically, state 2 might be better off initiating a conflict over the current dispute and fighting forever in that conflict than accepting the loss of  $\Delta$  per period in the current dispute. However, this is not the binding constraint. The primary constraint discussed in the text is the binding constraint. See footnote X below.

If the stakes in the dispute rise higher than the critical threshold,  $\bar{\Delta}_2$ , then state 2 would receive a higher income from backing down than from standing firm and waging a deterrent war. So it is better off backing down.

The value of the threshold is  $\bar{\Delta}_2 = w_1$ . This conclusion bears two implications. First, the higher is state 1's war cost per period ( $w_1$ ), the higher is state 2's threshold for the size of the stakes in the dispute above which it loses its deterrent capacity ( $\bar{\Delta}_2$ ). Second, if disputes vary in their impact on state 1's war cost, then this will generate variations in state 2's security policies across disputes.

Any dispute that *increases* state 1's war cost will *increase* state 2's critical threshold,  $\bar{\Delta}_2$ , and give it *more* time to take preparatory actions and initiate a deterrent war over the dispute.

Conversely, any dispute that *decreases* state 1's war cost will *decrease* state 2's critical threshold,  $\bar{\Delta}_2$ , and give it *less* time to take preparatory actions and initiate a deterrent war over the dispute.

### 3.2.2 State 1's threshold for the maximum size of the stakes in the dispute

A similar logic holds in reverse for state 1, although the mechanics are somewhat different because of the asymmetry between the two states in the dispute (e.g. state 1 gaining from the change in the status quo, state 2 losing from the change). Once state 2 initiates the deterrent war mentioned above, state 1 has two options: to join the conflict or to acquiesce to the demand to stop taking the action in question.

Figure 30 demonstrates the logic behind state 1's critical threshold. The x-axis shows the size of the stakes in the dispute ( $\Delta$ ). The y-axis shows state 2's incomes. The thick solid line shows its income once state 1 is taking the action that imposes an income loss on it in each period, e.g. the action that causes the dispute ( $q_{2p}^* - \Delta$ ). The thick dashed line shows its income from standing firm and going to war over the dispute ( $q_{2p}^* - w_2$ ). Here  $w_2$  is state 2's war cost per period. The thick solid line shows its income from backing down in the dispute and allowing state 1 to continue taking the action that imposes an income loss on it in each period ( $q_{2p}^* - \Delta$ ).

The critical threshold is at  $\bar{\Delta}_1$ . Once the size of the stakes in this dispute is above this threshold, state 2's income from standing firm and going to war over the dispute is greater than its income from backing down in the dispute ( $(q_{2p}^* - w_2) > q_{2p}^* - \Delta$ ). So it is better off standing firm and fighting forever over the dispute than backing down. And state 1 cannot restore deterrence even by fighting forever in the current dispute. So state 1's cost of waging a deterrent war in the current dispute would be incurred for naught. In this case state 1 is better off backing down in the current dispute without fighting (e.g. giving up its right to take the action that gave it an income gain of  $\Delta$  per period) than it would be initiating a deterrent war over the dispute.<sup>17</sup> By implication, state 1 must initiate a deterrent war *before*  $\Delta$  reaches this critical threshold.

The value of this threshold is  $\bar{\Delta}_1$ . Thus, the higher is state 2's war cost per period,  $w_2$ , the higher is state 1's threshold for the upper limit of the size of the stakes in the dispute,  $\bar{\Delta}_1$ , above which it loses deterrent capacity. By implication, any dispute which *increases* state 2's war cost will *increase* state 1's threshold and give state 1 *more* time to take preparatory actions and initiate a deterrent war. Any dispute which *decreases* state 2's war cost will decrease state 1's threshold and give state 1 *less* time to take preparatory actions and initiate a deterrent war.

[Assume that a certain number of periods transpires before it becomes clear that the dispute is irreconcilable. In this interval of time, state 1 gains income at state 2's expense at the rate of  $\Delta$  per period. If  $\Delta$  is too large, then state 1's gain during this interval is so large that, even if state 2 then initiates a conflict and continues to retaliate for an infinite duration of time, it will not be sufficient to render state 1 worse off, in retrospect, having caused the dispute than having remained at peace at the status quo distribution of income and rights to actions. That is, even if state 2 were to sustain a retaliatory conflict forever, state 1 would still be better off causing further disputes in which it gains  $\Delta$  per period than remaining at peace at the status quo distribution. So state 2's costs of retaliating in the current dispute would be incurred for naught, because its war aim of restoring deterrence cannot be achieved.

---

<sup>17</sup>Hypothetically, state 1 might be better off initiating a conflict over the current dispute and fighting forever in that conflict than accepting the loss of  $\Delta$  per period in the current dispute. However, this is not the binding constraint. The primary constraint discussed in the text is the binding constraint. See footnote X below.

In this case state 2 is better off accepting the loss of  $\Delta$  per period in the current dispute without fighting than it would be initiating a conflict over the dispute.<sup>18</sup> By implication, if the value of  $\Delta$  starts at zero and increases over time during the pre-conflict interval, then state 2 must initiate a retaliatory conflict *before*  $\Delta$  reaches this critical threshold. Let this threshold be called  $\bar{\Delta}_2$ . (See appendix for proof.)]

[Thus state 2 must initiate conflict before state 1's income from standing firm and being made to fight forever becomes *greater* than its income from backing down. To derive observable implications from this imperative, let us set state 1's income from backing down to be greater than its income from standing firm and having to fight forever:

$$\sum_{t=0}^{\infty} q_{1p}^* d^t > \sum_{t=0}^{t=1} (q_{1p}^* + \Delta) d^t + \sum_{t=1}^{\infty} q_{1w}^* d^t. \quad (18.2)$$

Two implications are evident by inspection. First, as  $\Delta$  increases in magnitude, it eventually reaches a critical threshold at which this inequality is no longer true (other things being equal). This is  $\bar{\Delta}_2$ , the threshold mentioned above. Second, the higher is state 1's income from war ( $q_{1w}^*$ ) relative to its income from the status quo ante ( $q_{1p}^*$ ), the lower is this critical threshold (and conversely). The critical threshold is thus determined by the difference:  $q_{1p}^* - q_{1w}^*$ . The smaller this difference is, the lower the critical threshold  $\bar{\Delta}_2$  is. Notice that state 1's income from the status quo ante ( $q_{1p}^*$ ) is its income from backing down in the dispute. So the critical threshold is determined by the difference between state 1's income from backing down in the dispute and its income from standing firm and going to war over the dispute.]

[Suppose that state 1 were to acquiesce in this way for a certain number of periods before joining the conflict. During this interval it would lose  $\Delta$  per period and state 2 would gain  $\Delta$  per period.<sup>19</sup> If  $\Delta$  is too large, then state 2's gain during this interval will be so large that even if state 1 then joins the conflict and continues to fight for an infinite duration of time,

---

<sup>18</sup>Hypothetically, state 2 might be better off initiating a conflict over the current dispute and fighting forever in that conflict than accepting the loss of  $\Delta$  per period in the current dispute. However, this is not the binding constraint. The primary constraint discussed in the text is the binding constraint. See footnote X below.

<sup>19</sup>Compared to the situation in which state 2 permits state 1 to continue taking the original action that shifts income its way by  $\Delta$  per period.

it will not be sufficient to render state 2 worse off, in retrospect, having stood firm in the dispute than having backed down. Then state 1 cannot deter state 2 from causing further disputes in the future even if state 1 prolongs the current conflict forever. Hence, by the same logic as above, state 1 must join or even initiate the current conflict before  $\Delta$  increases to this critical threshold. Let this critical threshold be called  $\bar{\Delta}_1$ .]

[Thus state 1 must initiate conflict before state 2's income from standing firm and being made to fight forever becomes *greater* than its income from backing down. Setting its income from backing down to be greater than its income from standing firm and having to fight forever, we have:

$$\sum_{t=0}^{\infty} (q_{1p}^* - \Delta) d^t > \sum_{t=0}^{t=1} (q_{2p}^* - \Delta) d^t + \sum_{t=1}^{t=2} q_{2p}^* d^t + \sum_{t=2}^{\infty} q_{2w}^* d^t. \quad (18.3)$$

The implications are the same as that above. First, as  $\Delta$  increases in magnitude, it eventually reaches a critical threshold at which this inequality is no longer true (other things being equal). Second, the higher is state 2's income from war ( $q_{2w}^*$ ) relative to its income from backing down in the dispute ( $q_{2p}^* - \Delta$ ), the lower is the critical threshold value of  $\Delta$  (and conversely). Now the critical difference is  $q_{2p}^* - \Delta - q_{2w}^*$ . Again this difference varies from one type of spillover problem to the next, and that yields observable implications for differences in state behavior and international outcomes.]

### 3.3 A modeling strategy for explaining conflicts caused by disputes over income

The modeling strategy is based on this critical difference:  $q_{1p}^* - q_{1w}^*$ . If the international spillover problem *increases* state 1's income from war, it *lowers*  $\bar{\Delta}_2$  and forces state 2 to go to war sooner rather than later. Conversely, if the spillover problem *decreases* state 1's income from war, it *increases*  $\bar{\Delta}_2$  and enables state 2 to delay its entry into the war. As we will see, the critical difference  $q_{1p}^* - q_{1w}^*$  varies from one type of spillover problem to the next. These variations produce observable implications for differences in state behavior and international outcomes from one type of spillover problem to the next.

And it generates an intuitive modeling strategy for explaining conflicts caused by changes

in the status quo.

When the magnitude of each state's critical threshold varies from one type of dispute to another, that generates observable implications for variations in a range of other variables:

- how quickly each state will assess the income threat posed by the other state;
- how quickly each state will issue diplomatic threats;
- how quickly each state will mobilize its army and initiate conflict;
- whether its diplomatic threats will be seen as credible;
- whether its military action will be sustainable;
- whether and when it will contract allies to render its diplomatic threats credible and its military action sustainable; etc.

## 4. The puzzle of territorial over-expansion, reframed

### 4.1 The consequences of *not* choosing the army size that maximizes national income

Figure 30 shows a scenario in which state 1 increases its army size above the optimal size. Again the  $x$ -axis represents state 1's allocation of its total population between the military and civilian sectors, and the optimal army size is labelled  $A_1^*$ . The  $y$ -axis on the left side of the figure represents state 1's national income. The  $y$ -axis on the right side of the figure represents state 1's share of the international system's territory. As state one increases its army size above the optimal level, its income decreases but its share of the system's territory increases. Figure 31 shows the impact of these changes on income per capita. Again the  $x$ -axis represents the allocation of the population between the military and civilian sectors. Now the left  $y$ -axis represents income per capita, while the right  $y$ -axis represents state 1's share of the system's territory. As state 1 increases its army size above the optimal level, income per capita decreases (because total income decreases while total population remains constant). The farther state one increases its army size above the optimal level, the more territory it gains but the lower its income per capita falls.

A state that expands territorially in this way eventually runs into a domestic political

constraint. Either the soldiers refuse to fight because their wage rate is too low or the civilians refuse to work because their wage rate is too low. If the leadership does not reduce the army size to increase income per capita, eventually the soldiers or the civilians stage a riot, rebellion or revolution to press the leadership to reduce the army size and increase incomes again. Or they will remove the leadership and replace it with a new leadership that promises to reduce the army size and increase incomes again. The reality of this domestic political constraint is demonstrated by many episodes in European history. In many times and places, high war taxes caused falling incomes and political discontents of various kinds. Leaders recognized this constraint and, when it appeared to be the binding constraint, revised their strategies accordingly. Given these realities, why would a state ever expand territorially by such a large margin that its leadership risked being removed from office by discontented soldiers or civilians? To answer this question it is necessary to distinguish between the long run determinants of labor productivity that generate equilibrium in the long run, on one hand, and the short run factors that shift labor productivity and produce different equilibria in the short run, on the other hand.

## 4.2 A sudden drop in civilian labor productivity

Suppose that state 1 experiences an economic depression that lowers the marginal productivity of its civilian labor.<sup>20</sup> If the marginal productivity of civilian labor decreases *significantly*, this poses a strategic dilemma for its policy makers. In an industrial economy, the production of goods and services requires two inputs, labor and physical capital, in fixed proportions.<sup>21</sup> If the amount of physical capital in the civilian economy needs to be reduced by a significant margin, it can take years to accomplish. Therefore, when civilian labor productivity decreases, policymakers need to know whether the decrease is expected to be temporary or permanent. If it is expected to be *temporary* – because the depression is expected to end after a few years – then the optimal strategy is: (1) to maintain the current level of physical capital in the civilian economy, (2) reduce the amount of labor in the civilian economy by a

---

<sup>20</sup>In the next chapter I explain how an economic depression lowers civilian labor productivity. Briefly, the decrease in exports lowers the production levels of domestic firms, thereby lowering the amounts of output and income generated by a given number of workers.

<sup>21</sup>Most physical plant and equipment require a full complement of employees to operate at full capacity.

small margin temporarily, and (3) increase the size of the army by a small margin temporarily. This might be called a "weather the storm" strategy. This strategy maximizes national income in the long run, given that the drop in civilian labor productivity is temporary. By contrast, if the decrease in civilian labor productivity is expected to be *permanent*, then the optimal strategy is: (1) to reduce the amount of physical capital in the civilian economy by a significant margin, slowly over time, through attrition and lower investment rates, (2) reduce the amount of labor in the civilian economy by a large margin permanently, and (3) increase the size of the army by a large margin permanently. This might be called a "wholesale transformation" strategy. This strategy maximizes national income in the long run given that the drop in civilian labor productivity is permanent. However, if policymakers caught in this situation transfer the civilian workers into the military sector too quickly, this strategy will create severe shortages of manufactured goods in the short run.

These scenarios are displayed in figures 30 and 31. In both scenarios, the marginal productivity *parameter* for civilian labor is assumed to decrease by 20%. Figure 30 displays the first scenario in which the productivity parameter decreases temporarily and then returns to its original level. In this case the marginal productivity *schedule* for civilian labor decreases by only a small margin, and only temporarily, before returning to its original level. Figure 31 displays the second scenario in which the productivity parameter decreases permanently. In this case the marginal productivity *schedule* for civilian labor decreases permanently by a large margin. (Why does the same decrease of 20% in the productivity parameter cause the schedule to decrease by a small margin in the first scenario but a large margin in the second scenario? The answer is that the mathematical formula for the schedule includes the optimal amount of physical capital in the civilian economy. In the first scenario the optimal amount of physical capital remains constant at its original long-run level. In the second scenario the optimal amount of physical capital decreases significantly to a lower long-run level.)

In the first scenario, the optimal strategy is to increase the army size temporarily by a small margin, from  $A_1^*$  to  $A_1^{**}$ , and then reduce it to its original level of  $A_1^*$ . In the second scenario, the optimal strategy is to increase the army size permanently by a large margin, from  $A_1^*$  to  $A_1^{***}$ , and leave it there for the long run. Given this distinction between a change in the short run equilibrium (in the first scenario) and a change in the long run



equilibrium (in the second scenario), let us reconsider the original question. Assume that state 1 experiences an economic depression that lowers the marginal productivity of its civilian labor. And assume that policymakers expect the decrease to be temporary because they expect the depression to end after a few years. Then their optimal strategy is the first strategy: increase the army size temporarily by a small margin, from  $A_1^*$  to  $A_1^{**}$ , and then reduce it to its original level of  $A_1^*$ . If they were to choose the second strategy instead, they would be engaging in a territorial *over-expansion* that would lower national income below its maximum feasible level of  $q_1^{**}$ .

Why would a state ever make this error? Why would a state suffering a *temporary* drop in civilian labor productivity due to a depression make the mistake of assuming it is a *permanent* drop in productivity? More precisely, if a state suffers a sudden, temporary decrease in labor productivity, why would it ever make the mistake of assuming the productivity loss is permanent and go to war over the problem – if this would only intensify international security competition, increase pressures to adopt the income-maximizing army size, and drive the state into territorial over-expansion?

The short answer is: A state would make this mistake if its drop in civilian labor productivity was caused by other states' tariff increases, and it expected those states to increase tariffs repeatedly in the long run. From this perspective, the root cause of the problem of territorial overexpansion in such cases is the irreconcilability of the international tariff dispute. The goal of the next chapter is to draw as many observable implications as possible from this hypothesis.

### 4.3 A sudden drop in military labor productivity

Suppose that state 1 experiences a provincial rebellion that lowers the marginal productivity of its labor in the military/governance sector.<sup>22</sup> If the marginal productivity of military/governance labor decreases *significantly*, this poses a strategic dilemma for the central

---

<sup>22</sup>In the next chapter I explain how a provincial rebellion lowers labor productivity in the military/governance sector. Briefly, when a province rebels against the central government, it increases the number of soldiers and policemen the government needs to employ to maintain domestic order in each hectare of territory in that province. This effectively lowers the number of hectares of territory that can be brought to order by a given number of soldiers and policemen.

government. The production of domestic law and order requires two inputs – government personnel and institutional infrastructure – in fixed proportions.<sup>23</sup> If the amount of institutional infrastructure is lowered by a significant margin (for example by the loss of an entire province to a secession movement), the optimal strategy is to reduce the number of government personnel significantly too. Therefore, when a provincial rebellion occurs, the central government needs to know whether its loss of control of that province is expected to be temporary or permanent. If it is expected to be *temporary* – because the rebels are expected to come to terms easily – then the optimal strategy is to "weather the storm" by: (1) decreasing the size of the army by a small margin temporarily, and (2) increasing the amount of labor in the civilian economy by a small margin temporarily. This strategy maximizes national income in the long run, given that the drop in labor productivity in the military-government sector is temporary. By contrast, if the decrease in labor productivity in this sector is expected to be *permanent*, then the optimal strategy is "wholesale transformation": (1) reducing the size of the army and the governance workforce by a large margin permanently, and (2) increasing the amount of labor in the civilian economy by a large margin permanently. This strategy maximizes national income in the long run given that the drop in labor productivity in the military-government sector is permanent. By implication, if policymakers fail to adopt this strategy after the permanent loss of one province to a secession movement, they will fail to maximize national income. That is, if policymakers fail to transfer soldiers and government workers into the civilian sector after losing a province, national income will fall significantly below its maximum feasible level, and the incomes of landowners in the remaining provinces will fall significantly below their maximum feasible level.

These scenarios are displayed in figures 30 and 31. In both scenarios, the marginal productivity *parameter* for labor in the military-government sector is assumed to decrease by 20%. Figure 30 displays the first scenario in which the productivity parameter decreases temporarily and then returns to its original level. In this case the marginal productivity *schedule* for civilian labor decreases by only a small margin, and only temporarily, before returning to its original level. Figure 31 displays the second scenario in which the produc-

---

<sup>23</sup>Most government institutions require a full complement of employees to operate at full capacity.

tivity parameter decreases permanently. In this case the marginal productivity *schedule* for civilian labor decreases permanently by a large margin. (Why does the same decrease of 20% in the parameter cause the schedule to decrease by a small margin in the first scenario but a large margin in the second scenario? The answer is that the mathematical formula for the schedule includes the optimal amount of institutional infrastructure in society. In the first scenario the optimal amount of institutional infrastructure remains constant at its original long-run level. In the second scenario the optimal amount of institutional infrastructure decreases significantly to a lower long-run level.)

In the first scenario, the optimal strategy is to decrease the army size temporarily by a small margin, from  $A_1^*$  to  $A_1^{**}$ , and then increase it to its original level of  $A_1^*$ . In the second scenario, the optimal strategy is to decrease the army size permanently by a large margin, from  $A_1^*$  to  $A_1^{***}$ , and leave it there for the long run. Given this distinction between a change in the short run equilibrium (in the first scenario) and a change in the long run equilibrium (in the second scenario), let us reconsider the original question. Assume that state 1 experiences a provincial rebellion that lowers the marginal productivity of its labor in the military-government sector. And suppose that policymakers expect the decrease to be permanent because they expect to lose control of the province permanently. Then their optimal course of action is the second strategy: Decrease the army size permanently by a large margin, from  $A_1^*$  to  $A_1^{***}$ , and leave it there for the long run. If they were to choose the first strategy instead, they would be engaging in a territorial over-expansion that would lower national income below its maximum feasible level of  $q_1^{***}$ .

Why would a state ever make this error? Why would a state suffering a *permanent* drop in military labor productivity due to a provincial secession make the mistake of assuming it is only a *temporary* drop in productivity? More precisely, if a state suffers a sudden, permanent decrease in military labor productivity, why would it ever make the mistake of assuming the productivity loss is temporary and go to war over the problem – if that would only intensify international security competition, increase pressures to adopt the income-maximizing army size, and drive the state into territorial over-expansion?

The short answer is: A state would make this mistake if its drop in military labor productivity was caused by provocation of the secession movement from a neighboring state,

it expected the provocation to continue and expand the secession movement to its other provinces, and the only way to stop this was to recapture the first province to secede. From this perspective, the root cause of the problem of territorial overexpansion in such cases is the irreconcilability of the international dispute over provocation of secessionists. The goal of the next chapter is to draw as many observable implications as possible from this hypothesis.

#### **4.4 Are states driven to war by the threat of income losses or the threat of power shifts?**

If this theory is correct, then hegemonic wars were caused by the threat of income losses due to irreconcilable disputes over income. While power did shift before and during each war, such power shifts did not cause any of these wars. The goal of the remainder of this part of the study is to derive and test the observable implications of this theory for variations in state behaviors prior to the outbreak of each war. Parts III-V of the study derive and test the observable implications for variations in state behaviors during and after each war.

### **5. Figures for this chapter**

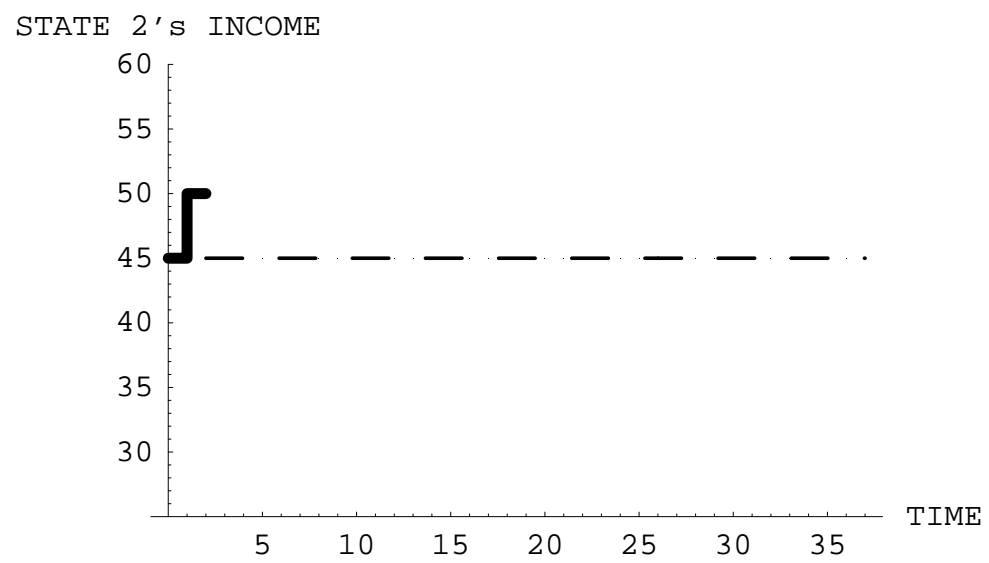


Figure 18.1: Dispute between state 1 and state 2 over income;  
Stakes in dispute = 5

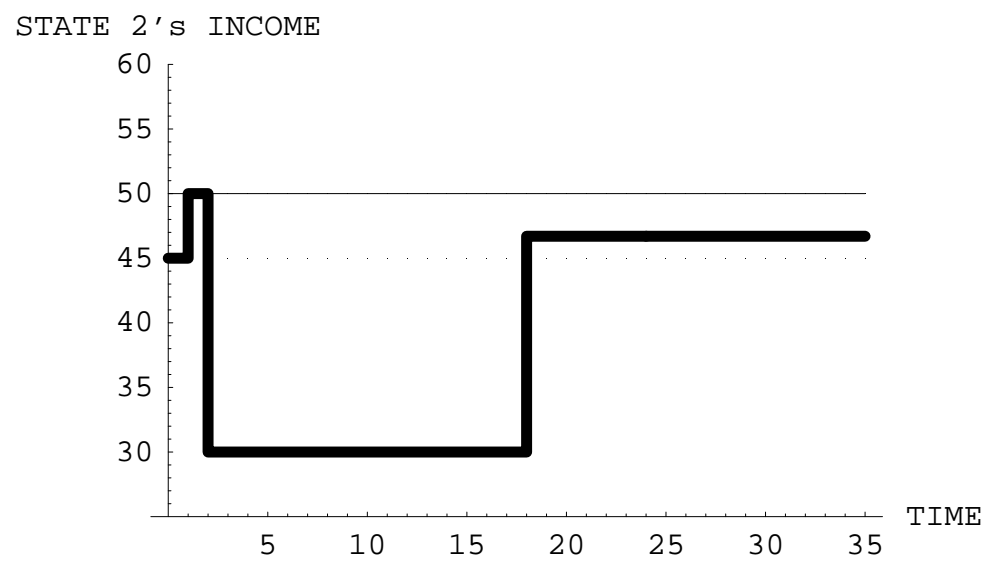
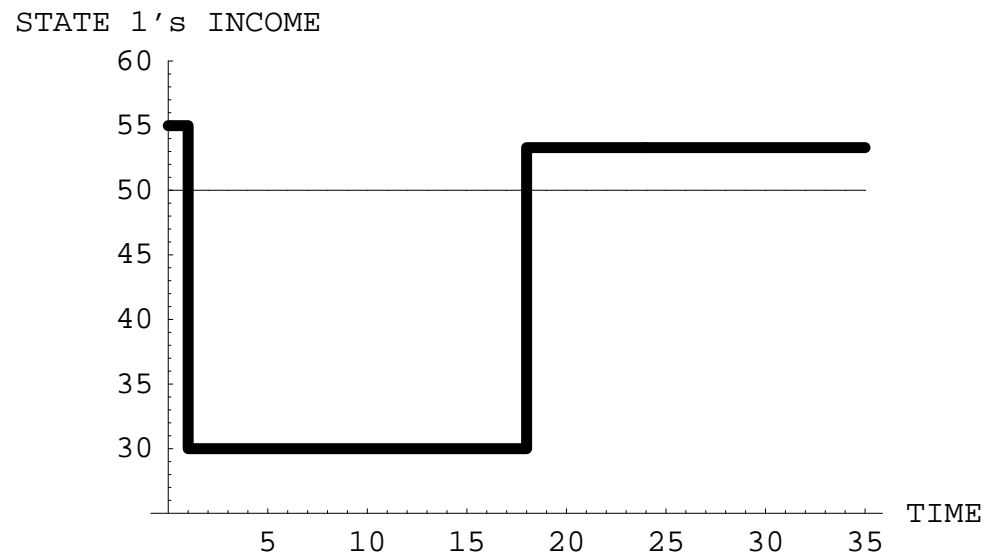
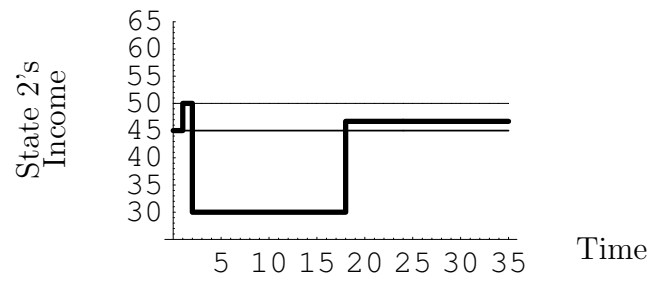
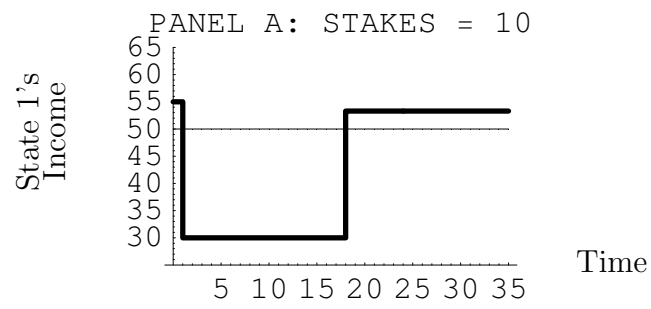


Figure 18.2: Deterrent war between state 1 and state 2 over a dispute



Panel A: Stakes = 5

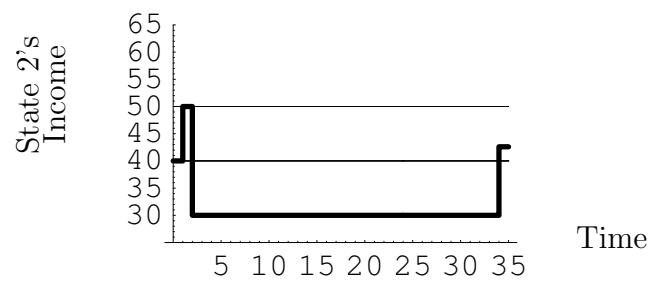
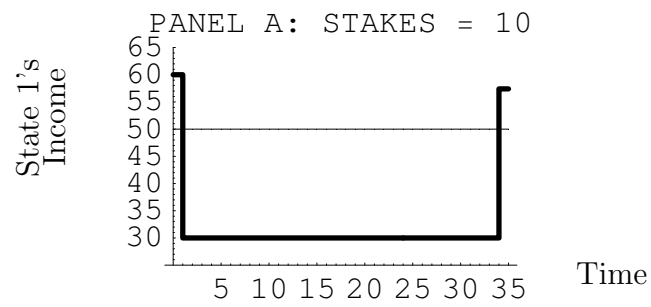


Figure 18.3

Panel B: Stakes = 10

# SECESSION CRISIS

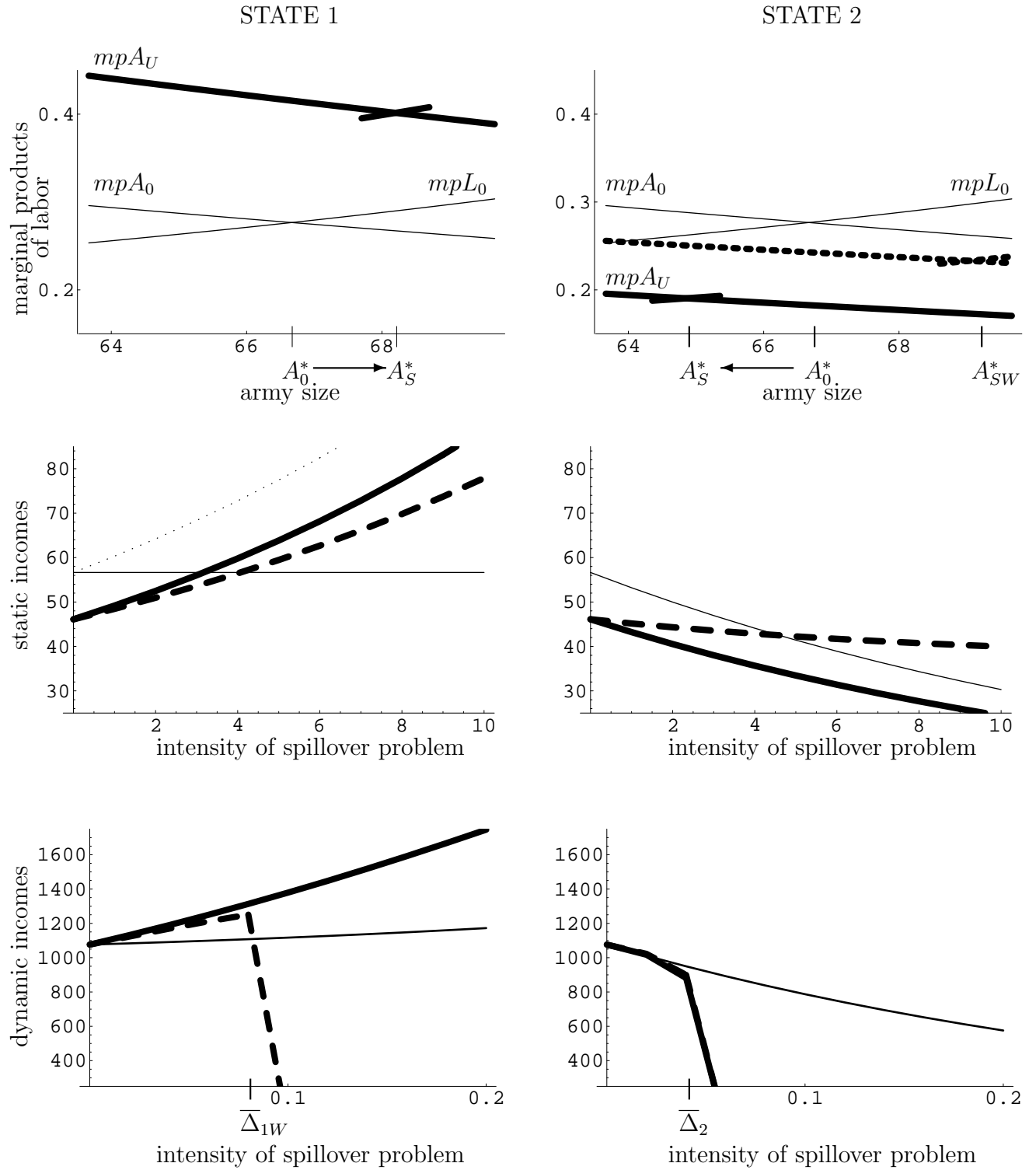


Figure 18.4: Impact of international spillover problem on each state's optimal army size, static income and dynamic income for the case of secession spillovers



## CHAPTER 19

### Observable implications of the hypothesis of income loss avoidance

There were seven hegemonic wars in Europe between 1550 and 1950. Each war erupted because one state had developed a new domestic activity that increased its income but decreased another state's income through some kind of international spillover effect. The second state's income loss caused a dispute between the two states. Did the first state have the right to continue its new domestic activity? Or did the second state have the right to be free of the income losses caused by the activity? The existing international laws and agreements between the two states did not resolve the dispute, because those pacts contained "incompatible rights". One pact gave one state the right to take the action in question while another pact gave the other state the right to a cessation of the action. In each case bargaining over the dispute broke down and war broke out.

To explain what caused these wars, two questions need to be answered.

1. Why did each state decide to go to war over the dispute rather than settle it peacefully?
2. Once each state decided to go to war over the dispute rather than settle it peacefully, why did it initiate military hostilities when it did rather than sooner or later?

In answer to the first question, each state decided to go to war because the legal incompatibility gave it uncertainty about the other state's intentions to initiate more disputes of a similar kind that would impose more income losses on it. Each state reasoned that it needed to go to war and impose more war costs on the other state than it might gain from initiating more disputes, or else the other state would have an incentive to initiate more. Thus each state went to war to deter the other state from imposing more income losses on it. In answer to the second question, each state initiated military hostilities when it did because the income stakes in the dispute were increasing toward its critical deterrence threshold.

As explained in the previous chapter, each state had to go to war before the income stakes reached its critical deterrence threshold, or else it would lose the capacity to deter the other state from imposing more income losses on it.

Each state faced the risk of losing its deterrent capacity if it did not go to war before the stakes rose above its threshold. In some cases the critical deterrence threshold was relatively low, so the income stakes in the dispute approached the threshold quickly as the international spillover problem intensified. And the war started quickly after the dispute first erupted. In other cases the critical deterrence threshold was relatively high, so the income stakes in the dispute approach the threshold slowly as the international spillover problem intensified. And the war did not start until long after the dispute first erupted. This contrast is explained by the impact of the international spillover problem on each state's optimal army size (as will be shown in this chapter). However, these shifts in relative power between the two states only determined the *timing* of each state's entry into military hostilities – whether to initiate military hostilities quickly after the dispute erupted or much later. They did not determine each state's choice between going to war and settling the dispute peacefully. This choice was determined by the threat of increasing *income* losses if the other state was not deterred from initiating more disputes that would impose more income losses.

In this chapter I delineate the causal assumptions behind this theory of hegemonic war and derive its observable implications. If the assumptions and implications are corroborated by the evidence in the following chapters, it will support the conclusion that these wars were caused by disputes over income rather than shifts in relative power. The theory thus provides the opportunity for a "critical test" between these rival explanations of hegemonic war. Insofar as the theory's assumptions and implications contradict those of other explanations of hegemonic war that have been offered, it also provides for "critical tests" of those other explanations as well.

# 1. A theory of wars caused by disputes over "incompatible rights"

## 1.1 Stylized facts and causal assumptions

In each of the historical cases, one state developed a new domestic activity that had an intended effect and an unintended effect. The intended effect was to increase its income. The unintended effect was to decrease another state's income through some kind of international spillover effect. The unintended spillover effect created a problem of "incompatible rights". For according to existing international laws and agreements between the two states, the first state had the right to engage in its new domestic activity, but the second state had the right to freedom from the spillover effect and the income losses. So the first state's *exercise* of its right to engage in its new domestic activity was perceived by the second state as a *violation* of its right to freedom from the spillover effect and the income losses. The same problem also occurred in reverse. While the second state had a right to immunity from the spillover effect and the income losses, it did not have the right – under existing international laws and agreements – to interfere in the first state's new domestic activity. So the second state's *exercise* of its right to freedom from these effects was perceived by the first state as a *violation* of its right to freedom from foreign interference in its domestic activities. This was the problem of "incompatible rights": One international law or agreement gave one state the right to *take* an action while another one gave the other state the right to a *cessation* of that action.

Because of this incompatibility, each state perceived the other state to be violating its rights under existing international laws and agreements. The mutual perceptions of wrongdoing provoked a dispute and an international crisis. The problem of "incompatible rights" prevented the dispute from being settled peacefully. I now state the causal assumptions and the observable implications of this theory more precisely. In what follows I call the state that developed a new domestic activity "state 1" and the state that was harmed by its international spillover effects "state 2".

In each historical case there were nine "facts on the ground" prior to the outbreak of the international dispute and the descent into war. I take these facts as the starting point of my

theory of what caused each dispute and war: the causal assumptions of this theory.

**Fact 1.** State 1 develops a new domestic activity that increases its income but generates a negative spillover effect on state 2.

**Fact 2.** State 2 suffers income losses from the spillover effect and demands that state 1 stop its new domestic activity.

**Fact 3.** Each state perceives that the other state is committing a violation of some international law or norm. State 2 perceives that state 1's new domestic activity is violating international law because of its international spillover effect. State 1 perceives that state 2's demand to stop the activity is an interference in its domestic affairs and, as such, a violation of international law.

**Fact 4.** Each state expects to incur income losses from the other state's violation if it is not stopped.

**Fact 5.** Each state perceives that the other state's illegal activity is increasing in magnitude or intensity.

**Fact 6.** Each state expects to incur increasing income losses if it does not stop the other state's illegal activity.

**Fact 7.** Each state identifies a legal justification for opposing the other state's illegal activity.

**Fact 8.** Each state concedes that the other state's legal case has some merit so concessions to the other state are required, but its own case has merit too and justifies its opposition.

**Fact 9.** Each state offers concessions to the other state on account of the merits of its case, but maintains its claim to deserve concessions in return on account of the merits of its own case.

This theory generates a number of observable implications for patterns of evidence that are expected prior to the outbreak of each war.

## 1.2 Observable implications of the theory

**Implication 1.** Each state issues a demand that the other state stop its illegal activity.

**Implication 2.** Each state issues a threat that if the other state does not stop its illegal activity, then it will initiate military hostilities to force the other state to stop.

**Implication 3.** Each state refuses to capitulate to the other state's demand on the ground that it will be less costly to wage a war over the issue than to allow the other state to continue to violate international law and impose more income losses on it.

**Implication 4.** Each state perceives the other state's refusal to concede as evidence that it is violating international law intentionally and it intends to continue to violate the law and impose more income losses.

**Implication 5.** Each state renews its demands, threats and intransigence to the other state's demands and threats.

**Implication 6.** Each state declares war on the other state and enters military hostilities.

If these observable implications are corroborated by the empirical evidence, then it will support this explanation of what caused each state's decision to go to war over the dispute rather than settling it peacefully.

## **2. Why each state initiated military hostilities when it did - rather than sooner or later**

Each of the seven hegemonic wars was caused by a dispute over an international spillover problem. The dispute became irreconcilable because of the legal incompatibility (as explained above). The dispute was contentious because the income stakes for each state were high. State 1's new domestic activity *increased* its labor productivity in the military or civilian sector, thereby increasing its income. But the international spillover effect *decreased* state 2's labor productivity in the military or civilian sector, thereby lowering its income. The particular impact on labor productivity depended, however, on the type of spillover problem that caused the war.

These wars were caused by three different kinds of international spillover problem. Some of the wars were caused by *secession* problems. Others were caused by *revolution* problems. And one war was caused by a *depression* problem. Each type of spillover problem had a different impact on the relative productivity of military and civilian labor in each state (e.g. state 1 and state 2).

1. Secession problem. In these cases state 1 developed a domestic governance reform that increased its income but inspired a similar reform movement across the border in state 2, which then erupted into a secession crisis in state 2. State 1's domestic governance reform increased its productivity of military labor, while the secession crisis in state 2

decreased its productivity of military labor.

2. Revolution problem. In these cases state 1 experienced a domestic political revolution that raised the incomes of its lower and middle classes but inspired a similar revolutionary movement across the border in state 2, which then lowered the incomes of its upper classes. State 1's revolution increased the productivity of military and civilian labor for its lower and middle classes, while the revolutionary movement in state 2 decreased the productivity of military and civilian labor for its upper classes.
3. Depression problem. In this case state 1 enacted trade tariffs and international lending restrictions that raised the incomes of its domestic firms and workers (above what they would have been otherwise), but exported its depression to state 2, thereby lowering the incomes of its firms and workers (below what they would have been otherwise). State 1's tariffs increased the productivity of its civilian labor, while the effects of those tariffs on state 2 decreased the productivity of its civilian labor.

Because each type of spillover problem had a different impact on the relative productivity of military and civilian labor in each state, each type of problem had a different impact on each state's optimal army size. These differences are illustrated in figure 19.1. Each panel in the figure shows one state's marginal products of military and civilian labor both before and after the spillover problem erupts, and the resulting change in its optimal army size. The *thin* solid lines represent the marginal products before the spillover problem erupts, while the *thick* solid lines represent the marginal products after the spillover problem erupts. Notice that one or both of state 1's marginal product curves risers in each type of spillover problem, thereby increasing its income. By contrast, one or both of state 2's marginal product curves falls in each type of spillover problem, thereby decreasing its income. Notice that the impact on each state's optimal army size is different in each type of crisis. State 1's optimal army size increases in a secession crisis but decreases in a depression crisis, while it remains unchanged in a revolution crisis. Conversely, state 2's optimal army size decreases in a secession crisis but increases in a depression crisis, while it remains unchanged in a revolution crisis.

These differences in marginal labor productivity across the three types of spillover problem drove state strategies in exactly the way predicted by the theory of income disputes outlined in the previous chapter. Each type of spillover problem had a different impact on

each state's optimal army size. Hence each type of spillover problem had a different impact on each state's peace-war income differential, which in turn had a different impact on the *other* state's critical deterrence threshold for the income stakes in the dispute beyond which it would lose deterrent capacity. And finally, each state's critical threshold determined its strategies for crisis bargaining and the initiation of military hostilities: how quickly or slowly it would assess the spillover problem, prepare for war, issue threats and enter military hostilities.

In the following subsections, I explain how each type of spillover problem affected these variables and derive the observable implications for differences in state strategy across the three types of spillover problem. In analyzing each type of problem, I assume that it was *state 1* that developed a new domestic activity that increased its income and *state 2* that suffered the international spillover effect and income losses as a result.

## 2.1 Secession Crisis

**Fact 1s.**<sup>1</sup> State 1 enacts a domestic governance reform to improve its local communal organization along religious or ethnic lines. The reform increases the productivity of state 1's labor devoted to military-governance functions, thereby increasing its income.

Consider the mechanism by which state 1's governance reform increases its labor productivity in the military-government sector. The reform reduces social frictions among community members and thereby reduces the amount of labor the central government needs to employ in the military-governance sector to maintain domestic order. This constitutes, in effect, a decrease in the government's marginal cost of maintaining order: The fewer soldiers and policemen it takes to maintain order on a hectare of territory, the lower is the marginal cost of labor devoted to maintaining order on each hectare. This can be described, equivalently, as an increase in the marginal productivity of labor devoted to maintaining order: The fewer soldiers and policemen it takes to maintain order on each hectare of territory, the more hectares of territory can be provided with order by a given number of soldiers and po-

---

<sup>1</sup>Recall **Fact 1** from above: "State 1 develops a new domestic activity that increases its income but generates a negative spillover effect on state 2."

licemen.<sup>2</sup> This increase in labor productivity is shown in the upper left panel of figure 19.2. In this panel the x-axis represents state 1's allocation of labor between the civilian economic sector and the military-governance sector (as in figure 19.1 above). The y-axis represents marginal labor productivity (as in figure 19.1 above). The thin solid line shows state 1's marginal product of labor in the military-governance sector before its domestic governance reform ( $mpA_{SQ}$ ). The thick solid line shows the impact of a 20% increase in military labor productivity: the marginal product schedule shifts *upward* (to  $mpA_U$ ).

**State 1's domestic governance reform increases its labor productivity in the military-governance sector, thereby increasing its optimal army size from  $A_1^*$  and  $A_{1s}^*$ .**

For example, in the Kingdom of Bohemia in the 16th and 17th centuries, local governance was reformed along the lines preferred by the Protestant majority, thereby reducing social frictions among members of all religions.

**Fact 2s.** State 1's domestic governance reform inspires a similar movement across the border in the provinces of state 2 with a similar religious or ethnic composition. When state 2's central government rejects the demands of the provincial reformists, they launch a provincial secession movement and declare independence from state 2. This secession crisis decreases the productivity of state 2's labor devoted to military-governance functions, thereby decreasing its income.

Consider the mechanism by which state 2's secession crisis decreases its labor productivity. The provincial secession causes a domestic conflict between the secessionists and the central government that has two adverse effects. First, the conflict increases the number of personnel the central government must employ to maintain order in those provinces. This constitutes, in effect, an increase in the government's marginal cost of maintaining domestic order.<sup>3</sup> Second, the conflict damages agricultural land and equipment and thereby reduces the amount of goods, services and income generated by a given area of territory. Hence a given number of personnel devoted to maintaining order generates less income in that area

---

<sup>2</sup>In the economic theory of production, the principle of "duality" implies that a marginal cost function is mathematically equivalent, or "isomorphic", to a marginal productivity function.

<sup>3</sup>The more personnel it takes to maintain order in a province, the higher is the marginal cost of maintaining order in that province.



than it did before the secession crisis broke out. To put it another way, it takes more personnel to generate the same amount of income from that area (compared to the situation before the secession crisis broke out). This also constitutes, in effect, an increase in the government's marginal cost of maintaining order.<sup>4</sup>

Both increases in the marginal cost of maintaining domestic order can be described, equivalently, as decreases in the marginal productivity of labor devoted to maintaining order: The more soldiers and policemen it takes to maintain order per hectare of territory, the fewer hectares of territory can be provided with order by a given number of soldiers and policemen. This decrease in labor productivity is shown in the upper right panel of figure 19.2. The x-axis represents state 2's allocation of labor between the civilian economic sector and the military-governance sector (as in figure 19.1 above). The y-axis represents marginal labor productivity (as in figure 19.1 above). The thin solid line shows state 2's marginal product of labor in the military-governance sector before the secession crisis erupts ( $mpA_{SQ}$ ). The thick solid line shows the impact of a 20% decrease in military labor productivity: the marginal product schedule shifts *downward* (to  $mpA_U$ ).

**State 2's domestic secession crisis decreases its labor productivity in the military-governance sector, thereby decreasing its optimal army size from  $A_2^*$  and  $A_{2s}^*$ .**

To continue with the same example, the domestic governance reform in the Kingdom of Bohemia, mentioned above, inspired a similar reform movement across the border in the Kingdom of Austria, where some provinces also had a high percentage of Protestants. When their demands for reform were rejected by Austria's central government, they attempted to secede from the Kingdom of Austria, thereby provoking a domestic secession conflict that destroyed property and required heavy troop deployments to restore domestic order.

### 2.1.1 Observable implications of the changes in optimal army size

**Implication 7s.** Since state 2's optimal army size decreases, it can issue full-bore diplomatic threats from the start of the crisis if necessary.

---

<sup>4</sup>The more personnel it takes to maintain order on a hectare of territory, the higher is the marginal labor cost of maintaining order per hectare of territory.

**Implication 8s.** Since state 2's optimal army size decreases, it can take full-strength military action from the start of the crisis if necessary.

**Implication 9s.** Since state 1's optimal army size increases, it cannot issue full-bore diplomatic threats from the start of the crisis (if threats are necessary). Rather it will delay in issuing diplomatic threats while it amasses capability and obfuscate the military threat it poses in the meanwhile.

**Implication 10s.** Since state 1's optimal army size increases, it will take hedging military action if immediate action is necessary. Otherwise it will delay in taking military action until it amasses the optimal capability.

**Implication 11s.** Since state 1's optimal army size increases and state 2's optimal army size decreases, the frontline in the war will be within state 2's borders.

### 2.1.2 The impact of state 1's peace-war income differential on state 2's critical deterrence threshold

First consider how state 1's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle left panel of figure 19.2. The x-axis represents the effectiveness of state 1's domestic governance reform in terms of the percentage increase in its labor productivity in the military/government sector (assumption 1s above). This is the productivity gain that state 1 would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the reform. This axis thus represents the size of the stakes in the international dispute for state 1. The y-axis represents state 1's single-period incomes from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state 2. State 1's income from peace remains constant as the stakes in the dispute rise, because if it backs down and abandons its domestic governance reform it forgoes the productivity gain from the reform and its income remains at the *status quo ante*. By contrast, its income from war increases steadily as the stakes in the dispute rise, for two reasons. First, when it stands firm in the dispute, it retains the productivity gain from the domestic governance reform. Second, when it goes to war it benefits from the favorable change in relative army size (e.g. the reform increases its own army size while the spillover effect decreases state 2's army size). Consequently, the

difference between state 1's peace and war incomes *narrows rapidly* as the income stakes in the dispute rise.

The impact on state 2's critical deterrence threshold is displayed in the bottom right panel of figure 19.2. The x-axis represents the intensity of the international spillover effect on state 2 in terms of the percentage decrease in its labor productivity in the military/government sector as its domestic secession spreads from province to province (assumption 2s above). This is the productivity loss that state 2 would suffer if it were to back down in the international dispute and acquiesce to state 1's demand to accept the spillover effect. This axis thus represents the size of the stakes in the international dispute for state 2. The y-axis represents state 2's long run incomes from peace and war. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 1. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 1. The income from peace declines steadily as the domestic secession movement spreads from province to province at home. The income from war declines even more steeply, however, because of the unfavorable change in relative army size (e.g. state 1's army size increases while state 2's army size decreases). As a result, state 2's long run income from standing firm and going to war *quickly falls below* its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence state 2's critical deterrence threshold is *low* compared to the threshold in the other types of spillover problem studied below. State 2's threshold is shown by  $\bar{\Delta}_2$  in the figure.

**State 1's peace-war income differential decreases rapidly as the income stakes in the international dispute rise. This trend lowers state 2's critical deterrence threshold compared to its level in the two other types of spillover problem.**

### **2.1.3 Observable implications: state 2's crisis bargaining and war initiation strategies**

This analysis generates a number of observable implications for the historical evidence on state 2's strategies.

**Implication 12s.** State 2 will assess the income threat posed by the international spillover problem quickly and impatiently.

**Implication 13s.** State 2 will issue a firm demand that state 1 stop the domestic activity that is causing the spillover effect before it becomes clear whether the spillover problem is serious.

**Implication 14s.** State 2 will issue a military threat to back up its demand quickly after the spillover problem arises.

**Implication 15s.** The purpose of state 2's military ultimatum will be to cause deadlock and forestall further negotiations with state 1.

**Implication 16s.** The duration of time from the emergence of the international spillover problem to state 2's decision to initiate war will be comparatively short (compared to the duration in the other types of spillover crisis, revolutions and depressions).

**Implication 17s.** State 2's demand to stop generating the spillovers will be seen as militarily non-credible by state 1.

**Implication 18s.** State 2's military action will be unsustainable without allies.

#### 2.1.4 The impact of state 2's peace-war income differential on state 1's critical deterrence threshold

Consider how state 2's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle right panel of figure 19.2. The x-axis represents the intensity of the international spillover effect on state 2 in terms of the percentage decrease in its labor productivity in the military/government sector as its domestic secession crisis spreads from province to province (causal assumption 2s). This is the productivity loss that state 2 would incur if it were to back down in the international dispute and acquiesce to state 1's demand to accept the spillover effect. This axis thus represents the size of the stakes in the dispute for state 2. The y-axis represents state 2's single-period incomes from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state 1. State 2's income from peace declines steadily as the stakes in the dispute rise because of its productivity loss from the international

spillover effect. Its income from war also declines steadily as the stakes in the dispute rise, but for a different reason: the unfavorable change in relative army size generated by the spillover problem (e.g. the decrease in its army size and the increase in state 1's army size). Consequently, the difference between state 2's peace and war incomes remains constant as the income stakes in the dispute rise.

The impact on state 1's critical deterrence threshold is displayed in the bottom left panel of figure 19.2. The x-axis reflects the effectiveness of state 1's domestic governance reform in terms of the percentage increase in its labor productivity in the military/government sector (causal assumption 1s). This is the productivity gain that state 1 would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the reform. This axis thus represents the size of the stakes in the dispute for state 1. The y-axis represents state 1's long run incomes from peace and war. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 2. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 1. The income from peace remains constant as the stakes in the dispute rise for the same reason that the single-period income remains constant. If it backs down and abandons its domestic governance reform, it forgoes the productivity gain from the reform. By contrast, the income from war increases steadily as the stakes in the dispute rise, for two reasons. First, when it stands firm in the dispute, it retains the productivity gain from the domestic governance reform. Second, when it goes to war it benefits from the favorable change in relative army size (e.g. the reform increases its army size and the spillover effect decreases state 2's army size). As a result, state 1's long run income from standing firm and going to war never falls below its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence it has no critical deterrence threshold.

**State 2's peace-war income differential remains unchanged as the income stakes in the international dispute rise. Hence state 1 has no critical deterrence threshold.**

### 2.1.5 Observable implications: state 1's crisis bargaining and war initiation strategies

Here are the observable implications for the historical evidence on state 1's strategies.

**Implication 19s.** State 2 will perceive state 1's potential opposition to the demand to stop generating the spillovers as militarily credible .

**Implication 20s.** If the moderates in state 1 propose to respond to state 2's demand with reassurances and accommodations, then the hardliners in state 1 will not force a rush to war by accusing the moderates of caving in to foreign threats.

**Implication 21s.** State 1 will respond to state 2's demand with reassurances and accommodations.

**Implication 22s.** State 1 will not issue demands and deadlines backed by military threats until after state 2 initiates military hostilities.

**Implication 23s.** State 1's military action will be sustainable without allies.

## 2.2 Depression Crisis

**Fact 1d.**<sup>5</sup> State 1 is experiencing an economic depression. Its government responds by enacting trade tariffs. The tariffs increase the productivity of its labor devoted to civilian economic production, thereby increasing its income. The tariffs reduce its demand for state 2's exports, however, thereby exporting its depression to state 2.

Consider how state 1's tariffs increase its labor productivity in the civilian economic sector. The tariffs increase the prices that domestic consumers pay for imported goods and induce them to buy domestically produced goods instead, thereby increasing the incomes of domestic firms and workers above what they would have been otherwise.<sup>6</sup> These income gains can be represented as an increase in the marginal productivity of labor devoted to civilian economic production: A given number of civilian workers produces more goods, services and income after the tariffs are enacted. This increase in labor productivity is shown in the upper left panel of figure 19.3. The x-axis represents the state 1's allocation of labor

---

<sup>5</sup>Recall **Fact 1** from above: "State 1 develops a new domestic activity that increases its income but generates a negative spillover effect on state 2."

<sup>6</sup>Technically, this is an income-shifting policy rather than an income-generating policy, because it simply shifts income from foreign firms and workers to domestic firms and workers. See FN.

between the civilian economic sector and the military-governance sector (as in figure 19.1 above). The y-axis represents marginal productivity (as in figure 19.1 above). The thin solid line shows state 1's marginal product of labor in the civilian economic sector before it enacts the tariffs. The thick solid line shows the impact of a 20% increase in civilian labor productivity: the marginal product schedule shifts *upward*.

**State 1's tariffs increase its labor productivity in the civilian economic sector, thereby decreasing its optimal army size from  $A_1^*$  and  $A_{1d}^*$ .**

For example, in the late 1920s, the economies of France, Great Britain and the United States went into depression. Their governments responded by enacting trade tariffs. These policies induced consumers to buy more domestically produced goods, thereby increasing the incomes of domestic firms and workers above what they would have been otherwise.

**Fact 2d.** State 1's tariffs reduce its demand for state 2's exports. The fall in state 2's exports decrease its productivity of labor devoted to civilian economic production, thereby decreasing its income.

Consider how state 1's policies decrease state 2's labor productivity. State 1's trade tariffs reduce the demand for state 2's export goods, thereby lowering production levels and reducing the incomes of its firms and workers. The income losses in state 2 can be described as a decrease in the marginal productivity of its labor devoted to civilian production: A given number of civilian workers produces less goods, services and income after the drop in exports. This decrease in state 2's labor productivity is shown in the upper right panel of figure 19.3. The x-axis represents the state 2's allocation of labor between the civilian economic sector and the military-governance sector (as in figure 19.1 above). The y-axis represents marginal productivity (as in figure 19.1 above). The thin solid line shows state 2's marginal product of labor in the civilian economic sector before state 1 enacts the trade tariffs. The thick solid line shows the impact of a 20% decrease in state 2's civilian labor productivity: the marginal product schedule shifts *downward*.

**State 2's export losses decrease its labor productivity in the civilian economic sector, thereby increasing its optimal army size from  $A_2^*$  and  $A_{2d}^*$ .**

To continue with the same example, Germany's economy went into depression in the late 1920s. The depression deepened considerably when France, Germany and the United States enacted trade tariffs. Germany's export losses lowered the incomes of its firms and workers.

### 2.2.1 Observable implications of the changes in optimal army size

**Implication 7d.** Since state 1's optimal army size decreases, it can issue full-bore diplomatic threats from the start of the crisis if necessary.

**Implication 8d.** Since state 1's optimal army size decreases, it can take full-strength military action from the start of the crisis if necessary.

**Implication 9d.** Since state 2's optimal army size increases, it cannot issue full-bore diplomatic threats from the start of the crisis (if threats are necessary). Rather it will delay in issuing diplomatic threats while it amasses capability and obfuscate the military threat it poses in the meanwhile.

**Implication 10d.** Since state 2's optimal army size increases, it will take hedging military action if immediate action is necessary. Otherwise it will delay in taking military action until it amasses the optimal capability.

**Implication 11d.** Since state 1's optimal army size decreases and state 2's optimal army size increases, the frontline in the war will be beyond state 2's borders.

### 2.2.2 The impact of state 1's peace-war income differential on state 2's critical deterrence threshold

First consider how state 1's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle left panel of figure 19.3. The x-axis reflects the effectiveness of state 1's tariffs in terms of the percentage increase in its civilian labor productivity (causal assumption 1d). This is the productivity gain that state 1 would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the tariffs. This axis thus represents the size of the stakes in the dispute for state 1. The y-axis represents state 1's single-period incomes from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state



2. State 1's income from peace remains constant as the stakes in the dispute rise, because if it backs down and abandons the tariffs it forgoes the productivity gain from them, so its income remains at the *status quo ante*. Its income from war also remains constant but for a different reason. When it stands firm in the dispute it retains the productivity gain from the tariffs; but when it goes to war it suffers from an unfavorable shift in relative army size (e.g. the tariffs reduce its optimal army size, but the spillover effect increases state 2's optimal army size). Consequently, the difference between state 1's peace and war incomes *remains constant* as the income stakes in the dispute rise.

State 2's critical deterrence threshold is displayed in the bottom right panel of figure 19.3. The x-axis represents the intensity of the international spillover effect on state 2 in terms of the percentage decrease in its civilian labor productivity from the foreign tariffs (causal assumption 2d). This is the productivity loss that state 2 would incur if it were to back down in the international dispute and acquiesce to state 1's demand to accept the tariffs and their spillover effects without interference. This axis thus represents the size of the stakes in the dispute for state 2. The y-axis represents state 2's long run incomes from peace and war. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 1. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 1. The income from peace declines steadily as the export losses mount and the domestic depression deepens. The income from war declines less steeply, though, because state state 2 goes to war it benefits from the favorable change in relative army size (e.g. the tariffs decrease state 1's army size decrease while the export losses increase state 2's army size). As a result, state 2's long run income from standing firm and going to war *never falls below* its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence state 2 has no critical deterrence threshold.

**State 1's peace-war income differential remains unchanged as the income stakes in the international dispute rise. Hence state 2 has no critical deterrence threshold.**

### 2.2.3 Observable implications: state 2's crisis bargaining and war initiation strategies

This analysis generates a number of observable implications for the historical evidence on state 2's strategies.

**Implication 12d.** State 2 will assess the income threat posed by the spillover problem slowly and patiently. Even if its income losses from the spillover effect become large, it will be slow and patient in assessing the problem.

**Implication 13d.** State 2 will adopt a flexible diplomatic position as to the concessions it wants state 1 to make to moderate the spillover effects. State 2 will maintain its flexible position even after its income losses from the spillover effect become large. State 2 will wait until its military preparations for war are complete before it issues firm demands for concessions from state 1 to moderate the spillover effect.

**Implication 14d.** State 2 will wait until its military preparations for war are complete before it issues military threats to back its demands.

**Implication 15d.** The purpose of state 2's military ultimatums will be to establish diplomatic pretexts for military actions it already decided to take.

**Implication 16d.** The duration of time from the emergence of the spillover problem to state 2's decision to initiate war will be comparatively long compared to the duration in a secession crisis or a revolution crisis.

**Implication 17d.** State 2's demand to stop generating the spillovers will be seen as militarily credible by state 1.

**Implication 18d.** State 2's military action will be sustainable without allies.

### 2.2.4 The impact of state 2's peace-war income differential on state 1's critical deterrence threshold

First consider how state 2's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle right panel of figure 19.3. The x-axis represents the intensity of the international spillover effect on state 2 in terms of the percentage decrease in its civilian labor productivity from the foreign tariffs (causal assumption 2d). This is the productivity loss that state 2 would incur if it were to back down in the international dispute and acquiesce to state 1's demand to accept the tariffs and their spillover effects without interference. This axis thus represents the size of

the stakes in the dispute for state 2. The y-axis represents state 2's single-period incomes from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state 1. The income from peace declines steadily as the stakes in the dispute rise because of its productivity loss from the international spillover effect. The income from war declines even more steeply as the stakes in the dispute rise. For not only does it suffer the productivity loss from the spillover effect, but it loses even more civilian production and income from transferring civilian workers into the military to raise its army size from  $A_2^*$  to  $A_{2d}^*$ . Consequently, the difference between state 2's peace and war incomes *widens* as the stakes in the dispute rise.

State 1's critical deterrence threshold is displayed in the bottom left panel of figure 19.3. The x-axis reflects the effectiveness of state 1's tariffs in terms of the percentage increase in its civilian labor productivity (causal assumption 1d). This is the productivity gain that state 1 would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the tariffs. This axis thus represents the size of the stakes in the dispute for state 1. The y-axis represents state 1's long run incomes from peace and war. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 2. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 1. The income from peace remains constant as the stakes in the dispute rise for the same reason that the single-period income remains constant. If state 1 backs down and abandons its tariff and lending policies it forgoes the productivity gain from the tariffs. The income from war increases steadily as the stakes in the dispute rise. For it retains the productivity gain from the tariffs and gains even more civilian production and income from transferring soldiers into the civilian economy to lower its army size from  $A_1^*$  to  $A_{1d}^*$ . As a result, state 1's long run income from standing firm and going to war *never falls below* its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence it has no critical deterrence threshold.

**State 2's peace-war income differential increases as the income stakes**

in the international dispute rise. Hence state 1 has no critical deterrence threshold.

### 2.2.5 Observable implications: state 1's crisis bargaining and war initiation strategies

Here are the observable implications for the historical evidence on state 1's strategies.

**Implication 19d.** State 2 will perceive state 1's potential opposition to the demand to stop generating the spillovers as militarily credible .

**Implication 20d.** If the moderates in state 1 propose to respond to state 2's demand with reassurances and accommodations, then the hardliners in state 1 will not accuse the moderates of caving in to foreign threats.

**Implication 21d.** State 1 will respond to state 2's demand with reassurances and accommodations.

**Implication 22d.** State 1 will not issue demands and deadlines backed by military threats until after state 2 initiates military hostilities.

**Implication 23d.** State 1's military action will be sustainable without allies.

## 2.3 Revolution Crisis

**Fact 1r.**<sup>7</sup> State 1 experiences a domestic political revolution that increases the lower class's share of the national income and it decreases the upper class's share. Its revolution generates ideological spillovers across international borders into state 2, thereby provoking a similar political movement there.

The revolution in state 1 *increases* the marginal productivity of its civilian and military labor in producing income for the lower class (which now controls the domestic political regime). These shifts in labor productivity are shown in the top left panel of figure 19.4. The x-axis represents state 1's allocation of labor between the civilian economic sector and the military-governance sector. The y-axis represents the marginal products of civilian and military labor in producing income for the lower class in state 1. The thin solid lines show

---

<sup>7</sup>Recall **Fact 1** from above: "State 1 develops a new domestic activity that increases its income but generates a negative spillover effect on state 2."

state 1's marginal products of labor before its domestic revolution. The thick solid lines show the impact of a 20% redistribution of national income from the upper class to lower class (e.g. the positive impact on the marginal products of labor in producing income for the lower class in state 1).

**Since the marginal products of civilian and military labor change by the same degree, the revolution leaves state 1's optimal army size *unchanged* (at  $A_1^*$ ).**

For example, in 1789, France experienced a domestic political revolution that increased its lower class's share of the national income at the expense of the upper class.

**Fact 2r.** State 1's revolution generates ideological spillovers into state 2 that provoke a revolutionary movement there, threatening to lower the upper class's share of the national income in state 2.

The revolutionary movement in state 2 *decreases* the marginal productivity of its civilian and military labor in producing income for the upper class in state 2 (which still controls the domestic political regime). These shifts in labor productivity are shown in the top right panel of figure 19.4. The x-axis represents state 2's allocation of labor between the civilian economic sector and the military-governance sector. The y-axis represents the marginal products of civilian and military labor in producing income for the upper class in state 2. The thin solid lines show state 2's marginal products of labor before the revolution erupts in state 1. The thick solid lines show the impact of a 20% redistribution of national income from the upper class to lower class (e.g. the negative impact on the marginal products of labor in producing income for the upper class in state 2). Since the marginal products of civilian and military labor change by the same degree, the revolution leaves state 2's optimal army size *unchanged*.

**Since the marginal products of civilian and military labor change by the same degree, the revolution leaves state 2's optimal army size *unchanged* (at  $A_2^*$ ).**

To continue with the same example, the French Revolution of 1789 generated ideological spillovers into Austria that provoked a political movement for economic redistribution there.

### 2.3.1 Observable implications of no change in the optimal army sizes

**Implication 7r.** State 2 will delay in issuing diplomatic threats until the revolution in state 1 appears to become radical.<sup>8</sup> Then state 2 will issue full-bore diplomatic threats (since its optimal army size is neither increased nor decreased by the spillover problem).

**Implication 8r.** State 2 will delay in taking military action until the revolution in state 1 appears to become radical.<sup>9</sup> Then it will take full-strength military action (since its optimal army size is neither increased nor decreased by the spillover problem).

**Implication 9r.** State 1 will not issue diplomatic threats until state 2 does. Then state 1 will respond with full-bore diplomatic threats (since its optimal army size is neither increased nor decreased by the spillover problem).

**Implication 10r.** Once state 2 takes military action, state 1 will respond with full strength military action (since its optimal army size is neither increased nor decreased by the spillover problem).

**Implication 11r.** Since state 1's optimal army size stays the same and state 2's optimal army size stays the same, the frontline in the war will be at state 2's borders.

### 2.3.2 The impact of state 1's peace-war income differential on state 2's critical deterrence threshold

First consider how state 1's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle left panel of figure 19.4. The x-axis reflects the degree of radicalization of state 1's domestic revolution in terms of the percentage increase in its labor productivity in producing income for the lower class (causal assumption 1r). This is the productivity gain that state 1's lower class would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the revolution. This axis thus represents the size of the stakes in the dispute for state 1. The y-axis represents the single-period incomes of state 1's lower class from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state 2. The income of state 1's lower class from peace remains constant as the stakes in the dispute rise. For

---

<sup>8</sup>See implication X below.

<sup>9</sup>See implication X below.

if it backs down and abandons the revolution it forgoes the productivity gain from it, so its income remains at the *status quo ante*. Its income from war increases as the stakes in the dispute rise, because when it stands firm it retains the income gains from the domestic revolution. Consequently, the difference between state 1's peace and war incomes *narrows* as the income stakes in the dispute rise. But it does not narrow as quickly as in the case of secession spillovers.

State 2's critical deterrence threshold is displayed in the bottom right panel of figure 19.4. The x-axis represents the intensity of the international spillover effect on state 2 in terms of the percentage decrease in its labor productivity from the revolution in state 1 (causal assumption 2r). This is the productivity loss that state 2's upper class would incur if it were to back down in the international dispute and acquiesce to state 1's demand to accept its domestic relation and the ideological spillovers without interference. This axis thus represents the size of the stakes in the dispute for state 2's upper class. The y-axis represents the long run incomes from peace and war for state 2's upper class. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 1. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 1. The income from peace declines steadily as the ideological spillovers intensify and the extent of domestic revolutionary unrest increases. The income from war declines just as steeply, because the spillovers do not change either state's optional army size. As a result, state 2's long run income from standing firm and going to war eventually falls below its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence state 2's critical deterrence threshold is *low* compared to the threshold in a depression crisis, but not as low as that in a secession crisis. State 2's threshold is shown by  $\bar{\Delta}_2$  in the figure.

**State 1's peace-war income differential decreases slowly as the income stakes in the international dispute rise. This trend gives state 2 a critical deterrence threshold that is lower then the threshold in a depression crisis but higher than the threshold and a succession crisis.**

### 2.3.3 Observable implications: state 2's crisis bargaining and war initiation strategies

This analysis generates a number of observable implications for the historical evidence on state 2's strategies.

**Implication 12r.** (a) State 2 will assess the income threat posed by the spillover problem slowly and patiently as long as it expects the domestic revolution in state 1 to remain moderate. (b) As soon as state 2 sees signs that state 1's revolution is radicalizing, state 2 will assess the income threat posed by the spillover problem quickly and impatiently.

**Implication 13r.** (a) Before it becomes clear whether the spillover problem is serious, state 2 will adopt a flexible diplomatic position on the concessions it wants from state 1 to moderate the domestic activity causing the spillover effects. (b) As soon as state 2 sees signs that state 1's revolution is radicalizing, state 2 will issue a firm demand that state 1 stop the domestic activity causing the spillover effects.

**Implication 14r.** (a) State 2 will not issue military threats to back its diplomatic position as long as it expects the domestic revolution in state 1 to remain moderate. (b) As soon as state 2 sees signs that state 1's revolution is radicalizing, state 2 will issue military threats to back its demands.

**Implication 15r.** The purpose of state 2's military ultimatums will be to elicit concessions from state 1 to stop the domestic activity causing the spillover effects.

**Implication 16r.** The duration of time from the emergence of the spillover problem to state 2's decision to initiate war will be intermediate (e.g. longer than the duration in a secession crisis, but shorter than the duration in a depression crisis).

**Implication 17s.** State 2's demand to stop generating the spillovers will be seen as militarily non-credible by state 1.

**Implication 18r.** State 2's military action will be unsustainable without allies.

### 2.3.4 The impact of state 2's peace-war income differential on state 1's critical deterrence threshold

Now consider how state 2's peace-war income differential changes as the income stakes in the international dispute rise. This trend is displayed in the middle right panel of figure 19.4. The x-axis reflects the degree of radicalization of state 1's domestic revolution and, consequently, the intensity of the ideological spillovers into state 2, in terms of the percentage decrease in its labor productivity in producing income for its upper class (causal assumption



2r). This is the productivity loss that state 2's upper class would *incur* if it were to back down in the international dispute and accept the ideological spillovers passively. This axis thus represents the size of the stakes in the dispute for state 2. The y-axis represents the single-period incomes of state 2's upper-class from peace and war. The thin solid line shows its single-period income from backing down in the international dispute and remaining at peace within its borders. The thick solid line shows its single-period income from standing firm in the international dispute and going to war against state 1. The income of state 2's upper-class from peace declines steadily as the stakes in the dispute rise because of its productivity loss from the international spillover effect. The income from war declines just as steadily as the stakes in the dispute rise, because the international spillover effect continues in this case as well. (Even though state 2 stands firm in its demand that state 1 moderate its domestic revolution, standing firm in this demand does not, in itself, actually moderate state 1's domestic revolution.) Consequently, the difference between state 1's peace and war incomes *remains constant* as the income stakes in the dispute rise.

State 1's critical deterrence threshold is displayed in the bottom left panel of figure 19.4. The x-axis reflects the degree of radicalization of state 1's domestic revolution in terms of the percentage increase in its labor productivity in producing income for the lower class (causal assumption 1r). This is the productivity gain that state 1's lower class would *forgo* if it were to back down in the international dispute and acquiesce to state 2's demand to abandon the revolution. This axis thus represents the size of the stakes in the dispute for state 1. The y-axis represents the long run incomes from peace and war for state 1's lower class. The thin solid line shows its infinite-horizon income from backing down in the international dispute and remaining at peace with state 2. The thick solid line shows its infinite-horizon income from standing firm in the international dispute and going to war against state 2. The income from peace remains constant as the stakes in the dispute rise for the same reason that the single-period income remains constant. If it backs down and abandons the revolution it forgoes the productivity gain from it, so its income remains at the *status quo ante*. The income from war increases as the stakes in the dispute rise, because when it stands firm it retains the income gains from the domestic revolution. As a result, state 1's long run income from standing firm and going to war never falls below its long run income from backing down and remaining at peace as the income stakes in the dispute rise. Hence it has no critical

deterrence threshold.

State 2's peace-war income differential increases as the income stakes in the international dispute rise. Hence state 1 has no critical deterrence threshold.

### 2.3.5 Observable implications: state 1's crisis bargaining and war initiation strategies

Here are the observable implications for the historical evidence on state 1's strategies.

**Implication 19r.** State 1's opposition to the demands will be seen as militarily credible by those in state 2 who believe state 1's revolution is moderate, but non-credible by those in state 2 who believe state 1's revolution is radical.

**Implication 20r.** If the moderates in state 1 propose to respond to state 2's demands with reassurances and accommodations, then the hardliners in state 1 will accuse the moderates of caving in to foreign threats.

**Implication 21r.** State 1 will respond to state 2's demands with confrontation and intransigence.

**Implication 22r.** State 1 will issue demands with deadlines backed by military threats as soon as state 2 perceives that state 1's revolution is radicalizing.

**Implication 23r.** State 1's military action will be unsustainable without allies.

## CHAPTER 20

### Case study: The outbreak of the Thirty Years War

The war is traditionally explained as a consequence of the Protestant Reformation. This explanation raises a number of questions. Why was there such a long delay from the start of the Protestant Reformation in 1517 to the outbreak of the war in 1618? Which aspect of the Reformation took so long to develop to the point that it caused a war? If this aspect was more or less developed by 1600, then why did none of the crises in Austria and Bohemia between 1600 and 1611 cause a war? What was so different about the crisis of 1618? Why couldn't the two sides in the 1618 dispute split the difference peacefully and thereby save the costs of war?

Previous analysts of the Thirty Years War have given many answers to these questions. Their answers are refuted by the historical evidence. The theory presented in the previous chapters provides new answers to these questions. In the following analysis I present the evidence that supports these answers as well as the evidence that refutes the answers given by previous analysts of the war.

In section 1, I summarize the spread of the Reformation into Austria and Bohemia and the rise of the Catholic counterreformation in response, paying particular attention to legal developments that gave each side the impression it was engaged in an "all-or-nothing" contest that put its survival at stake. In section 2, I analyze the crises in Austria and Bohemia between 1600 and 1611 to determine their similarities and differences in comparison to the crisis of 1618. This close analysis enables a number of previous explanations of the Thirty Years War to be refuted and provides support for the new explanation offered here. In section 3, I outline the legal developments and disputes after 1611 that eventually became causes of war in 1618. In section 4, I analyze the diplomatic crisis that began in March of 1618 and led to the outbreak of war in August of 1618, providing the evidence that corroborates the implications developed in the previous chapter. In section 5, I analyze the continued

diplomatic bargaining during the first year of the war to show that the evolution of each side's bargaining position supports the new explanation of the war offered here. In section 6, I address the question of why Austria chose to continue the war even after it became clear that she could not achieve a quick victory, providing evidence that refutes alternative rival hypotheses and supports the new explanation offered here.

## **1. The Protestant Reformation generates distributional conflict**

The seeds of the Thirty Years War were sown 100 years earlier with the outbreak of the Protestant Reformation. In 1517 Luther broke with the Catholic Church over spiritual issues and problems of church corruption. He inspired a "protestant" movement that competed with the Catholics for the allegiance of peasants, merchants, nobles and rulers. As the Protestant movement spread through Germany, it posed a threat not only to the Catholic religion but also to the material interests of Catholic nobles and rulers because it created distributional conflicts. Who would control church lands and institutions? Who would exercise jurisdiction over subjects and receive the profits of jurisdiction? How would political power be distributed within noble assemblies – and between assemblies and central governments? How would court patronage be distributed among the nobility?<sup>1</sup> To avoid losing in these distributional struggles, the Catholics tried to roll back the Protestants and eliminate them from Germany. From the 1520s to the 1550s, the opposing camps fought military battles for control of territories, governments and populations. The Habsburg government eventually ran out of money and stopped its military campaigns, leaving large territories in Germany with Protestant rulers, populations and military organizations for self-protection.

The opposing camps signed the Peace of Augsburg in 1555. The settlement gave each territorial ruler in Germany the right to choose the confession of his subjects: Catholicism or Lutheranism. The settlement provided terms for dividing church lands between the two confessions. It also provided for *toleration* – the fair treatment of subjects who practiced a different confession than their ruler. They had the right to emigrate and take their movable assets with them. If they stayed, they would be tolerated as long as they practiced their

---

<sup>1</sup>MacHardy 2003, p. 47.

religion in private. The settlement also gave rulers and clergy the right to convert from Catholicism to Lutheranism. The resulting period of peace lasted for decades. Yet the Augsburg settlement did not resolve all of the issues at stake. Over time local agreements on church lands and toleration began to unravel. Open conflict broke out on occasion when one side or the other saw violence as the only way to vindicate its rights.<sup>2</sup>

### **1.1 Reformation and counter-reformation in the Kingdom of Austria**

The Protestant movement also spread widely in Austria after 1517. The population was receptive to the new confessions while the Habsburg rulers displayed toleration and avoided the strict posture advocated by the hard-line Catholics – at least for the first 60 years. Refugees from the German conflicts entered Austria in large numbers in search of toleration. The spread of Protestantism in Austria produced little strife of the kind that had afflicted Germany.<sup>3</sup>

The Protestant confessions spread through Austria informally at first, in part because some of the clergy and masses still hoped for reconciliation with the Catholic Church. Local communities adopted Protestant preachings and practices without central direction or uniformity. By 1550 Austrian society became largely Protestant. The expansion gathered speed and institutional form as the new confessional churches continued to grow and their leaders saw a need to obtain formal recognition from territorial rulers.<sup>4</sup> The Habsburg rulers in Vienna and the provincial capitals remained Catholic, however, not only for religious reasons but for financial and political reasons. They received much financial support from the Catholic Habsburg government in Spain. And they had greater chances of election to the throne of the Holy Roman Empire – spanning Germany, Austria and northern Italy – if they were Catholic.<sup>5</sup> At the same time the Habsburg rulers remained highly dependent on the Protestant provincial nobilities for war finance to fight the Ottoman Turks. Hence

---

<sup>2</sup>MacHardy 2003: 58; Pages, 1970: 36-9; Steinberg, 1966: 22-23.

<sup>3</sup>Asch 1997: 48-51; Evans, 1979: 3-7, 13, 19-20, 25, 39-40; MacHardy 2003: 47-48, 51-51. "After the 1520s little religious violence broke out and serious persecution was very rare throughout the century, Ferdinand's attack on the Bohemian Brethren for a few years from 1547 being the clearest case." (Evans 1979, p. 13)

<sup>4</sup>Asch 1997: 48-51; Evans, 1979: 3-7, 13, 19-20, 25, 39-40; MacHardy 2003: 47-48, 51-51.

<sup>5</sup>MacHardy 2003: 52-53.

the Habsburgs had to tolerate the Protestants and grant them shares of the distributional prizes mentioned above. Indeed it was the Habsburg ruler in Vienna who, as Emperor of the Holy Roman Empire, brokered the Peace of Augsburg in 1555. In the 1560s and 1570s, the need for war finance forced the Habsburgs to make substantial concessions to the Protestant nobilities of Austria. Between 1568 and 1572 Emperor Maximilian II granted verbal guarantees of freedom of worship to the Protestants of Lower, Upper and Inner Austria. While these promises were not constitutionally binding, they reflected the humanist and ecumenical approach of the Habsburg dynasty as well as its lack of bargaining power.<sup>6</sup>

By 1580 ninety percent of the nobility of Lower Austria were Protestant – and this was the heartland of Habsburg power. The numbers were nearly as high elsewhere.<sup>7</sup> The Habsburgs' dilemma was particularly acute in the area of government administration in both Vienna and the provinces. The government relied on a corps of noble administrators drawn from the provinces who were increasingly Protestant. The Austrian Habsburg rulers saw the Protestant revolts in the Netherlands and France as warnings about the power of non-Catholic officials to paralyze a government if their demands were not met. If such officials seized the provincial centers of power, it would threaten the central government itself.<sup>8</sup>

With the death of Maximilian and the accession of Rudolph in 1576, the regime initiated a series of counterreformation campaigns designed to stem the tide.<sup>9</sup> The regime made it harder for Protestants to become government administrators and easier for Catholics to do so.<sup>10</sup> The regime made it more difficult for Protestants to enter the nobility. Church lands previously alienated to the Protestants were restored to the Catholic Church.<sup>11</sup> These campaigns produced only modest gains for the Catholics from 1578 to 1594.<sup>12</sup> Nonetheless they heightened the Protestant nobility's sense that their survival was threatened. It was

---

<sup>6</sup>Asch 1997: 48-51; Evans, 1979: 3-7, 13, 19-20, 25, 39-40; MacHardy 2003: 47-48, 51-51.

<sup>7</sup>MacHardy 2003: 53-57; 136-137, 149.

<sup>8</sup>Asch 1997: 53; MacHardy 2003: 57-58, 184-187.

<sup>9</sup>Evans, 1979: 42. "[The Counter-Reformers] tackled the task in bold fashion, adopting extreme postures and eschewing any truce. But there was, of course, much more to their approach than any mere calculation of tactics: they embodied a fresh spiritual mood which relished the challenge."

<sup>10</sup>Asch 1997: 53; MacHardy 2003: 57-58, 184-187.

<sup>11</sup>MacHardy 2003: 53-57; 136-137, 149.

<sup>12</sup>Evans, 1979: 43-5.

not only a matter of religious freedom but a matter of economic resources and the social status and political power needed to secure them.<sup>13</sup>

The most vigorous campaign to eliminate Protestants and restore the Catholic religion was launched in Inner Austria in 1598. Led by the Habsburg governor of the region, Archduke Ferdinand, the campaign re-catholicized the entire region.<sup>14</sup> When Ferdinand's campaign succeeded, the Catholics in the central government in Vienna wondered whether his method might be duplicated elsewhere in the Austrian Monarchy. Later they would see him as an attractive candidate for the succession to the Habsburg throne. Ferdinand's success in Inner Austria depended on circumstances unique to the region, however. Elsewhere in Austria the Habsburg rulers were more dependent on the Protestant nobilities and peasants – and both groups were more able to strike back in defense of their rights. Hence the Habsburgs could not go as far in their counterreformation campaigns without provoking open conflict. The regime began a campaign in upper and lower Austria in 1587. By 1595 the Protestant peasants of Upper Austria were in open revolt against both their Catholic Habsburg rulers (over religious issues) and their Protestant noble lords (over tax issues).<sup>15</sup> After the regime suppressed the revolt by force, it began more campaigns in Upper Austria and Lower Austria.<sup>16</sup> By 1604 the campaigns aroused such hostility that the Protestants of both regions united to stand firm and reject any further curtailments of their religious rights.<sup>17</sup> The Habsburg regime and its supporters were not willing or able to retreat, however. They hardened their position and looked for opportunities to advance it by legal means.

## **1.2 Reformation and counter-reformation in the Kingdom of Bohemia**

Developments followed a different course in Bohemia. Prior to 1526 it had been an independent kingdom. In the 1400s it had experienced a religious reformation of its own (decades before the Lutheran Reformation began). The followers of Jan Hus broke with the Catholic

---

<sup>13</sup>Asch 1997: 53; MacHardy 2003: 57-58, 184-187.

<sup>14</sup>Asch 1997: 51; Evans, 1979: 45; MacHardy 2003: 61-62; Maland 1980: 61; Mortimer 2015: 85-86; Polisensky, 1971: 51-2; Steinberg 1966: 36.

<sup>15</sup>MacHardy 2003: 62-63.

<sup>16</sup>Evans, 1979: 45.

<sup>17</sup>Mortimer 2015, pp. 90-92.

Church over doctrinal disputes and established a distinct confession that was eventually tolerated by the traditional Catholic clergy and masses. The "Hussite" movement opened the doors of religious education to the population at large, "widening [their] mental and physical horizons" and giving them a greater chance of "playing a part in all departments of public life".<sup>18</sup> The religious toleration policies of the kingdom's government and the Catholic Church enabled cooperation across confessional lines to extend beyond the religious sphere to the economy, political system and society at large. By 1500 a majority of the population had adopted the new *Utraquist* confession of the Hussite movement and the social fabric of Bohemia was strong enough to bear religious difference and doctrinal disagreement without conflict.<sup>19</sup>

In 1526 the noble estates of Bohemia elected Ferdinand I, Habsburg king of Austria, to be the king of Bohemia as well. They elected Ferdinand primarily for international reasons: to combine the military power of the two kingdoms as a stronger bulwark against the Turkish empire to the east and the French kingdom to the West. Bohemia retained its status as a kingdom with its own monarchy and government. The Habsburg regime in Prague was headed by the Grand Chancellor of Bohemia who ruled in the king's name and was appointed for life. The Chancellor was accountable not only to the king, however, but also to the noble estates, because their political rights vis-a-vis the government were guaranteed by ancient customary law. While their assembly, the Bohemian diet, needed royal approval to convene, the king needed the diet's approval to raise taxes and issue certain kinds of laws.<sup>20</sup>

The exact *extent* of the diet's right to withhold approval would prove to be a critical area of dispute in the crisis of 1618. The ambiguity arose from a fundamental tension in the relationship between the two kingdoms. The advent of Habsburg rule in Bohemia in 1526 created the appearance that Bohemia had lost its sovereignty and been subsumed in

---

<sup>18</sup>Polisensky, 1971: 23-4.

<sup>19</sup>Polisensky, 1971: 23-4. "The fact of religious tolerance and its consequences for society had become so characteristic of life in the Czech lands, and were so rooted in all sections of the population, that even the changes which followed the linking of the Bohemian state with the Alpine lands of the Habsburgs [in 1526] left the new ruler no opportunity to upset the balance, at least in the first years of his reign. [Cooperation across confessional lines] became uniformly accepted by the beginning of the 16th century and was a firm foundation... for the maintenance of religious peace and calm furthering of domestic affairs of state." (Polisensky, 1971: 24.)

<sup>20</sup>Pages, 1970: 33, 42; Polisensky, 1971: 25.



the Austrian orbit. Yet the condominium of the two kingdoms was intended simply to enable military power aggregation by formalizing the Bohemians' obligations to pay taxes and provide forces for the joint defense – so the problems of alliance burden sharing and unity of command would not have to be solved anew with each foreign military threat. The Bohemian nobles selected the Habsburg Dynasty to play this sovereign role because they expected to benefit from the increased security. The king's authority to raise defense taxes and command the army was amalgamated with his other authorities, however, which benefited the Bohemian nobles less (or even harmed them). Their rights to control the kingdom's government and administration in domestic matters went back centuries, so they were no more inclined to cede these ancient rights to Ferdinand of Habsburg after 1526 than to any previous native king. In this sense their territories remained an independent and sovereign kingdom despite their institutionalized military alliance with Austria and the veneer of suzerain authority it layered over their domestic sovereignty.

The Habsburg king's regime in Prague, operating through the Bohemian Chancery, presided over all five provinces of the kingdom (Bohemia, Silesia, Moravia and Upper and Lower Lusatia). Each province retained its own government administration and noble diet whose members had the rights to hold administrative office, make policies and withhold approval of royal initiatives.<sup>21</sup> The Bohemian Diet believed its role was to preserve the relationship between king and subjects (noble and commoner alike) by preventing the king from adopting measures that would harm the subjects' interests. The Diet's instruments in this task included informing the king of measures that were provoking adverse reactions and maintaining an independent military capability to defend the subjects' rights by force if necessary. Between 1527 and 1611 the Bohemian diet defeated a series of royal proposals to change fiscal and administrative practices on the grounds that they would undermine noble rights. The most important victory concerned the Royal Bohemian Exchequer which Ferdinand created in 1527. He and his successors – Maximilian and Rudolph – tried repeatedly over the next 80 years to transfer the fiscal authority of noble officeholders to the Exchequer. The Diet prevented such reforms in large part to ensure that the Exchequer's authority would not surpass that of the Bohemian Chancery – because the Chancery was obligated

---

<sup>21</sup>Polisensky, 1971.

to respect the nobles' rights and could be held accountable by the diet.<sup>22</sup> The exact *extent* of the diet's right to hold the Chancery accountable was not clear, however, and this would also become a critical point of dispute in the crisis of 1618.

With this political and institutional background in mind, it is possible to understand how the outbreak of the Lutheran Reformation in 1517 affected Bohemia. The new "protestant" teachings and practices made their way into the kingdom but caused little trouble for the first three decades. The long-standing norms of mutual toleration and normal social intercourse between the Catholic minority and the non-conformist majority remained strong. When the Catholic camp in Germany scored military victories against the Protestant camp there in the 1540s, it emboldened the Bohemian King, Ferdinand I, to conduct a persecution campaign against the Bohemian Protestants from 1547 to 1552. But as the Catholic camp in Germany faltered and settled for peace in 1555, so Ferdinand was also forced to recognize that his regime simply lacked the power to suppress Protestantism in Bohemia. Over the course of the 1500s the majority of the Bohemian nobility and peasants became Protestant. The remaining Catholic clergy were not zealous to pursue counterreformation campaigns against them. Nor was the king in a position to promote the counterreformation in Bohemia, because he needed the financial support of the Bohemian noble estates to pay an annual tribute to the Turks (agreed at the Peace of Constantinople in 1562). Indeed the new bargaining power of the Bohemian Protestants enabled them to extract a formal recognition of their confession from King Maximilian in 1575.<sup>23</sup>

In the absence of domestic proponents, the counterreformation in Bohemia was initiated by foreigners. A set of Catholic clergy from Germany and Italy immigrated to the kingdom in the 1550s and began the process of reviving the Church organization and converting the nobles and peasants back to Catholicism. Their efforts produced little progress at first because the Habsburg regime played a relatively passive role and the Bohemian traditions of nonconformity (to Catholic doctrine) and toleration of nonconformity ran deep. In the 1580s the native Catholic magnates joined with the Spanish ambassador and the Papal nuncio to form a "Spanish party" with the goals of capturing control of the Bohemian government from

---

<sup>22</sup>Polišenský 1971a, pp. 27–29, 47.

<sup>23</sup>Asch 1997: 48-49; Evans, 1979: 47; Polisensky, 1971: 27-29, 46-9.

the Protestant nobility and advancing the counterreformation. The party's leader became Chancellor in 1599 and initiated a series of aggressive measures – centralizing control of the administrative agencies at the expense of the upper nobility, giving top government offices to members of the Spanish party, convincing Rudolf to dismiss his Protestant advisors and employ only Catholic ones, and reviving restrictions on the Bohemian Brethren (one of the Protestant confessions). In a portent of reactions to come, one of the Protestant leaders urged the Bohemian Protestant estates to withhold tax payments from the government to force it to moderate some of these measures.<sup>24</sup>

The counter Reformation movement became more effective throughout the Austrian Monarchy after 1600. The clergy and their noble supporters became more organized and determined. The Catholic nobles, although a minority of the nobility as a whole, formed their own Catholic estates to promote the counterreformation along with the clergy. This in turn induced the Protestant nobility to form their own estates.<sup>25</sup> Caught between the two sides, Emperor Matthias faced a dilemma. On one hand, his natural inclination was to side with his co-religionists and the Roman Church. On the other hand, the realities of power and material interest often dictated that he play the role of neutral arbiter, peacemaker or even ally of the Protestants in showing the Catholic minority that it simply lacked the power to force the Protestants to capitulate. Matthias oscillated repeatedly between these two poles.

## **2. Crises and settlements that set precedents for 1618**

Between 1605 and 1611, five crises erupted in the Habsburg lands due to frictions between the Catholic Habsburg government and the Protestant provincial nobilities. Four of the crises were settled without any fighting while one crisis was settled after a year of fighting. The subsequent crisis of 1618, which erupted into the Thirty Years War, was remarkably similar to these crises. The question naturally arises: Why did the crisis of 1618 end so differently? The answer becomes clear by comparing the six crises in detail in light of the

---

<sup>24</sup>Evans, 1979: 41, 47-8. Pages, 1970: 42. Polisensky, 1971: 13, 49, 51, 66-7. Steinberg, 1966: 34.

<sup>25</sup>Evans, 1979: 63-5.

theory presented in the previous chapters.<sup>26</sup>

## 2.1 The Hungarian crisis and settlements of 1606

The Kingdom of Austria had long acted as a bulwark to defend Europe from the Ottoman Turks to the east. The border line between Austria and the Turks lay in the Kingdom of Hungary. The Turks controlled the eastern part of the kingdom while the western part was a province of the Austrian Monarchy.<sup>27</sup> The Austrian government still paid tribute to the Turks for the province, however. In 1593 King Rudolf of Austria declared war on the Turks to end the tributary arrangement and establish full Austrian sovereignty over the province. The Austrian effort was hampered by disputes among the provincial nobilities composing the coalition. After ten years of inconclusive fighting against the Turks, the Catholic nobility of Hungary, together with Rudolf's Catholic advisers and the Spanish party in Prague, convinced him to launch a military campaign against the Hungarian Protestants. He agreed to the plan for political reasons rather than the religious purpose of persecuting Protestants. He simply wanted to restore Austrian sovereignty over Hungary. It provoked a strong reaction.<sup>28</sup>

The Hungarian Protestants launched a military counter-offensive in alliance with Transylvania and the Turks in 1605. They moved westward across Hungary and into Moravia, a province of the neighboring Kingdom of Bohemia. They pressed the Moravian Protestants to join their resistance to Habsburg religious persecution – and hoped that Protestants elsewhere in the Kingdom of Bohemia would join too. However, the Moravian Protestants sided with the Moravian Catholics to fight off the "foreign" incursion. Together they called for military support from the other main provinces of the Bohemian kingdom, and the joint force, under Rudolf's suzerainty, fought the Hungarian Protestants to a standstill. The Bohemian position was untenable, however, because Rudolf needed more tax revenues to fund

---

<sup>26</sup>Mortimer (2015) provides many of the essential facts on the six crises, but does not analyze the similarities between the crises or ask why the crisis of 1618 ended differently despite the similarities. My conclusion about the cause of the crisis of 1618 and the war that followed differs substantially from Mortimer's conclusion, as noted below.

<sup>27</sup>Pages 1970, pp. 23–4, 33–4.

<sup>28</sup>Evans, 1979: 51–52; Maland 1980: 18–19; Mortimer 2015: 90, 93–94; Pages, 1970: 33–4; Polisensky, 1971: 14, 67.

the force yet the Austrian Protestant Estates refused to grant them. After a year of inconclusive warfare, his forces were at the point of rebellion for lack of pay and Hungary was at the point of rebellion against the oppressive rule of the Habsburgs. Rudolf's brother, Archduke Matthias, was more concerned about the threat posed by the Turks, the Hungarians' ally. He gathered a family council to intervene and give him authority to negotiate a peace agreement.<sup>29</sup>

Matthias reached a settlement with the Hungarian Protestants in June 1606. The accord granted them religious freedom and the political safeguards needed to secure it. Habsburg authority in Hungary was transferred from Rudolph to Matthias. Day-to-day control of the royal government would be exercised by a Hungarian noble elected by the Hungarian Estates. Offices in the royal government would be reserved for Hungarians. The Jesuits would be expelled from the kingdom. And a peace settlement would be reached with the Turks.<sup>30</sup>

## **2.2 The Bohemian crisis and settlements of 1607-8**

Within a year, Rudolf started planning an expedition to overturn these settlements. Foreseeing another disaster for the House of Habsburg, Matthias gathered the Estates of Hungary, upper Austria and lower Austria – consisting mostly of Protestant nobles – and secured their agreement to provide military support to defend the Hungarian settlements. His alliance was joined by Moravia, a province of the Kingdom of Bohemia ruled by Rudolf, when leaders of its upper nobility staged a provincial coup in the Moravian capital against the ruling Catholic clique. Their tactics were eerily similar to those that the Bohemians would use in the crisis of 1618. The wealthiest Catholic noble in the province committed to fund troops for Matthias' alliance, and then "storm[ed] into the Imperial governor's palace with a following of like-minded noblemen and turn[ed the provincial governor] out of office." The coup leader was then appointed head of a new provincial government by an emergency meeting of the Moravian noble Estates.<sup>31</sup>

---

<sup>29</sup>Mortimer, 2015: 95-6. Polisensky, 1971: 66-9, 71.

<sup>30</sup>Evans, 1979: 52. Maland, 1980 :19. Mortimer, 2015: 96-7. Polisensky, 1971: 73. Steinberg, 1966, 34.

<sup>31</sup>Mortimer, 2015: 97-100. Polisensky, 1971: 75.

Mathias, supported by his large alliance of Habsburg lands, invaded the province of Bohemia and approached the seat of Rudolph's power in Prague. The goal was either to obtain Rudolph's acceptance of the Hungarian settlements or to depose him from the throne of the Bohemian kingdom. Rudolph could not stand up to Matthias without political and military support from the noble Estates of Bohemia. Since the nobles were mostly Protestant, however, they opposed Rudolph's goal of overturning the Hungarian settlements – and resented his previous efforts to restrict the Protestant religion in Bohemia. "Well aware that they had the option of switching their support to Matthias, they agreed to concede nothing to Rudolph until he accepted their list of demands."<sup>32</sup> Under this pressure Rudolph agreed to abandon his planned expedition and accept the Hungarian settlements of 1606 in return for continued political and military support from the Bohemian Estates. He was not prepared to concede religious freedom to the Bohemian Protestants on any terms, however.<sup>33</sup>

The Protestants wanted not only freedom of worship but political and institutional safeguards to secure this freedom. They wanted the rights: to elect "defenders" to protect the rights of Protestants, to control the government agency for religious affairs beyond the Catholic Church, to reserve all government offices for Bohemians, to reserve half of all government offices for Protestants, and to veto property acquisitions by the Jesuit Society (a branch of the Catholic clergy). When the Protestant estate members sought to present these demands to Rudolph, he tried to limit the size of their delegation. They ignored him and brought all of their membership to the palace to hear the leader read their demands to him.<sup>34</sup>

Rudolph was determined not to concede any of the demands – and the Bohemian Estates wanted Matthias' invasion force to leave the province as soon as possible. So they agreed to defer the negotiations over these issues for a few months (see below). In the meanwhile they reached an agreement directly with Matthias to increase his authority in the Habsburg lands comprising his alliance (at Rudolph's expense). The accord, struck in June of 1608, elevated

---

<sup>32</sup>Mortimer, 2015: 101.

<sup>33</sup>Mortimer, 2015: 100-1. Polisensky, 1971: 75.

<sup>34</sup>Mortimer, 2015: 101-2.

Mathias to the positions of King of Hungary, Margrave of Moravia and ruling prince of upper and lower Austria – leaving Rudolf only with authority in the remaining provinces of the Kingdom of Bohemia. Even these provinces nominated Mathias to be Rudolf's successor as King of Bohemia. In return for this "free election" as the next king, Matthias committed to respect the "existing" rights and privileges of the Bohemian Estates. He also gave the Moravians verbal promises to respect their religious freedom.<sup>35</sup>

It is worth summarizing the key elements of this crisis, because nearly all of them would recur in the crisis of 1618.

1. The top official in a province of the Kingdom of Bohemia was thrown out of office by force and replaced by a rival.
2. The new officer took control of the government and entered an alliance with a military opponent of the ruling King of Austria. (The king was Rudolf, the opponent was Mathias.)
3. The opponent's military alliance included upper and lower Austria, the homelands of the Austrian king.
4. Facing such an overpowering alliance, the King sought military support from the noble Estates of his homeland, the large majority of which were Protestant.
5. In return for their support, the Protestant estates demanded concessions on religious issues and threatened to switch their allegiance from the ruling king to his opponent if he did not meet the demands.
6. To obtain their support the king met some but not all of the demands.

There were three key differences between the crisis of 1608 and the crisis of 1618. In 1608 the provincial coup did not change the distribution of constitutional authority between the provincial nobility and their royal overlord. None of the other provinces in the Kingdoms of Bohemia or Austria experienced a similar coup or change in the distribution of constitutional authority.<sup>36</sup> And the Austrian king did not appeal to Spain for military support. In 1618, as we will see, the provincial coup changed the distribution of constitutional authority between

---

<sup>35</sup>Mortimer, 2015: 102. Polisensky, 1971: 75, 77. Steinberg, 1966, 34.

<sup>36</sup>Even though they joined the alliance against the king immediately.

nobility and king and threatened to inspire similar changes in other provinces.<sup>37</sup> And the Austrian king appealed to Spain for military support.

The king backed down in 1608 but stood firm in 1618. The question naturally arises: Did he stand firm in 1618 because of the change in the distribution of authority in one province – and the threat that it would spread to other provinces – or the expectation of military support from Spain?

### **2.3 The Austrian crisis and settlement of 1609**

The Protestant Estates in Matthias' lands – Hungary, Moravia, upper and lower Austria – now made a secret agreement to withhold their oath of allegiance from him until he committed to respect their religious freedoms. They claimed they had a customary right to confirmation of their rights and privileges before their recognition of a new ruler. So they demanded clarification of the previous concessions granted by Maximilian four decades earlier to preclude the regime from violating them further.<sup>38</sup> Matthias split this alliance by offering the Moravian and Hungarian Estates stronger guarantees than he was willing to offer the Austrian Estates. The Moravian Estates caved in and gave Mathias their oath of allegiance for two reasons: Matthias committed to allow day-to-day control of the Moravian government to be exercised by a Moravian noble, and the Moravians needed a powerful sovereign overlord to defend the province from "foreign" military attacks.<sup>39</sup>

The Protestant nobilities of upper and lower Austria continued to withhold their oath of allegiance until their demands for religious freedom and political guarantees were met. They presented the demands to Matthias at the Estates meeting of September 1608 in Vienna – which included both Protestants and Catholics. After he gave vague assurances and insisted that their oath of allegiance should come first, the Protestant members, who formed a large majority, closed the meeting and established their own Estates – separate from the Catholics – in a different town. The newly formed Protestant Estates of upper and lower Austria then created a Confederation, formalized by the signing of a defensive

---

<sup>37</sup>Even though they remained loyal to the king for a year after the crisis erupted.

<sup>38</sup>MacHardy, 2003: 65.

<sup>39</sup>Evans, 1979: 52. Mortimer, 2015: 103.



alliance, and prepared to defend their religious rights by force.<sup>40</sup> While the members from upper Austria had already called up their provincial militia and started recruiting troops, now the members from lower Austria followed suit. Matthias and his chief minister perceived a threat that the Austrian Protestants would withdraw from the Austrian monarchy and establish a republic with military support from the Bohemian Protestants (who were still under Rudolf's authority).<sup>41</sup> To meet this military threat from within, Matthias obtained troops from Hungary (by making further concessions to the Hungarian Protestants) and issued a hard deadline to the Austrian Protestants to give their oath of allegiance. He argued that Maximilian had granted the previous concessions four decades earlier on the condition that they not harm the Catholic religion, yet the clarifications and concessions now demanded by the Protestants would harm it.<sup>42</sup> He also claimed that their demands on religious issues were "private matters" that they had no right to raise in a negotiation over their terms for giving the oath of allegiance to a new ruler.<sup>43</sup>

Although each side was prepared to go to war, neither side had any other allies (beyond Bohemia or Hungary). So they reached a compromise agreement in March of 1609. Matthias committed to respect the Protestants' existing religious rights codified in previous agreements with Maximilian in the 1570s; to grant new religious rights beyond those granted by Maximilian; to allow the Protestants a share of the royal government offices; and to create an impartial court to settle all legal disputes over the distribution of church land between the two confessions (Catholicism and Lutheranism). In return the Austrian Protestants recognized Matthias' succession as the king of Austria and gave their oath of allegiance. Notably, the Protestants had demanded confessional *parity* in appointments to all government offices, but Matthias only promised that he would give the important offices to the nobility (rather than others) without regard to confession (rather than preferentially to Catholics).<sup>44</sup>

---

<sup>40</sup>MacHardy, 2003: 65.

<sup>41</sup>MacHardy, 2003: 65.

<sup>42</sup>MacHardy, 2003: 65.

<sup>43</sup>Evans, 1979: 52. MacHardy, 2003: 72. Mortimer 2015: 103-4. Steinberg, 1966, 34.

<sup>44</sup>MacHardy, 2003: 65.

## 2.4 The Bohemian crisis and religious constitution of 1609

The noble Estates of Bohemia convened in January of 1609 to negotiate with Rudolf over the issues of religious freedom and political guarantees. By May they had failed to extract any concessions from him, in part because the "Spanish party" in Prague was pressing him to stand firm. He then prohibited further meetings of the noble Estates on this issue. In response the Protestant nobles convened their own Estates meeting – separate from the Catholics. When Rudolf prohibited such meetings and showed signs that he was preparing to suppress the meetings by force, they stood firm and prepared to defend their right to assemble by force. Then they presented Rudolf with a new declaration that affirmed their loyalty to him and promised their demands were limited. In return he permitted a new meeting of the Estates to discuss the religious issues. When the meeting came, however, he made no concessions.<sup>45</sup>

Stiffening their resolve, the Protestant Estates gave Rudolf a draft law codifying their demands together with a declaration that they intended to arm and defend themselves against any violators of their religious rights. They made a plan to appoint a governing council of their own, call up the militia and raise taxes to support it. They also received signals from two other provinces of the Kingdom of Bohemia – Silesia and Moravia – indicating they would help in the effort to defend Protestant religious freedom. In response to this threat Rudolph agreed to grant some but not all of the Bohemian Protestants' demands. When the hard-liners among them stood firm on the remaining demands, however, Rudolf withdrew even these concessions. In response the Protestants appointed thirty Directors, called up the militia, and started raising funds to support it. The Directors set up a provisional government in Prague to manage the growing military force, maintain order in the city, and provide defense against any attack by Rudolf's forces or the Catholic nobility. Under this pressure Rudolf granted the Protestants of the provinces of Bohemia and Silesia a written charter of their rights: the *Letter of Majesty* (1609).<sup>46</sup>

The main clause of the charter awarded all Protestants in Bohemia and Silesia the right to

---

<sup>45</sup>Mortimer 2015: 105-7. Polisensky, 1971: 49-50.

<sup>46</sup>MacHardy 2003: 65-66. Maland 1980: 20-21. Mortimer, 2015: 107-9. Steinberg, 1966: 34-35. Each province received its own *Letter of Majesty*, but the two charters were similar.

practice the Bohemian Confession (an amalgam of the two main Protestant denominations) and to be free of coercion to convert to Catholicism by any overlord. A second clause gave each confession the right to exercise control of its church lands and revenues. A third clause granted Protestants living on royal domains the right to build churches on these domains. This included both the royal agricultural Estates and the royal cities. The Protestants interpreted this clause to include the ecclesiastical domains as well, since kings of Bohemia had always exercised manorial rights over the ecclesiastical domains and raised revenues from them in time of need. Moreover, the king's manorial rights over the ecclesiastical domains had been affirmed by his own Chancellor in Prague just a few years earlier. So it seemed clear that this clause of the *Letter of Majesty* permitted the construction of Protestant churches on ecclesiastical lands.<sup>47</sup> However, "Catholics and the [Habsburg] dynasty denied this claim, on the grounds that Church land was only protected, not owned, by the crown."<sup>48</sup> This ambiguity would provoke the disputes that eventually caused the Thirty Years War.

The charter also recognized the need for political mechanisms to manage religious disputes between the Protestant and Catholic Estates of Bohemia. Each camp was guaranteed the right to elect and convene its own confessional Diet without the king's approval.<sup>49</sup> If disputes arose between Catholics and Protestants, each Diet had the right to appoint a set of "defenders" to advocate for its constituents' religious rights and negotiate with the other Diet's "defenders". Each set of defenders had the rights to convene its confessional Diet for consultations on how to handle a crisis and appeal to the king in intractable disputes. In this event the king would submit the dispute to an arbitration panel having an even balance of Catholic and Protestant members. Finally, in an ancillary agreement between the Catholic and Protestant diets, each Diet committed that its constituents would respect the lands, churches and religious practices of the other's constituents.<sup>50</sup>

In the view of most Czech historians, the *Letter of Majesty* gave the Bohemian Protestants more rights than the Augsburg settlement of 1555 gave Protestants elsewhere in the Holy

---

<sup>47</sup>Bireley, 2014: 91. Gindely, 1884:43-44. Steinberg, 1966: 34.

<sup>48</sup>Evans, 1979: 66-67.

<sup>49</sup>Previously the diet needed royal approval to convene (as mentioned above).

<sup>50</sup>Gindely, 1884:45. Mortimer, 2015: 134-5. Pages, 1970: 44-45. Polisensky, 1971: 50. Steinberg, 1966: 34.

Roman Empire.<sup>51</sup> The charter codified the Bohemian traditions of toleration in religious matters, dispute settlement by negotiation rather than conflict in political matters, and equal partnership between monarch and nobility in the constitutional sphere.<sup>52</sup> While the Catholics accepted the *Letter of Majesty* on the surface, the most powerful ones, including the king's officers in Prague, opposed it and looked forward to an opportunity to overturn it.<sup>53</sup>

## 2.5 The Bohemian crisis and succession accord of 1611

Two years later Rudolf, the King of Bohemia, tried to regain authority over the lands he had lost to Matthias in 1609. He contracted a military force of foreign mercenaries and allowed it to enter the province of Bohemia, presumably in preparation to advance into Matthias' lands to the east. The force was commanded by Rudolf's cousin Leopold, who may have been planning to coerce the Bohemian noble Estates into nominating him as Rudolph's successor. The threat was perceived immediately by Matthias and the Habsburg lands under his authority – and by the Bohemian Estates. Some of the Estates leaders suspected the Prague Spanish party of promoting the military incursion. The Estates asked their sovereign Rudolf for authorization to raise a professional army to defend their province from the force *whose invasion he had authorized*. After he denied the request, they started recruitment and taxation on their own authority. Rudolph *then* approved their actions and allowed them to call up the native Bohemian militia as well. Amazingly, he also asked for funds to pay the force of invading foreign mercenaries (whether to retain them for his own protection or to induce them to disband and go home it was not clear). After a few weeks of guarded standoff around Prague, the Bohemian noble Estates elected a council of thirty Directors and issued a declaration that their actions were justified, they remained loyal to Rudolf, and they were acting in his name. A week later a large military force sent from Moravia by Matthias arrived at Prague.<sup>54</sup>

---

<sup>51</sup>Pages (1970: 45) summarizes the historiography in this way.

<sup>52</sup>Polisensky, 1971: 50, 79-80.

<sup>53</sup>MacHardy 2003, pp. 65–66; Maland 1980, pp. 20–21; Pagès 1970, p. 45; Polišíenský 1971a, p. 76.

<sup>54</sup>Mortimer 2015: 110-15. Polisensky, 1971: 76.

In the end Rudolf was forced to cede his authority over the provinces of Bohemia and Silesia to Matthias. The manner in which royal authority was transferred set important precedents for the succession of Ferdinand in 1617 and the crisis of 1618. The Bohemian Estates *invited* Matthias to Prague, indicating that they both sought and approved the crown's transfer from Rudolf to Matthias.<sup>55</sup> Equally crucial, the *Letter of Majesty* remained in effect after the transfer of royal authority.

### 3. Disputes within Bohemia that would provoke the crisis of 1618

Soon questions about the rights of Bohemian Protestants to build churches and choose pastors led to disputes. The Protestants on the ecclesiastical domain of *Braunau*, exercising their rights under the *Letter of Majesty*, built a church there. Their defenders issued an opinion that the domain was certainly part of the royal domains, because "the Emperor had frequently mortgaged church possessions or pledged them as securities."<sup>56</sup> The abbot disagreed, however, and when the issue was brought to Matthias he ruled against the Protestants. They rejected his ruling and refused to turn over the key to the church. Then the Archbishop of Prague declared a church built by Protestants on the ecclesiastical domain of *Klostergrab* to be illegal. Perhaps in anticipation of similar opposition the Archbishop boarded up the church, demanded that the subjects convert to the Catholic confession, and threatened to punish them if they attended Protestant services in the region. Then Matthias decreed that all vacant pastor positions on the royal domains would be filled by the Archbishop of Prague, a Catholic. Since the Archbishop would eventually replace all Protestant pastors with Catholic ones, this decree essentially eliminated the right to practice the Protestant confession guaranteed by the *Letter of Majesty*.<sup>57</sup>

The Protestant Estates in Bohemia tolerated these abuses for years. But by 1614 they concluded that the best solution in the long run would be to replace the Catholic Habsburg

---

<sup>55</sup>Maland 1980: 60. Mortimer 2015: 116-7. Steinberg 1966, 35.

<sup>56</sup>Pages, 1970: 50-51.

<sup>57</sup>Evans, 1979: 63. Gindely, 1884:44-45. Mortimer, 2015: 134. One historian suggests that if the Protestants had been more skilled at crafting legal agreements, they could have avoided such conflicts with the Catholics. "The Protestants of the time showed their complete incapacity for the framing of important and comprehensive laws." (Gindely and Ten Brook 1884, p. 46)

Dynasty with a Protestant one at the next available legal opportunity. In the meanwhile they took advantage of Matthias' need for new tax revenues to fund more defenses against the Turks to pressure him to convene a general diet in Prague. They lodged their complaints at the Diet in 1615, but by this point Mathias and his counselors were determined to promote the Catholic counter-reformation regardless of previous commitments. Matthias rejected the complaints and yet the diet still had to approve the new tax revenues and accept liability for another tranche of government debts. This outcome showed that Habsburg rulers were no longer granting new religious concessions – or even just compliance with previous concessions – in return for financial support for military defense.<sup>58</sup> Acting through his Chancellor in Prague, Matthias then took further measures that directly violated the *Letter of Majesty*. He required the residents of royal domains to convert to Catholicism and the magistrates of royal cities to admit Catholics to borough citizenship. He also required royal cities to allow their city counselors to be appointed by the Royal Chamberlain – knowing that he would only appoint Catholics.

### 3.1 A counter-reformation regime assumes control in Bohemia

As Matthias approached death in 1617, he bequeathed the thrones of Austria and Bohemia to his cousin, Archduke Ferdinand of Styria. The prospect of Ferdinand's succession as King of Bohemia posed a serious threat to the Protestant Estates of Bohemia, since he had been a staunch promoter of the counterreformation in Styria. As a condition of electing him, therefore, they required him to reaffirm the *Letter of Majesty*. His Jesuit (Catholic) counselors advised that while it would have been wrong for him to *issue* such a charter, once it had been issued it would be right for him to *reaffirm* it. Accordingly Ferdinand accepted the Protestants' condition – at least on the surface. They also required him to commit that he would not take control of the Bohemian government until Matthias died.<sup>59</sup>

Despite these commitments, Ferdinand and his advisers and officials soon looked for opportunities to get around the *Letter of Majesty*. Indeed some of them had never accepted

---

<sup>58</sup>Polisensky, 1971: 77-8, 85.

<sup>59</sup>Asch 1997: 52-54; Gindely, 1884:42; Limm 1984: 14; Maland 1980: 21, 61-62; Mortimer, 2015:130; Pages 1970: 43-49; Steinberg 1966: 34-36.

the charter in the first place.<sup>60</sup> In the cities and towns they packed the ruling councils with Catholics even where the majority of the population was Protestant. They granted new borough citizenships only to Catholics. They reduced borough autonomy over local issues – both religious and nonreligious – by requiring borough councils to obtain royal government approval of decisions. And they instituted censorship of publications attacking the counterreformation. In the countryside they began a campaign to deny the Protestant clergy the salaries they customarily received from local parish revenues.<sup>61</sup> Following their lead, the Catholic nobility in the countryside pressured subjects to convert to Catholicism.<sup>62</sup>

### 3.2 Stalemate: Each side sees capitulation as ruinous and violence as necessary

In December 1617, the Habsburg royal court moved from Prague back to Vienna, its traditional base. Emperor Matthias, his Chancellor and King Ferdinand "turned control over to a largely Catholic Regency Council *with instructions to provoke the [Protestant] Estates*."<sup>63</sup> At this point the disputes over the Protestant churches at *Braunau* and *Klostergrab* resurfaced in a way that heightened tensions throughout the kingdom. The Protestants at *Braunau* had been ordered to surrender their church years earlier, but refused. They were ordered to surrender it again and refused. After their representatives were arrested and imprisoned, they still refused. Mathias sent a royal commission to persuade them but they stood firm. The commission tried to seize the church, but "was barred by a mob wielding sticks and stones".<sup>64</sup> Although the Protestants retained control of the church, their representatives remained in prison.<sup>65</sup> Each side could claim a victory of sorts from standing firm – but had to accept a loss of possibly greater value because the other side had stood firm too.

Perhaps to avoid similar stalemates, the Archbishop of Prague handled the Protestant church at *Klostergrab* differently. Years earlier he had locked it up and begun fining subjects

---

<sup>60</sup>Asch 1997: 52-54; Evans, 1979: 63, 65; Limm 1984: 14; Maland 1980: 21, 61-62; Pages 1970: 43-49; Steinberg 1966: 34-36.

<sup>61</sup>Gindely 1884: 47-49. Pages, 1970: 49-50.

<sup>62</sup>Mortimer, 2015, 132-3. Polisensky, 1971: 76-7.

<sup>63</sup>Guttman, 1988:763 (*italics added*). Maland 1980: 62. Mortimer, 2015: 133. Pages, 1970: 50.

<sup>64</sup>Mortimer, 2015:134.

<sup>65</sup>Bireley, 2014: 91. Mortimer, 2015:134.

for attending Protestant services in the region. Now he had the church destroyed. The move provoked outrage across Bohemia and beyond.<sup>66</sup>

Each side had reached the point where it feared it would lose everything if it made any further concessions. The Catholics had reached this point decades earlier when the Protestant share of the population surpassed eighty percent. In response the Catholics initiated the counterreformation campaigns. Now the Protestants were at this point – fearing that if they surrendered any more of their religious rights they would lose everything. How is it that *each* side perceived that it was *completely* vulnerable to the other side’s advances? Did each side perceive that relative power was shifting against it? Or did each side perceive that the other side was advancing through non-power strategies, such as religious pedagogy and legal maneuvers? Whatever the source of its sense of vulnerability, each side felt the only way to prevent the other side’s advance was to use force.

#### **4. The spillover problem erupts: Constitutional reformism in Bohemia creates a risk of secessionism in Austria**

In March of 1618, the Protestant Estates of Bohemia decided to exercise their constitutional rights codified in the *Letter of Majesty* nine years earlier. They selected a Council of Defenders to advocate for their religious rights, negotiate with the Catholics, and appeal to the king if necessary. The defenders in turn convened a confessional Diet of representatives from all regions in the province of Bohemia. After considering a wide range of grievances about Habsburg religious policy – including the reservation of top government office for Catholics<sup>67</sup> – the Diet concluded that Protestantism was under threat and a resolute response was necessary. It formulated a comprehensive appeal for redress and sent it to Emperor Matthias.<sup>68</sup>

The Emperor and his chief minister asserted that the Protestant diet had no jurisdiction over the question of the Protestant churches at *Braunau* and *Klostergrab*, so their convening

---

<sup>66</sup>Bireley, 2014, 91. Mortimer, 2015:134. Pages, 1970: 51. "This senseless piece of violence touched off the rebellion." (Pages, 1970: 51)

<sup>67</sup>Bireley 2014, p. 91.

<sup>68</sup>Mortimer, 2015:135.



of the Diet constituted an illegal challenge to Habsburg authority. The chief minister immediately concluded that their challenge must be met firmly *on procedural grounds* without addressing the *substantive* issues they were raising. Acting on his advice (and possibly even the orders of the Bohemian Chancellor in Prague), Emperor Mathias sent a letter to the royal regents in Prague – to be delivered to the Protestants – maintaining that their diet lacked jurisdiction, was opposed to the Emperor’s authority, and was therefore prohibited from meeting again. The letter deferred judgment on the substantive issues until some later date when the Emperor might return to Prague, but without specifying when he might do so. The letter also threatened that in the meanwhile the Emperor would punish the Protestant Bohemian leadership if they continued to flout his authority.<sup>69</sup> In short, the Habsburg government was demanding that the Bohemian Protestants first give up their constitutional right to assemble and voice grievances, which had been guaranteed by the *Letter of Majesty*, and then the Emperor might address their concerns.

The harshness of the letter surprised even one of the Catholic regents in Prague who had refused to sign the *Letter of Majesty* years earlier. The Protestants assumed that the regents must have written the letter themselves, since Emperor Mathias would never take this position unless his advisers recommended it. The Protestants were vexed and determined to stand firm. The chief minister sent them another letter that sounded much more moderate but maintained the original demands that they relinquish their right to assemble and defer their substantive concerns indefinitely. Claiming that their next meeting had already been voted onto the calendar – so they could not cancel it – the Protestant diet met in Prague on May 21 and 22, 1618. The Diet concluded that the Habsburg demands simply proved that Protestant constitutional and religious rights in Bohemia were under threat. The Diet members decided to meet with the main Catholic regents on the next day to present their list of grievances and obtain redress. The nobles planned to frame the issue as one of religious freedom rather than political opposition to Habsburg authority. They simply wanted a reaffirmation of the religious rights granted by previous Habsburg rulers and the constitutional right to assemble and voice grievances in defense of those rights.<sup>70</sup>

---

<sup>69</sup>Mortimer, 2015:135. Pages, 1970: 51-2.

<sup>70</sup>Mortimer, 2015:135-8. Pages, 1970: 52.

#### 4.1 Negotiating for constitutional safeguards of religious rights: The ddefenestration of Prague

The diet's strategists had three goals for the meeting with the Habsburg regents. First, the strategists wanted to prove that the regents had been persecuting Bohemian Protestants with the aim of eliminating them. Second, the strategists wanted to highlight that repeated efforts to hold the regents accountable by nonviolent means had failed, so the only way to stop them was to resort to violence. Third, the strategists wanted to demonstrate that the Protestant nobility of Bohemia was *willing* to resort to violence to stop the regents. To achieve these aims the strategists conducted the meeting like a trial. With all of the diet members packed into the regents' chambers in Prague, the strategists read the charges against the regents, presented the evidence, and asked the diet members for their verdicts.

To prove that the regents were persecuting Bohemian Protestants, the strategists focused on the regents' actions that violated the *laws and customs* of the kingdom. The first charge was that the regents themselves had issued the prohibition on further diet meetings to discuss religious issues. The strategists argued that this prohibition violated the terms of the *Letter of Majesty* which guaranteed the right to convene the diet. When the regents denied having issued the prohibition, however, the strategists knew they could not prove the regents had issued it. So the strategists turned to their second charge: the king's officials had pursued policies designed to persecute Protestants *intentionally*. This was proven, the strategists argued, by the fact that two of the regents in the room had refused to sign the *Letter of Majesty* years earlier. Their refusal showed that the Habsburg government of Bohemia was reserving the right to violate the Protestant rights codified in the charter. The implication was that as long as the two regents held out with impunity, the rights of Bohemian Protestants would be at risk. So the Diet needed to punish the regents to signal that it would not tolerate royal officials who violated Protestant rights – and punish them forcefully to signal it was resolved to use force to defend Protestant rights. The assembled diet members returned judgment by denouncing the regents and demanding punishment. The strategists obliged the assembly by throwing the regents out of the window of their chambers.<sup>71</sup>

---

<sup>71</sup>Limm 1984: 14; Maland 1980: 62-63; MacHardy 2003: 68-70; Pages 1970: 50-53; Parker 1997: 43;

Perhaps the most notable feature of this trial was that the Bohemian strategists did *not* impose this punishment based on the first charge (that the regents themselves had issued the prohibition on further diet meetings). Apparently the strategists felt constrained *only* to impose punishments based on charges they could actually prove to the diet's satisfaction. This feature of the trial sheds light on a crucial debate among later historians. Some have argued that the leaders of the Bohemian Protestant nobility were more radical than the majority of the nobles; and the leaders dragged the majority into a war they did not want. That hypothesis is contradicted by the observation that the leaders of this trial felt constrained by the majority opinion of the noble diet as a whole. As long as the leaders were constrained in this way, it is irrelevant whether their *preferences* were more radical than the majority's preferences, because their *actions* were constrained by the majority's preferences. The implication is that if there was a war, it was because the majority of the nobles were willing to stand firm on their demands even if it led to a war.

The negotiating positions and actions of the Bohemian Protestants following the "defenestration of Prague" revealed their preferences. Although the diet was willing to go to war to defend Protestant rights, its goal was *not* to overthrow the Habsburg government of Bohemia. On the contrary, the diet wanted to resume its contractual relationship with the King of Bohemia and his remaining officials – albeit on restructured terms – and was proceeding by a combination of force and reason that seemed likely to obtain this outcome (e.g. the same combination that had worked well in the previous crises). The diet was, in effect, negotiating for a constitutional reform. It sought the right to remove and replace royal officials who violated Protestant rights. In time it also sought the right to limit the king's options in appointing nobles to government offices to exclude candidates likely to violate Protestants' rights. By removing the two regents forcefully, the diet signaled its resolve to secure this reform by force if necessary.

Following the defenestration, the diet members immediately "constituted themselves a *Landtag*" and elected a Directory of thirty nobles to replace the ten Habsburg regents.<sup>72</sup>

---

Petran 1998: 86-87; Polisensky 1971: 98-100; Steinberg 1966: 36. It is not clear from the English language histories whether the idea to throw the regents out of the window originated with the strategists or the Diet members at large. In any case the ritual of throwing offenders out a window had a long history in the kingdom as a sign of disapprobation and resolve.

<sup>72</sup>Gindely, 1884:64-65, 74. Pages, 1970: 54.

To signal that they were not repudiating Habsburg royal authority in Bohemia but only exercising the right to replace royal officers guilty of violating the *Letter of Majesty*, the Diet installed the thirty new Directors "in Matthias's name".<sup>73</sup> The new Directory quickly took control of the administrative and fiscal machinery of the Bohemian government in Prague and started to raise an army of mercenaries to defend the religious rights of Bohemian Protestants. The Directory was endorsed by most of the Protestant nobility in the province and even by some of the Catholic nobility (after they were promised religious toleration). The region loyal to the Directory expanded quickly to encompass most of the northern and central parts of the province.<sup>74</sup>

#### **4.2 The Austrian response: Fear of secessionism at home prompts a quick ultimatum**

The defenestration of Prague turned a problem of Protestant rights within the Kingdom of Bohemia into a cross-border spillover problem. For the Habsburgs had to wonder: If they allowed Protestants in the Kingdom of Bohemia to remove and replace *their* royal officials by force, then perhaps the Protestants in the Kingdom of Austria might start removing and replacing *their* royal officials by force. If so, then the foundations of Habsburg authority in Austria itself could crumble.

Archduke Ferdinand and chief minister Klesl convinced Emperor Matthias to send an envoy to Prague to gather information and stiffen the resolve of the regents to stand firm.<sup>75</sup> After meeting with the Bohemians on June 6, the envoy sensed that they were strongly opposed to the Habsburg demands and would resort to force to defend their rights rather than submit. He knew the Habsburg government lacked the necessary force to prevail in a war, however.<sup>76</sup> He concluded that the Bohemian opposition to Habsburg demands was militarily credible (implication 6.1s). He decided to return to Vienna and advise the Emperor to make concessions and search for a peaceful compromise that might be acceptable to the

---

<sup>73</sup>Quote is from Maland, 1980: 63. Pages, 1970: 54.

<sup>74</sup>Gindely, 1884:65-66. Bireley, 2014: 91.

<sup>75</sup>Bireley, 2014: 91.

<sup>76</sup>Gindely, 1884:74-75.

Bohemians.<sup>77</sup> In particular he would advise the Emperor to promise to respect the *Letter of Majesty* and stop claiming that he had been respecting it.<sup>78</sup>

The Bohemian Directory in Prague was divided politically between a moderate group and a hardline group.<sup>79</sup> The moderates genuinely wanted to reach a compromise with Vienna that would leave the Habsburg king with real power in Bohemia. The hardliners wanted a degree of autonomy for Bohemia that would leave the Habsburg king of their kingdom a mere figurehead.<sup>80</sup> The moderates prevailed in the debates over how to deal with Vienna. The Directory responded to Vienna's demands with reassurances and accommodations (implication 8s). The Directors reassured Vienna by announcing that the Bohemian people only wanted a reaffirmation of their religious freedoms, not political independence. To accommodate Habsburg concerns, the Directors professed that the people were willing to retain Ferdinand as their King in Bohemia as long as he guaranteed their religious rights. As if to prove it the Directors blamed the dispute on Ferdinand's officials rather than the king himself. The Directory did not issue any military threats or set any deadlines for Vienna to meet its demands (implication 9s). At this stage the Directors only took two steps that might block a compromise solution with Vienna. They banished the Jesuit (Catholic) clergy from Bohemia for all time – a move for which there was legal precedent (the Hungarian settlement of 1606 had banished the Jesuits from that kingdom for all time). And they confiscated the property of some of their staunchest Catholic opponents.<sup>81</sup>

The Habsburg government in Vienna assessed the threat posed by the Bohemians quickly and injudiciously (implication 1s). The Emperor was willing to wait for the envoy's return from Prague. While the Emperor was open to using force against the Bohemians if necessary, he believed the problem could be solved through negotiation and compromise.<sup>82</sup> "It was

---

<sup>77</sup>Bireley, 2014: 91.

<sup>78</sup>Gindely, 1884:75.

<sup>79</sup>For discussion of the religious divisions among the Bohemian nobility and the political divisions among their leaders, see Evans (1979: 54, 56), Polisensky (1971: 79) Pages (1970: 43-4).

<sup>80</sup>Gindely 1884: 86-87.

<sup>81</sup>Pages 1970: 54-5. Petran 1998: 87.

<sup>82</sup>Bireley, 2014: 92. Bireley's conclusion that Matthias was willing to negotiate is based on a letter that Matthias wrote to Ferdinand, saying that he was "inclined at least to attempt negotiations". By contrast, MacHardy (2003: 109) writes, "Soon after the defenestration of May 1618, both Habsburg rulers rejected

only in obedience to Klesl and Ferdinand's counsels that he had adopted the measures which so irritated the Bohemian Protestants."<sup>83</sup> He wanted to assess the problem slowly and judiciously. But Archduke Ferdinand and his advisers were not willing to wait for the envoy's return. They sensed that the Bohemians might withdraw from the Austrian Monarchy and establish a republic, and if they did so, Habsburg power would decline irreversibly.<sup>84</sup> The Protestant religions would prevail throughout Germany and it would be impossible to restore the Catholic religion. Ultimately it was not a problem of religion but one of government authority and the economic and political value it held for those who exercised it. The regime's chief minister, Cardinal Khlesl, explained the issue in a summary of the position of Ferdinand's faction.

So the Bohemian and Austrian states have brought their sovereign lords into such a servitude that the sovereigns can scarcely take action in the lands that they have inherited or exercise their sovereign authority except in so far as their subjects allow it, so that there remains to the sovereigns only the name of sovereign authority, and its exercise remains with the non-Catholic subjects.<sup>85</sup>

Remarkably, the memorandum claimed that even if military action against Bohemia failed, it would be better to lose authority in Bohemia that way than to lose it by capitulation without a war.

"You cannot lose anything [by initiating war]. If assaults of the [Protestant noble] estates as enumerated continue, you will certainly lose eventually with great harm and derision. [By contrast] should you lose [militarily], since the outcome of war

---

bargaining with the Protestant estates on the grounds that their survival was at stake." MacHardy may be right in her broader conclusion: "The available records of public expressions and private correspondence of the Chief Minister and the two emperors do not disclose fundamental differences in their representation of the Protestant opposition, or in motives in this conflict." (MacHardy, 2003: 108-9.) Nonetheless, Bireley's direct evidence shows that Matthias was willing to negotiate. This conclusion is supported by evidence that MacHardy herself cites elsewhere: "In October 1618, four months later, Matthias himself approached the Duke of Saxony to negotiate with the Bohemians." (MacHardy, 2003: 249, n. 105.)

<sup>83</sup>Gindely, 1884:74.

<sup>84</sup>Bireley, 2014: 91-2.

<sup>85</sup>Bireley, 2014: 92. Bireley says the memo was most likely written by Khlesl.

is uncertain, you would lose with honor and not shamefully.”<sup>86</sup>

As one advisor said, even if the Habsburg regime lost the war, it would only lose something whose possession no longer had any value.<sup>87</sup>

In responding to the perceived threat to its authority, the Austrian regime was not simply spearheading the religious interests of the Roman Catholic Church. On the contrary there had always been conflicts of interest between the regime and the Roman church despite their common religion. And the regime’s rulers and clerical advisors had always put the regime’s interests first.<sup>88</sup>

Within weeks of the defenestration of Prague, Ferdinand’s faction concluded that the Emperor should use military force against the Bohemians.<sup>89</sup> Ferdinand perceived that the Bohemian opposition to Habsburg demands was not militarily credible (implication 6.2s). Perhaps on the basis of his previous success in his counterreformation campaigns in Inner Austria, he expected the Bohemian opposition to crumble quickly under military pressure.

Before it became clear how serious the spillover problem was, Ferdinand’s faction endeavored to issue a firm demand to the Bohemians backed by a military threat (implications 2s and 3s). Without waiting for the envoy to return from Prague, Ferdinand’s faction convinced the Emperor to send a stern diplomatic communique to Prague. The communique promised that the *Letter of Majesty* would be observed *as it had been observed to date*, meaning that Bohemian Protestants would *not* be permitted to build churches on ecclesiastical land. The communique then demanded that the Bohemians “return to quiet” and threatened to impose punishments on them if they refused.<sup>90</sup>

---

<sup>86</sup>Bireley, 2014: 93. The first three bracketed phrases were added by the present author. The fourth one appears in brackets in Bireley.

<sup>87</sup>Gindely 1884: 75.

<sup>88</sup>As one historian wrote, “Sixteenth-century Habsburgs reserved the right to prevent papal interference, even in matters considered purely spiritual by Rome. The emperors resisted the growing influence of papal nuncios, both as acute diplomats and as spokesmen for papal pronouncements. Neither Rudolf nor Matthias had any great sympathy for the Society of Jesus, the Jesuit clergy; even at Graz and Innsbruck its impact on the dynasty can be exaggerated. There are other signs too of a rising generation of aulic Christians, who indeed serve Catholicism, but serve their own masters better. Ultimately they were rivals to a traditional Roman-Catholicism, and while seeming to be its allies they harnessed its energies for their own purposes.” (Evans, 1979: 59-62.) The quoted section of Evans’ book also refers to the situation after 1600.

<sup>89</sup>Bireley, 2014: 91; Gindely 1884: 75.

<sup>90</sup>Gindely, 1884:75-76.

### 4.3 The Bohemian response: Firm yet restrained

The demand to "return to quiet" essentially reiterated the earlier demand that the Bohemians disband the Council of Defenders and the Diet. This demand remained a non-starter in Prague, because the *Letter of Majesty* guaranteed the right to convene both bodies. The effect of the Habsburg ultimatum was to cause deadlock and an end to negotiations (implication 4s). The Bohemians stopped offering proposals for a negotiated settlement and simply reiterated that their cause was just and their military preparations were justified.<sup>91</sup> This outcome may have been intentional on the part of Ferdinand and his advisers – since they had already decided they wanted war.

The Directory in Prague then established diplomatic contacts with the other provinces of the Kingdom of Bohemia to convince them to join its campaign for religious freedom. The province of Moravia to the east remained neutral.<sup>92</sup> The province of Silesia adopted a somewhat more favorable posture. The Silesian Diet advised the Bohemians to pursue a negotiated settlement, but recommended that they only accept one if their conditions were met by the Habsburg government in Vienna.<sup>93</sup> This posture suggests that the Silesians wanted the Bohemians to secure the religious freedom of both provinces by peaceful means, but if this was not possible, the Silesians would join the Bohemians in using force to secure it. To signal its resolve on both points, the Silesian Diet voted to raise an army of 6000 troops but to keep the force on Silesian soil for the time being – in the hope that the Directory in Prague would reach a negotiated settlement with Vienna.

### 4.4 Austria initiates war without allies

Prior to the eruption of the crisis in May, Austria already had 3000 Spanish troops under its control (for use in winding down a separate conflict in Friuli). The troops were scheduled to be dismissed on June 1. After the defenestration in Prague, however, Mathias asked Spain's ambassador in Vienna for military support. The ambassador pledged to continue funding

---

<sup>91</sup>Gindely, 1884:76.

<sup>92</sup>Pages, 1970: 55-6.

<sup>93</sup>Gindely 1884: 83-84; Polisensky 1971: 106.



the 3000 troops for the time being but declined to make any other commitments, because he knew the Spanish empire was so overstretched that it could barely raise the funds to maintain its current commitments. At the same time, though, he "[held] open the possibility that [the Spanish King] would give sympathetic consideration to any request for help which might be made." Austria's ambassador in Madrid did not wait for instructions from his government: he requested military support from the Spanish government immediately after hearing of the defenestration in Prague.<sup>94</sup>

Within the Habsburg government, Ferdinand's faction wanted Mathias to switch from a diplomatic strategy to a military strategy and prepare for war. However, the regents in Prague were advising Mathias to make concessions and reach a peaceful settlement if the Bohemians agreed to disarm.<sup>95</sup> Mathias' chief minister, Cardinal Khlesl, thought a compromise settlement would be feasible and desirable, "and avoided doing anything that might exacerbate feelings."<sup>96</sup> Ferdinand sensed that Mathias would stay on the diplomatic track as long as his current advisers remained in place. So "after some infighting at court" – in which Ferdinand accused Khlesl of treason – Ferdinand's faction mounted a palace coup and took control of Habsburg policy.<sup>97</sup>

Ferdinand knew that a military victory over the Bohemians would be impossible without allies.<sup>98</sup> Yet he did not wait to receive an answer from Spain on whether it would provide military support. He and his advisers quickly decided to initiate war against the Bohemians.<sup>99</sup> <sup>100</sup> Anticipating hostilities, the Bohemian army moved into the southern part of the

---

<sup>94</sup>Brightwell, 1979: 416. Bireley (2014: 96) says it was 3000 Spanish troops. Pages (1970: 55-6) says it was 6000.

<sup>95</sup>Bireley, 2014: 91.

<sup>96</sup>Pages, 1970: 54.

<sup>97</sup>Bireley, 2014: 93-94. Evans, 1979: 65. Pages, 1970: 55-6.

<sup>98</sup>Bireley, 2014: 93. Brightwell, 1982a: 120.

<sup>99</sup>Asch 1997: 54-55; MacHardy 2003: 70; Mortimer 2010: 24; Mortimer, 2015: 160-1; Pages 1970: 54, 56-57; Petran 1998: 87, 89; Polisensky 1971: 102; Steinberg 1966: 37.

<sup>100</sup>Spain decided in June and July of 1618 to provide limited military support to Austria (see below). The Spanish government communicated these decisions to the Austrian ambassador in Madrid in late July. (Brightwell 1979, p. 425) This news cannot have reached Vienna until August. Austria's decision to go to war was made in Vienna in late July.

One might object that the 3000 troops committed by the Spanish ambassador in Vienna constituted an "ally". Yet Ferdinand and his advisers had no indication from the Spanish government that it would continue

province to force the main cities there, which were largely Catholic, to join the Protestant cause.<sup>101</sup> While the Habsburg regime in Vienna might have seen the move as a threat to lower Austria, the regime had more fundamental reasons to launch a military campaign against Bohemia. The Habsburg army invaded Bohemia in August.<sup>102</sup> The duration of time from the emergence of the spillover problem to the Habsburg decision to initiate war was comparatively short: two months (implication 5s).

## 5. Mobilizing for war

The battles and skirmishes in the fall of 1618 were inconclusive. Neither army had a decisive power advantage so neither seized much territory – although each plundered many civilian areas.<sup>103</sup> The opposing armies failed for different reasons, however. The Bohemians failed because they had *not yet* tapped their domestic base of military support: the Protestant majorities in the kingdom's five provinces who were threatened by Habsburg rule. The Austrian Habsburgs failed, in contrast, because their domestic base of military support was small and shrinking. The stalemate was bound to break in favor of the Bohemians as their domestic support solidified and the Habsburg's domestic base shrank further.

### 5.1 Domestic foundations of relative power

After the war began in August of 1618, it took some time for the Bohemian cause to reach its military potential. At first the provisional government in Prague had difficulty mobilizing the provincial militia force to full strength. The Directors were inexperienced at military organization.<sup>104</sup> The Directory became more effective over time, however, and a second

---

to fund these troops. Even if they had received such an indication, the 3000 troops were simply the domestic force the Austrian Habsburgs had needed to maintain their existing commitments in their own lands – even before the Bohemian crisis erupted. In this sense these troops did not constitute a foreign ally at all, but simply a domestic force for maintaining domestic order.

<sup>101</sup>Bireley, 2014: 91.

<sup>102</sup>Mortimer, 2015: 161.

<sup>103</sup>Pagès 1970, p. 56.

<sup>104</sup>Steinberg (1966: 37-8) says that some of the Protestant nobles in the countryside resisted the taxes needed to fund the force.

round of mobilization and taxation was more successful. As the fighting continued, the other provinces of the kingdom began to see the value of authorizing military force to secure their religious freedoms. In October of 1618, the province of Silesia released half of its force into action to defend the province of Bohemia from the Habsburgs.<sup>105</sup> While the province of Moravia remained neutral for the time being, sentiment among its Protestant nobility was shifting in favor of joining the Bohemian side.

The Habsburg rulers in Vienna faced the opposite trends. The Protestant Estates of Austria, comprising the vast majority of the kingdom's landowners, refused to provide any funding for the Habsburg war effort. Driven by reluctance to oppose fellow Protestants – and reluctance to support a losing cause – their refusal left the government to rely on the much smaller Catholic Estates and the wealth and credit of its own ministers. Austria's only ally, the overstretched Spanish empire, continued to fund the 3000 troops it had been underwriting before the Bohemian crisis erupted, but could not send any reinforcements for the foreseeable future (see below). The Austrian branch of the House of Habsburg was on its own. Worse, its main base of domestic military support, the Austrian Protestant Estates, harbored the same demands for religious freedom that the Bohemian Protestant Estates had just begun fighting for.

The net balance of forces and interests pointed toward a compromise settlement. The Bohemians were growing in power, but preferred to reach an accommodation with Vienna by offering reassurances and concessions. The Habsburgs preferred to impose a solution on Prague by force, but were losing the power to do it with every passing day – as their own Protestants stood aloof and stiffened their resolve to stand firm like the Bohemians. The crisis showed every sign of moving toward the same kind of settlement that had ended the previous crises: A compromise solution driven by the fact that neither side had the power *and* the interest to impose terms on the other side by force.

## **5.2 Austria's request for decisive Spanish intervention is rejected**

Immediately after the defenestration of Prague, Austria's ambassador in Madrid requested military support from the Spanish government (see above). The Spanish Council of State

---

<sup>105</sup>Bireley, 2014: 94-95.

debated the question of whether to provide military support in June and July of 1618 – before Vienna made the decision to go to war. The Spanish Council was deeply divided. On one side of the debate, the Spanish King’s most trusted advisor argued that if Spain provided support, it would only embolden Austria to stand firm in the diplomatic negotiations and get into a war. Yet neither country could afford the force needed to win a war. Once a war began, though, it would be too late to reach an agreement with the Bohemians. He therefore advised the Spanish King to refuse Austria’s request for military support in the hope that this would induce Vienna to settle with the Bohemians. On the other side of the debate, Spain’s former ambassador to Austria, now a member of the Spanish Council of State, argued that Spain had to continue funding the 3000 troops already under Austria’s control, because if it stopped the funding this would ruin the Austrian King’s reputation. In the end the Spanish King tried to combine the two positions. He decided to continue funding the 3000 troops, but to make no further commitments of military support and instead to press Mathias to reach a negotiated settlement with the Bohemians.<sup>106</sup> As one historian summarized the conclusion of the Spanish King in Council:

”[Mathias] must seek to escape from his difficulties by a judicious blending of force and reason. Too great a show of force [without reason] would only extend the war; yet reason alone [without force] would encourage the rebels to hold out for terms which could only be accepted with dishonor. Philip decided to help Mathias to fight a war, but at the same time [Spain’s ambassador] was to urge Mathias to work for a settlement in Bohemia ’by means of a peaceful agreement’.”<sup>107</sup>

Although these decisions were made in June and July of 1618, news of them did not reach Vienna until late in the summer of 1618 – after Ferdinand’s regime had already decided to go to war and the fighting had already begun.<sup>108</sup> Spain would not make any further commitments of military support to Austria for another year (e.g. in June of 1619).<sup>109</sup> By

---

<sup>106</sup>Brightwell, 1979: 426.

<sup>107</sup>Brightwell, 1979: 426.

<sup>108</sup>Brightwell, 1979: 425.

<sup>109</sup>Brightwell, 1982a: 125. See below.

that time the net balance of power between Austria and Bohemia had shifted greatly toward Bohemia due to the domestic factors mentioned earlier.

Spain's commitment to continue funding the 3000 troops raises an important question. Did this commitment affect Austria's decisions to initiate war and to continue fighting once the war began? The Spanish commitment could not have affected Austria's decision to initiate war, because news of the commitment did not reach Vienna until after Ferdinand and his advisers had already decided to initiate war. One might object that Ferdinand and his advisers *expected* Spain to make the commitment and this expectation affected their decision. This objection begs the question. Even if Ferdinand's regime did initiate war in the expectation of receiving Spanish military support, why didn't his regime settle for peace quickly after receiving the news that Spain would not provide any military support beyond the 3000 troops it was already funding, and it was only funding these troops to enable Ferdinand's regime to stave off military disaster for long enough to reach a compromise settlement with the Bohemians? If Austrian decision-making had been motivated by the Spanish commitment, then Austria would have settled for peace quickly after the war began. It did not. Instead it stood firm diplomatically and fought on for another year even though it did not receive any further military support from Spain or any other state. This course of action cannot have been motivated by the Spanish commitment. It must have been motivated by some other factor.

This conclusion is bolstered by the observation that the Spanish commitment declined in military value quickly. Two months after Austria received the Spanish commitment, Bohemia received a contingent of 2000 troops sent by a principality in southern Germany affiliated with the Protestant cause. This development reduced the net value of the Spanish commitment from 3000 to 1000 troops. Yet Austria fought on. If a Spanish supply of 3000 troops was not enough to motivate Austria to stand firm diplomatically and fight on for another year, then a net supply of 1000 troops certainly was not enough to motivate it.

## **6. Fighting while bargaining: The terms of the dispute are refined by combat**

After the fighting began in August of 1618, the diplomatic negotiations between Austria and Bohemia continued. Each side clarified and refined its negotiating position over time. The proclamations of the opposing leaders reveal three phases in their thinking about the nature of the dispute and the conditions for its resolution. In the first phase each side focused on offering concessions in the hope of attracting the other to a settlement. In the second phase it became clear that their disagreement was not simply about whether one side had violated a rule but about what constituted a "violation" in the first place. In the third phase the terms of the dispute were finalized into diplomatic and constitutional forms that reflected this absence of mutual understanding.

Although the three phases were sequential in time, the lines between the phases were not perfectly clear. Each phase coagulated into the next phase like the three forms of water in a lake that is gradually freezing over for the winter – gas, liquid and solid. In the summer, there is a gaseous fog on the surface of the lake in the mornings. By the fall, there are fewer mornings with a fog on the surface and some mornings with chunks of ice on the surface – but they melt during the day. By the middle of winter, of duties the entire surface is frozen all day long and there is never any fog on the ice cap. While the last phase is very different from the first, the transition from each phase to the next is gradual and reveals multiple forms of water at the same time. So it was with the three phases of the opposing leaders' thinking about the dispute of 1618. The third phase was very different from the first, yet the transitions between phases were gradual and exhibited multiple perspectives at the same time. It is worth considering each phase in turn because this provides evidence on the underlying causes of the dispute and highlights the impact of military developments on the disputing process, thereby helping to resolve debates about what caused the crisis of 1618 and the Thirty Years War.

## 6.1 Phase 1: Offering concessions

Soon after the war began each side reiterated the minimum conditions that the other side would have to meet to bring an end to military hostilities. Each side also offered concessions in the hope of inducing the other side to meet its conditions. For his part, King Ferdinand conceded that he was willing to guarantee the Protestants' religious rights provided they lay down their arms first. This condition was essential from his perspective, because the Bohemians had evicted royal officials from office and resorted to arms to enforce the evictions. If he permitted their actions to stand, it would establish a precedent that would give them the power to destroy the very foundation of Habsburg royal authority in Bohemia. Hence Ferdinand had to demand that they cease their armed enforcement of the evictions and lay down their arms.

For their part, the Bohemians conceded that they remained loyal to Ferdinand and were willing to retain him as the king of Bohemia provided he guarantee their religious and political rights. This condition was essential from their perspective, because the evicted royal officials had violated the rights of Protestants: curtailing their freedom of worship, seizing their church property and denying them an appropriate share of government offices. If the Bohemians permitted such violations to stand, it would set a precedent that would grant royal officials the power to destroy the very foundations of Protestant existence in Bohemia. Hence the Protestants had to demand guarantees of their religious and political rights before they could lay down their arms.

Neither side's concessions were sufficient to induce the other side to meet the relevant condition. What is notable, however, is that each side wanted to negotiate, kept its concession offer open, and hoped it would induce the other side to settle. The Bohemians had been willing to negotiate all along. The Habsburg regime in Vienna, controlled by Ferdinand's faction, had avoided negotiations after the defenestration in May. Within a few months, however, "realizing the weakness of their position, both Matthias and Ferdinand now agreed to negotiate with the rebels with the help of Saxon mediation, but on condition that the rebels dismissed the troops they had recruited and dissolved the government that they had formed. As its position further deteriorated Vienna gradually backed off from these conditions."<sup>110</sup>

---

<sup>110</sup>Bireley, 2014: 94. Cf. MacHardy, 2003: 111, n. 105.

Various third parties attempted to mediate between the two sides.

## **6.2 Phase 2: The crux of the dispute – disagreement over what constitutes a “violation”**

In the dispute of 1618, three political assets were at stake:

1. The government’s authority to issue laws provided it obtained the consent of the noble estates.
2. The government’s authority to make policies provided they did not violate the nobles’ customary rights and privileges.
3. The government’s authority to control appointments to government office provided the nobles received an appropriate share of the appointments.

The rights to these assets were governed by Bohemian laws for which there were precedents going back centuries. In the dispute of 1618 these precedents were ambiguous as to what should be the exact distribution of rights and obligations between the two sides. Each side interpreted the precedents in its own favor, so each side claimed it had the *sole* right to the asset. Couching the rights in this way, only one side in the dispute could have each right. Why couldn’t the opposing sides simply divide the right into partial shares, giving each side partial rights? Neither side could do so because neither side knew whether the other side had created the dispute intentionally or unintentionally. That is to say, for all each side knew, the other had created the dispute intentionally so agreeing to a settlement that gave the other side partial rights would allow the opponent to gain on net from the dispute. That would give the opponent an incentive to create more disputes intentionally until the opponent had gained *all* of the rights to the asset in question. In concrete terms:

1. If the king gave the Bohemians the right to approve or veto *one* of his regime’s religious laws, then they would intentionally create more disputes until they had the right to approve or veto *all* of his regime’s religious laws. On the other hand, if the Bohemians gave up the right to approve or veto *one* of the regime’s religious laws, then the regime would intentionally create more disputes until the Bohemians had given up the right to approve or veto *any* of the regime’s religious laws.
2. If the Bohemians allowed the regime to institute *one* religious policy that violated their



customary rights, then the regime would intentionally create more disputes until it was violating *all* of their customary rights. On the other hand, if the regime abandoned *one* policy because it was violating the Bohemians' customary rights, then they would intentionally create more disputes until the regime had to abandon *every* policy because of a claim that it violated their customary rights.

3. If the Bohemians permitted the king to retain *two* regents who had refused to sign the *Letter of Majesty* and respect Bohemians' religious rights, then the king would intentionally create more disputes until he had a full cadre of regents and administrators who held the same preferences and took the same actions as those two. Yet if the king permitted the Bohemians to control *any* government appointments, then they would intentionally create more disputes until they controlled *all* government appointments.

In the following subsections I present the historical evidence that: (1) the Bohemian laws governing these political assets were ambiguous; (2) each side interpreted the laws and legal precedents in its own favor; (3) each side was uncertain about the other side's intentions in causing the dispute; and (4) each side's uncertainty required it to use force to prevent the other side from benefiting from intentional opportunism. I also consider alternative rival hypotheses for why the two sides could not split the difference in the dispute peacefully and present the historical evidence that refutes those hypotheses.

### 6.2.1 Ambiguous legal precedents

The law was ambiguous in three areas. First, the king had the sole right to *issue* new laws, but the noble estates had the right to *consent* to new laws. The estates interpreted this to mean that they had the right to grant *or withhold* their consent to new laws. The king interpreted it to mean simply that the estates had a right to be *consulted* before he issued a new law. But because their right to be consulted had been delegated by the king himself – from his own princely authority (*iura majestatis*) – in the end he could issue a new law over their objections, and they had no right to resist his authority over the issue.<sup>111</sup>

Second, the king had an obligation to respect the nobles' customary rights and privileges when issuing new laws. The nobles interpreted this to mean that the king did not have the

---

<sup>111</sup>MacHardy, 2003.

right to change customary laws at will or issue new statutory laws that violated customary laws. But the Habsburg Kings asserted two exceptions to this rule as legal codification and judicial centralization proceeded in the 16th century. The nobles had to provide proof that a previous king had actually granted the customary right or privilege in question. And the king still had the right to change it or violate it if that was required to protect the "public good". And again the nobles did not have the right to resist his authority over such issues.<sup>112</sup>

Third, the king had the right to control appointments to government offices, but for some government offices he had a customary obligation to appoint only members of the noble estates. As the Protestant confessions spread into Bohemia in the 16th century, the nobles interpreted this to mean that the king had a customary obligation to appoint *Protestant* members of the noble estates to some offices. But none of the Kings had ever granted this concession in writing since the outbreak of the Protestant Reformation in 1517, and over time the Habsburg Kings began to view Catholic officeholders as essential for the "public good".<sup>113</sup>

The question naturally arises: Why couldn't the two sides simply split the difference in the dispute? Why couldn't they reach an agreement to give the king *partial* control over lawmaking, policymaking and government appointments and the Bohemians *partial* control – provided they accept the regime's authority?

### **6.2.2 Expansive interpretations of the law generate disagreement over what constitutes a violation**

The Habsburg regime took an expansive view of its rights in these areas. Hence the regime saw any attempt by the Bohemian Protestants to interfere with these rights as illegitimate, and maintained it had every right to punish the Bohemians until they stopped interfering. The Bohemian Protestants, in turn, took a more restrictive view of the regime's rights and consequently a more expansive view of their rights to resist the regime. They saw attempts by the king and his officers to claim unrestricted rights in these areas as illegitimate, and maintained they had every right to resist the regime until it stopped asserting unrestricted

---

<sup>112</sup>MacHardy, 2003.

<sup>113</sup>MacHardy, 2003.

rights. In short, the two sides disagreed over what constituted a "violation" and what constituted a "just enforcement action" in response. The Bohemians held that the regime had violated their religious and constitutional rights. The regime held that it had not committed any violation so the the Bohemians' resistance constituted sedition and rebellion.

There were historical and legal precedents for *each* side's position in the dispute. These precedents had been in place for decades or centuries. And so had the gaps and ambiguities in these precedents. So why did these issues only cause a war in 1618 and not before? The answer is that no previous situation created an "overlap" between the opposing sides' legal rights in the same way that the conjuncture of 1618 did.

### **6.2.3 Alternative rival hypotheses on what blocked a settlement and forced the war to continue**

One traditional hypothesis is that Catholic rulers always feared any kind of rebellious movement by their Protestant subjects. For once a rebellion broke out, the Protestants might persevere to the point of deposing the Catholic ruler and his regime entirely. Hence the perception that a rebellion was starting led Catholic rulers to escalate to all-out war immediately.<sup>114</sup> This hypothesis is refuted by the evidence on Protestant rebellions in Austria between 1595 and 1611 (see above). While in some cases the Habsburg regime crushed the rebellion completely, in other cases the rebellion was settled by a compromise, yet in *neither* set of cases did the rebellion lead to an all-out war that lasted for decades.

Another traditional hypothesis is that each side saw the stakes in the dispute as "all or nothing". Some Catholic political theorists maintained that a territory could only be ruled if its entire population adhered to a single religion. As long as there were some Protestants in the territory, rule was impossible.<sup>115</sup> In turn some Protestant resistance theorists maintained that once a ruler became tyrannical, the subjects had a right to depose him entirely.<sup>116</sup> With such "all or nothing" stakes, even a minor legal dispute could lead to an all-out war.

This hypothesis rests on the assumption that the conflict of interest between the Catholics

---

<sup>114</sup>MacHardy: 2003: 109.

<sup>115</sup>MacHardy: 2003: 110.

<sup>116</sup>MacHardy: 2003: 98-101.

and Protestants was "indivisible" – that there was no way to divide the political and economic assets at stake in the conflict and give one share to the Catholics and another share to the Protestants. This assumption is refuted by the evidence on inter-confessional relations in Europe between 1517 and 1688. The Peace of Augsburg in 1555 divided the relevant assets in a manner that was more or less sustainable for decades. The settlements in Austria between 1605 and 1611 divided the relevant assets without any major wars. Most historians agree that by 1617 the inter-confessional conflict of interest was settling down because each side preferred peace – even at the expense of making painful concessions – to a repeat of the wars of 1517 to 1555.<sup>117</sup> Finally, the Peace of Westphalia in 1648 divided the relevant assets in a manner that was quite sustainable. Within decades of the settlement of 1648, inter-confessional conflicts of interest had subsided in the popular consciousness to the point where they were simply no longer a cause of war. In sum the historical evidence does not sustain the assumption that the stakes between Catholics and Protestants were "indivisible".

So why did the Bohemian crisis of 1618 cause an all-out war that lasted for decades? Why couldn't the two sides figure out a way to divide the relevant assets and devise a compromise settlement? The evidence from the many settlements between 1648 and 1688 suggests there were many ways that the Austrians and Bohemians *could* have divided the stakes in the dispute of 1618. Why were they unable to arrive at such divisions?

The standard explanation is that the opposing sides settled for peace in 1648 because everyone was "tired of war", whereas in 1618 neither side was "tired" yet. While this hypothesis sounds intuitive, it also predicts that each of the crises between 1605 and 1611 would lead to a long war. Yet none did. What was it about the crisis of 1618 that made it so much harder for the opposing sides to divide the stakes in the dispute? The answer becomes evident by examining the concept of *asset division costs*.

Let us assume that dividing the stakes in the dispute is a costly process itself. So it is only worth settling for peace and incurring the *asset division costs* once the costs of war have become even higher. Before that point is reached it is not worth settling for peace. This hypothesis implies that the higher are the asset division costs in a particular dispute, the longer the war will last before it reaches the point where the war costs are even higher.

---

<sup>117</sup>Wilson 2008.

From this vantage, the crisis of 1618 produced a much longer war because it stemmed from a legal dispute in which the asset division costs were much higher than they were in other crises. Why were they much higher in 1618? The answer is straightforward.

#### **6.2.4 Uncertainty about intentions requires each side to use force**

The dispute was caused by each side's uncertainty about the other side's intentions in the legal dispute. The Habsburg regime could not tell whether the Bohemians were challenging its political authority *intentionally* to overthrow it or only *unintentionally* as a byproduct of their effort to defend their religious rights. For all the regime knew, the Bohemians were challenging its authority intentionally to overthrow it. So it had to impose war costs on the Bohemians until they agreed to accept its authority. In practice this meant continuing the war until the Bohemians accepted the king's authority to control appointments to government office and his officers' authority to issue laws and policies without obtaining their consent to each one. In turn, the Bohemians could not tell whether the regime was *truly unaware* of the evidence proving it had violated their religious rights, or *intentionally bluffing it was unaware* when it was really aware. For all the Bohemians knew, the regime was intentionally bluffing it was unaware of the evidence when it was really aware. So the Bohemians had to impose war costs on the regime until it accepted government administrators who would acknowledge evidence of government wrongdoing when it existed and reverse policies proven wrongful. In practice this meant continuing the war until the regime gave the Bohemians the right to control appointments to government office – at least to the extent necessary to secure appointees who would acknowledge evidence of government wrongdoing and reverse policies proven wrongful.

This characterization of the dispute implies a specific pattern of diplomatic discourse between the opposing leaders: a sequence of accusations, denials, reassurances, diagnoses of the "real" problem, and claims about the right to use force to resolve the problem. This pattern is observable in the historical record.

The Habsburg regime accused the Bohemian Protestants of intentionally aiming to overthrow it. The Bohemians denied the accusation and reassured the regime that (1) they remained loyal, (2) they were only trying to secure their religious rights, and (3) the real

problem was that the regime was refusing to acknowledge the evidence proving it had violated their rights. The regime denied the accusation and reassured the Bohemians that (1) it remained just, (2) it was only trying to serve them, and (3) the real problem was that they were intentionally trying to overthrow it. The regime claimed it had the right to punish the rebels until they accepted its authority. The Bohemians claimed they had the right to resist the regime until it stopped claiming the right to trample their religious freedoms and ignore the evidence of its crimes with impunity.

This characterization of the terms of the dispute explains the discourse of the opposing leaders. Thus the Habsburg regime saw its court appointments and religious policies over the years as legitimate exercises of the king's authority. The Habsburg King had the right to choose his officials, and they in turn had the right to exercise royal authority as they deemed appropriate. So all of the policies promulgated by the regime over the years had been legitimate. From the Bohemians' perspective, however, those policies constituted violations of their rights under the ancient traditions and customs of the kingdom – and that made the regime a tyranny that no longer deserved their obedience and allegiance. Hence they had the right to resist: their eviction of the royal officers and resort to arms constituted legitimate resistance to the unlawful actions of the royal officers. The resistance was justified because all previous efforts to achieve accountability by nonviolent means had failed. By contrast, the Habsburg regime held that the evictions of royal officers constituted sedition and the resort to arms constituted rebellion – both criminal acts under the existing legal system. Hence the regime had the right to punish such criminality: its military invasion of the Kingdom of Bohemia constituted a legitimate punishment of rebels who were defying the *Bohemian* king's authority. The punishment was justified because all of the regime's nonviolent calls for the Bohemians to cease-and-desist had been ignored.

### **6.3 The expected balance of power becomes a reality**

As the military campaigning season of 1618 ended, each side was more entrenched in its diplomatic position than ever and resolved to fight on. Over the winter each side looked for allies but found none outside the kingdoms of Bohemia and Austria.

In March 1619 Emperor Matthias died and Ferdinand assumed that he was the successor

as king of Bohemia. Accordingly he confirmed the *Letter of Majesty* to the Bohemian Protestants, commanded his army to suspend hostilities, and invited the Bohemians to send envoys to Vienna for negotiations. He also confirmed the rights of the Protestants of Silesia and upper Lusatia. Although Ferdinand was publicly committed to pursue negotiations, he privately admitted that his willingness to make concessions was limited and he still hoped to defeat the Bohemians by military force.<sup>118</sup> Perhaps sensing this, the directory in Prague refused his confirmation of the *Letter of Majesty*. In a sign that the Bohemian problem was spilling over into Austria, the Protestant estates of upper Austria refused to recognize Ferdinand's succession as ruler of Austria. And to signal their willingness to defend their religious rights by force, they established a directory, held up Ferdinand's revenue from his domain lands in upper Austria, and entered a defensive Confederation with Bohemia (on April 4).<sup>119</sup> Emboldened by their firm stance, the Protestant estates of lower Austria refused to recognize Ferdinand's succession as ruler of Austria unless he guaranteed their religious and constitutional rights. They also held that until he did so an *interregnum* existed and they had the right to rule the land.<sup>120</sup>

The other provinces of the kingdom of Bohemia – Silesia, Moravia and Lusatia – had remained on the sidelines up to this point. Now they joined the main province – Bohemia proper – in a defensive Confederation. The province of Moravia signed an alliance treaty with Bohemia (province) on May 11.<sup>121</sup> In response the Spanish government released 7000 troops from the Netherlands to march to Ferdinand's aid.<sup>122</sup>

The Habsburg regime now faced the most serious military threat up to this point in the war. Not only was the regime at war with the Kingdom of Bohemia, a realm rich in resources, but the majority of nobles in the Kingdom of Austria was taking up arms against the regime as well.

In June of 1619, a Bohemian force of 15,000 troops, commanded by Count Heinrich

---

<sup>118</sup>Bireley, 2014: 95.

<sup>119</sup>MacHardy, 2003: 71. Polisensky, 1971: 106.

<sup>120</sup>MacHardy, 2003: 71.

<sup>121</sup>Pages, 1970: 62.

<sup>122</sup>Bireley, 2014: 96.

Thurn, advanced deep into lower Austria and prepared to besiege Vienna. Since the city was defended by only 300 Habsburg troops, Ferdinand ordered a nearby garrison to dispatch a relief force. Just as the Bohemian force besieged Vienna, Ferdinand was meeting in his palace with the *Austrian* Protestant Estates to negotiate the conditions under which they would recognize his succession as king of Austria. He stood firm and was saved at the last minute by three strokes of military luck. First, the relief force arrived and sent the members of the Protestant Estates scurrying for cover. (They soon quit the negotiations with Ferdinand in Vienna and formed their own confessional estates in a nearby town.<sup>123</sup>) Second, a force of 2800 Hungarian troops arrived to fend off the Bohemian siege army. Third, another branch of the Bohemian army, under Count Ernst Mansfeld, had been trying to cut off the 7000 Spanish troops arriving from the Netherlands. But Mansfeld's force was defeated by Ferdinand's main army in Bohemia. In response the Directorate in Prague commanded Thurn to lift the siege of Vienna and retreat to Bohemia. <sup>124</sup>

Meanwhile, the Spanish Council of State was debating whether to increase its military support to Austria (from the level it had committed the previous August). On June 1, 1619, it decided to send 4000 troops to Austria; and on June 18 it decided to send another 2000 troops.<sup>125</sup> The Council made these decisions before finding out that the Bohemian army had advanced into Austria and laid siege to Vienna. In turn, Ferdinand decided to stand firm at the siege of Vienna before he found out that the Spanish Council of State had decided to increase its military support.

The traditional view is that the Bohemian advance into Austria and siege of the Vienna was a high risk gamble by the Bohemian military leader, Mathias Thurn, and it was bound to fail because his force was not well organized, disciplined or equipped. On this view the Directorate's decision to lift the siege and call Thurn's force back to Bohemia marked a recognition that he had overreached. Another interpretation is equally plausible, however. If it were not for the arrival of the 7000 Spanish troops from the Netherlands, the Bohemian army would have won the war at the siege of Vienna. There are two ways to see this point.

---

<sup>123</sup>Bireley, 2014: 98.

<sup>124</sup>Bireley, 2014: 96-97.

<sup>125</sup>Brightwell 1982, 125, 134, note 42.



First, if the 7000 Spanish troops from the Netherlands had never arrived, then Mansfeld would not have needed to divert his force to intercept them. Then Thurn's force could have maintained the siege of Vienna and extracted peace terms from Ferdinand. Second, even though the 7000 Spanish troops did arrive and Mansfeld's force did need to cut them off, his force might not have been defeated in the process. Then again, Thurn's force could have maintained the siege of Vienna and extracted peace terms from Ferdinand. Either way one looks at it, the Bohemian army came very close to winning the war: The only thing that saved Ferdinand's regime was the arrival of the 7000 Spanish troops from the Netherlands.

This fact turns the traditional view on its head and returns the focus to the more fundamental sources of relative power between Bohemia and Austria. Given the sub-optimal condition of the Bohemian army, it could *only* have advanced into lower Austria and laid siege to Vienna on the basis of an underlying power advantage due to a more fundamental factor. This factor was the broad base of its domestic political and financial support – relative to the narrow base of the Habsburgs' domestic support. From this perspective the need to lift the siege of Vienna was just a stroke of bad military luck. It might well have gone the other way. The key point is that the Bohemian army was *able* to advance deep into Austria and lay siege to Vienna, whereas the Habsburg army was not able to advance deep into Bohemia and lay siege to Prague.

#### **6.4 Phase 3: Finalizing the terms of the dispute**

During the first seven months of the war the Bohemians remained willing to accept Ferdinand as their king on more or less the same terms that applied before the war – provided he adhere to the terms as they interpreted them. They signaling their willingness to reach a settlement on such terms repeatedly. By the spring of 1619, however, they realized that Ferdinand would not accept those terms. He was committed to using his authority to suppress the Protestant religion and exclude its members from government office by force. Facing this threat, even the moderates among the Bohemian Protestants now preferred a political constitution that would deny the king real power and legitimize the nobility's right to resist a tyrannical king

by force (so the nobles would not be exiled for treason if they resisted one).<sup>126</sup>

In July of 1619, the five provinces of the crown of Bohemia formed a Confederation and issued "articles of Confederation" to specify the rights and privileges they were committed to defend by force.<sup>127</sup> The document specified a much more restricted role for the king and a much wider role in policymaking for his subjects. The most important clauses concerned the three areas of legal ambiguity discussed above:

1. The king's obligation not to issue new laws that violated the subjects' customary rights;
2. The nobility's right to approve or veto any new laws issued by the king; and
3. The nobility's right to control appointments to government offices.

The Bohemians still professed to be willing to accept Ferdinand as their king, but he would never accept the new terms.

On August 16, the Protestant Estates of upper and lower Austria joined the Confederation.<sup>128</sup> On August 19 the Bohemians deposed Ferdinand on the grounds that his election to the throne of their kingdom years earlier had not been done constitutionally.<sup>129</sup> The Bohemians elected Frederick, the ruler of a neighboring principality, as King of Bohemia. In response a number of Catholic rulers formed an alliance with Habsburg Austria to oust Frederick, defeat the constitutional movement and repress Protestantism in the Austrian Monarchy.<sup>130</sup>)

## **7. Rational calculations at the military abyss: Why Austria fought on**

If the crisis of 1618 had been like any of the previous crises between 1605 and 1611, it would have ended at the siege of Vienna in June 1619 – or soon thereafter. Ferdinand would have

---

<sup>126</sup>Steinberg 1966: 38. The radicals among the Bohemians wanted to do away with the monarchy and institute a republic.

<sup>127</sup>MacHardy, 2003: 72.

<sup>128</sup>MacHardy, 2003: 72.

<sup>129</sup>Bireley, 2014: 100.

<sup>130</sup>Steinberg 1966: 38-39.

recognized that he lacked the power to prevail and he would have settled for peace on the best terms he could get. As one historian noted, "Compromise was not completely out of the question for Ferdinand, as he had shown in negotiating the terms of his coronation in Prague a year earlier, but for him it was a last resort, involving wrestling with his religious conscience."<sup>131</sup> If there was ever a time that called for a "last resort", it was in June of 1619. The Bohemian army had besieged Vienna and trapped Ferdinand in his palace without a military defense. His position was practically identical to that of Rudolph in Prague in 1608: Overpowered by a large invasion force of Protestants, backed by a large majority of domestic Protestants, all wanting religious concessions. Why didn't Ferdinand concede defeat and settle for peace like Rudolph had?

A number of hypotheses are plausible, yet only one hypothesis is supported by the evidence – the one implied by the theory presented earlier.

The traditional hypothesis is that Ferdinand was a more committed religionist: he held out simply out of religious conviction. If true, this hypothesis implies that religious conviction outweighed calculations of relative power and material interest in Ferdinand's mind – so he would fight on even when the realities of power and interest dictated capitulation. This hypothesis is refuted by a few pieces of evidence. In 1629 Ferdinand found himself overpowered by Protestants demanding religious concessions – and he conceded defeat and settled for peace on their terms. One might object that this was only after 10 years of war that left Austria exhausted and fully aware of its powerlessness. Yet there is also evidence from Ferdinand's decisions in 1618 and 1619. First of all, even his clerical policy advisers gave decisive weight to power realities when the situation dictated it. Ferdinand may have shared the preferences of his clerical policy advisers as the traditional hypothesis claims, but their preferences were *realpolitik* in the final analysis. The Habsburg regime's chief minister, Cardinal Khlesl "planned Matthias's policies, including the strategic retreat of 1606-9. He also played the peacemaker vis-a-vis the increasingly minatory demeanor of German Catholics."<sup>132</sup>

A second hypothesis is that Ferdinand stood firm only when he had allies with the power

---

<sup>131</sup>Mortimer 2015, 161.

<sup>132</sup>Evans, 1979: 62.

to back his convictions. Thus, he stood firm in 1619 because he had a Spanish force at his disposal and expected Spain to send more troops soon. By contrast he capitulated in 1629 because he was out of allies and had no expectation that any new allies would appear. The evidence refutes this hypothesis too. In July of 1618 Ferdinand decided to initiate war even before hearing whether Spain would provide military support (beyond the 3000 troops already under Austrian control). For the next year Ferdinand stood firm diplomatically and continued to pursue the war even though Spain said it could not afford to send any reinforcements and exerted pressure on him to settle with the Bohemians. In June of 1619 Ferdinand stood firm against the siege of Vienna even before hearing whether Spain would provide further military support.

A third hypothesis is that Ferdinand stood firm because his state – Austria – was declining in power relative to the opposing state – Bohemia. So he had to fight a war to grab what he could grab before power shifted, or else his state would be vanquished.<sup>133</sup> If true, this hypothesis implies that Austria would have launched the war *and won it by preventing* Bohemia's power from rising to the point where it could vanquish Austria. The evidence says exactly the opposite. Austria did launch the war before Bohemia's power rose. Yet far from preventing Bohemia's rise and winning the war, Austria nearly lost everything when Bohemia's power continued to rise to the point that it could besiege Vienna.

A fourth hypothesis is that Ferdinand stood firm because he was uncertain about Bohemia's true power level. So Bohemia could bluff that it was more powerful than it really was – to gain more than it could ever gain based on its actual power. To deter such bluffing, according to this theory, Ferdinand's Austria would have to fight a war that would impose more costs on Bohemia than it might gain from such bluffing. Then it would be deterred from bluffing again in the future.<sup>134</sup> If true, this hypothesis implies one of the following two patterns as the war continued and Bohemia's war costs mounted – depending on whether Bohemia had been bluffing or telling the truth about its power before the outbreak of the war. On one hand, if Bohemia had been bluffing about its power, then as its war costs mounted, it would *reduce* its claims about its power and its diplomatic demands. On the other hand,

---

<sup>133</sup>See Powell (2006) for a general theory of war along these lines.

<sup>134</sup>See Fearon (1995) for a general theory of war along these lines.

if Bohemia had been telling the truth about its power, then once it had fought for long enough to prove that it was powerful enough to reap net gains from its power claims and diplomatic demands, Austria would stop fighting for good. The historical evidence refutes both implications. As the war continued and Bohemia's war costs mounted, it maintained its claims about its power and increased its diplomatic demands – yet Austria became *more* committed to fight.

Each of these hypotheses might be amended by suggesting that when the main antagonists – Bohemia and Austria – factored in the possibility of gaining allies, their decisions are explained by the hypothesis. Yet such amendments simply beg the questions underlying these hypotheses.

### **7.1 The period of fighting without allies comes to an end**

From the eruption of the war in August 1618 to the siege of Vienna in June 1619, each side fought essentially without allies. After that siege each side started to gain allies and a new phase of the war began. There were wide swings in territorial holdings from one side to the other, depending on which side had just gained or lost an ally. Yet each side's war aims were set by the summer of 1619. The Habsburg regime was fighting for a Bohemian political constitution that would give the king the right to issue laws without the nobles' consent and maintain complete control over government appointments. The Bohemians were fighting for a political constitution that would protect them from royal efforts to issue repressive laws without their consent and appoint government officers intent on destroying Protestantism. Neither side lost sight of its aims during the "Bohemian phase" of the war (from 1618 to 1621).

After the battle of the White Mountain in November 1620, Ferdinand simply equated Protestantism with disloyalty to the government. Since his counterreformation campaign in Inner Austria in 1598, he denied that Protestants had ever acquired "real public rights". Now he denied that they had any place in Austrian society and set out to eliminate them.<sup>135</sup>

This posture was dictated by the problem of Ferdinand's uncertainty about the Bohemi-

---

<sup>135</sup>Evans, 1979: 67-71.

ans' intentions: Were they opposing his authority intentionally to overthrow it or unintentionally as a byproduct of their goal of securing their religious freedoms? In the face of this uncertainty Ferdinand's optimal strategy, as shown by the theory presented earlier, was to impose war costs on the Bohemians until they stopped opposing his authority. *In his mind they would oppose his authority as long as they remained Protestant.* So the only way to stop them was to eliminate the Protestant confessions in his territories. Indeed he was right that they would oppose his authority as long as they remained Protestant. Yet that was *not* because they aimed intentionally to overthrow his authority, but rather because his regime was using its authority to deny their religious freedoms. If there had been some way to secure those freedoms while still accepting his authority, the Bohemian Protestants would have accepted it. In fact the history of toleration policies after 1648 proves there *were* ways that a ruler of one religion could coexist in the same territory with subjects of another religion. These institutional and organizational modalities were not conceivable in 1620, however.

Ferdinand's perceived need to eliminate Protestantism in his territories drove his political and military strategies from 1618 to 1621 – by which time he had largely succeeded in his goal. His commitment to this goal required him to form alliances with Bavaria and Spain, however, that provoked balancing responses and simply expanded the war. His commitment to *sustain* the suppression of Protestantism in his territories also required him to maintain these alliances and join his allies in battle wherever they had to fight against the balancers (for their own reasons). Without his regime's military participation his allies would have had to drop out of the war.<sup>136</sup> In this sense Ferdinand's need to eliminate Protestantism in his territories drove the entire war until 1635, at least, and possibly until 1648.

---

<sup>136</sup>By the same token, without their participation his regime would have had to drop out of the war.

## CHAPTER 21

### Case study: The outbreak of the French Revolutionary Wars

#### 1. Overview

##### 1.1 What is to be explained?

Scholars have long debated whether revolutions cause wars, and if so, why. A central case in the debate is the French Revolution of 1789 which erupted into a war with Austria in 1792. This chapter provides a new explanation of the war that differs from previous explanations. My theory explains more of the facts than the previous theories do and explains factors that the previous theories assumed to be independent causes of the war.

##### 1.2 The Contractual Realist explanation of the war

This theory assumes that the war was caused by the following factors:

1. the French revolutionaries' domestic goal of reducing the authority of the French monarchy and nobility;
2. the spread of this revolutionary ideology from France into Austria;
3. the Austrian elite's fear that it might cause a similar revolution in Austria that would reduce their authority and incomes;
4. the legal contradiction between the two states' rights under international law:
  - the Austrian government's right to defend the French king's divine right to rule (according to the pan-European law of monarchical government);
  - the French revolutionaries' right to domestic political independence and freedom from foreign interference in their domestic lawmaking process;

5. a second legal contradictions between the two states' rights under international law:
  - each state's right to host political émigrés from the other state;
  - each state's right *not* to be threatened with invasion by militant political émigrés harboring beyond its borders;
6. a third legal contradictions between the two states' rights under international law:
  - France's right to full jurisdiction over all land within France's borders;
  - the German princes' right to freedom from French encroachment on their estates in Alsace;

Because of the legal contradictions between their respective rights, each state's *exercise* of its legal right was perceived by the other state as the *violation* of its legal right. And each state's *claim* to have its right was perceived by the other state as an *intention* to violate its right, thus giving each state the suspicion that the other state was violating its right intentionally. If either state could have been certain that the other state was violating its right unintentionally, then it would have been better off making the concessions demanded by the other state than initiating a war. But because each state was uncertain whether the other state was violating its right intentionally or unintentionally - and suspected it was doing it intentionally - each state was better off rejecting the other state's demand and going to war over the matter than settling the matter peacefully, simply to protect itself from further intentional violations by the other state.

Each point in the escalation was triggered by one state's need to stand firm in defense of its legal rights under international law in the face of the other state's firm stance in defense of its overlapping (contradictory) legal rights under international law. Each time this occurred each state made some of the concessions demanded by the other state, but did not make all of the concessions demanded by the other for fear that giving up its rights would embolden the other state to violate its rights further. If there has been no legal contradiction between the two states' rights, each state would have been able to make all of the concessions demanded by the other without such fear. Thus it was the legal contradictions between their rights that caused the escalation point and the culmination of the process in war.

According to this theory, these factors alone were sufficient to cause the war. The other



causes suggested by the previous theories in the literature either fail to appear in the historical record of evidence at all or did not act as independent causes because they were driven by these factors.

### **1.3 Review and critique of previous explanations of the war**

The previous explanations divide into four different schools of thought.

#### **1.3.1 The "clash of principles" school**

This school assumes that the revolutionary principle of popular sovereignty in France was simply incompatible with the principle of absolute monarchy that undergirded the Austrian regime. A war between France and Austria was inevitable simply because the two principles of government differed so greatly. This difference made war inevitable for two reasons, according to this school. First, the mere existence of a French regime based on popular sovereignty threatened to inspire a similar domestic revolution in Austria. Second, the popular sovereignty of the French legislature provoked international legal disputes with Austria when the legislature tried to extend French jurisdiction into border areas whose feudal jurisdictions overlapped with German jurisdictions protected by the Holy Roman Empire (led by Austria).

This school of thought is correct to focus on the international spread of French ideology and the international legal disputes between France and Austria. The school is incorrect, however, in assuming that Austria had to oppose France simply as a matter of principle. While previous popular revolutions in the Dutch Netherlands and England did cause major wars, the French Revolution erupted into a war with Austria before the Revolution stepped over either of the lines that caused those previous wars. The English Revolution of 1688 only provoked a war when the English legislature deposed a sitting English King. The Dutch revolt of 1568 only expanded into a war with Spain after the Dutch repudiated their king and declared a republic – not only because they went republican but also because they were seceding from the Spanish empire. Between the outbreak of the French Revolution in 1789 and eruption of war in 1792, the revolutionaries neither deposed their king nor declared a republic. On the contrary, they reached a domestic constitutional settlement in 1791 that

established a constitutional monarchy and induced Austria and Prussia to conclude that the threat of revolutionary contagion was defused. The putative "clash of principles" was resolved by this settlement.

Rather it was the international legal disputes between France and Austria that caused the war – and not merely as a matter of principle. Many international legal disputes were resolved peacefully - even in the bloody 18th century. If it were only a matter of principle, then France and Austria would have resolved their disputes peacefully by splitting the difference over their principles. Each state was willing to make serious concessions to the other state in respect for its principles. This is evident from the observation that each state *did* make serious concessions to the other - to signal its willingness to reach an accommodation and its resolve to avoid war by any means possible. Rather, war was made inevitable by the existence of *legal contradictions* between the two states' rights under international law, such that each state's exercise of its right under one law was perceived as a violation of the other state's right under another law. Each state reasoned that if it were to give up its right, then the other state would continue to violate its right. Yet each state's resolve to stand firm for this reason struck the other state as a threat to continue violating its right! If it were not for such legal contradictions and the mutual political threats that they caused, France and Austria could have resolved their disputes peacefully.

In one of these contradictions, the French revolutionaries claimed the right to have domestic political independence and complete freedom of domestic lawmaking – including the right to abolish the French monarchy and establish a republic if it eventually came to that. Austria claimed that the pan-European law of monarchical government gave the French king the divine right to rule in France and Austria the freedom to protect that right – including the use of threats of force to induce the French revolutionaries to moderate their domestic political agenda. Although each piece of news suggesting that the French revolutionaries might abolish the monarchy and establish a republic raised fears of contagion in Austria, this factor alone did not cause the war. The escalation to war was also caused by two other legal contradictions: (1) the contradiction arising from the overlapping jurisdictions in the border areas and (2) the contradiction between each state's right to harbor political émigrés from the other state and each state's right not to be threatened by militant political émigrés

harboring beyond its borders. If it had not been for these legal causes of escalation, the French constitutional settlement of 1791 would have stabilized the revolution in France by enough to avoid a war with Austria.

### 1.3.2 The "balance of power" school

This school of thought assumes that the French Revolution created uncertainty about the balance of relative military power between France and Austria, inducing each state to overestimate its power and underestimate the other's power. Expecting military victory, each state became willing to initiate a war that it would have refused to initiate with a more accurate estimate of relative power. This explanation is incorrect on both a logical level and an empirical level. Logically, if the war had erupted due to inaccurate estimates of relative military power, then the war would have stopped quickly after the actual battles revealed to both sides the true balance of power.<sup>1</sup> Empirically, while there is evidence that some politicians in each state overestimated their state's power and underestimated the opposing state's power, there is evidence that other politicians had accurate estimates of the balance of power that made them argue strongly against initiating a war. In both Austria and France the politicians who argued against war for balance of power reasons received the strongest popular support – and only lost support when the international legal disputes between the two states intensified and rendered war necessary for reasons that had nothing to do with the balance of power. This pattern of evidence is consistent with the assumptions of my model. First, my model assumes that the revolution caused no shift in relative military power (or no shift large enough to give either state the expectation that initiating a war was preferable to reaching a peaceful accommodation). Second, my model assumes that the revolution did not cause any uncertainty about relative military power between the two states. Third, each state's uncertainty about the other state's intentions was caused *not* by any shift in relative military power or uncertainty about relative power, but rather by the legal contradictions that provoked the international legal disputes.

---

<sup>1</sup>Powell (2006) notes that wars caused by uncertainty about relative power should end quickly once the fighting resolves the uncertainty.

### 1.3.3 The "balance of threat" school

This school of thought assumes that Austria felt *politically* threatened by the spread of French republican ideology throughout Europe. And France felt *politically* threatened by the prospect of a counterrevolutionary invasion by Austria. According to this school each state overestimated the political threat from the other and, in conjunction with its overly optimistic estimate of the balance of relative military power, reasoned that it would be more secure initiating a war to combat the political threat than reaching a peaceful accommodation to defuse the threat. This explanation is incorrect not only for the reasons that the previous schools of thought are incorrect, but also for deeper logical and empirical reasons. Logically, there is always a peaceful settlement that both sides would prefer to a costly war,<sup>2</sup> even in cases where one state is imposing costly spillover effects on the other state. Empirically, there is strong evidence that such a peaceful settlement was reached when the French King accepted the constitution offered by the French legislature in 1791, thereby creating a constitutional monarchy and inducing Austria and Prussia to regard the threat of Revolutionary contagion as defused. Tensions between France and Austria over the spread of French ideology only escalated after that point because of the legal contradictions between their rights under international law. If it had not been for these contradictions, the constitutional settlement would have been sustainable. My model explains how the legal contradictions alone were both necessary and sufficient for a war over the spread of French ideology to occur. Without these contradictions, war would not have occurred. With them, war would have occurred even in the absence of any other factor (aside from the ideological spillovers themselves).

### 1.3.4 The "domestic motives" school

This school assumes that the revolutionary faction with strong domestic motives for the war with Austria – the *Brissotin* faction – convinced the rest of the legislative assembly, the Paris mob and the French people to initiate a war that they would not have initiated otherwise. According to this logic, if that faction had not had domestic motives for war - and had not made its case to everyone else so convincingly – there would have been no war. This school is

---

<sup>2</sup>Fearon (1995) makes this point generally.

incorrect for both logical and empirical reasons. Logically, it assumes that the international legal disputes between France and Austria were not sufficient in and of themselves to cause a war between them. Yet my model demonstrates that these disputes alone *were* necessary and sufficient to cause a war. In other words, even if no revolutionary faction in France had had any domestic motives for war, the international legal disputes alone would still have caused the war. The model explains key features of the historical evidence that cannot be explained by the domestic motives school (and even contradict this school).

Empirically, this school is contradicted by key evidence on the faction with domestic motives for war. First of all, the *Brissotin* faction broadcast its domestic rationales for a war loudly and regularly during the first two years of the revolution, but was rejected by the vast majority of listeners. Secondly, once this faction started to gain traction in domestic debates in the revolution's third year, its position was based on international motives as much as domestic motives.

The domestic motives school does not claim that the *Brissotin* faction's domestic rationales for war were alone sufficient to convince a majority of the legislative assembly or the nation to initiate a war. It does contend, however, that their domestic rationales were necessary to push a majority "over the hump" into initiating a war – and without hearing these domestic rationales presented so forcefully, the majority would not have initiated a war. This hypothesis ignores a number of other possibilities that might explain why a majority in the legislative assembly became convinced to initiate a war after hearing this faction's domestic rationales for war. First of all, in their the *Brissotin* faction members also gave international rationales for war, such as Austria's violations of France's rights under international law. Perhaps these rationales were the factor that convinced the majority to initiate a war. If this is how it happened, then the domestic motives for war were simply "piggybacking" on the international motives that were alone necessary and sufficient to provoke the assembly into a war.

In this scenario it was the strength of France's grievances against Austria that drove the nation "over the hump" into war. The *Brissotin* faction may have made the case against Austria more strongly than other factions that did not have domestic motives. Yet this does not mean that the *Brissotin's* case against Austria was convincing to the assembly because

of their domestic motives for war. It was the strength of French grievances against Austria that made the case against Austria so strong and convincing to the assembly. Put another way, if the case against Austria had not been so strong, the domestic motives for war would not have gotten the *Brissotin* faction anywhere with the assembly – just as their domestic motives *had* not gotten them anywhere over the previous two years. Thus the key question is whether the French grievances against Austria – alone – could have been necessary and sufficient to drive France into a war. To answer this question my model formalizes the logic by which such grievances can be necessary and sufficient to cause a war. The case study in this chapter provides the historical evidence that corroborates the model's observable implications.

While the *Brissotin* faction argued for a war for their domestic reasons, many other politicians and public opinion makers argued against a war for a number of reasons. Between 1789 and 1791 this debate was won overwhelmingly by those who were opposed to a war. This fact suggests that the vast majority of the assembly and the nation wanted to avoid a war for exactly the reasons the antiwar leaders said war should be avoided. It was only when the international legal disputes between France and Austria escalated in 1791 that the antiwar majority changed its mind. This constitutes *prima facie* evidence that these disputes caused the about-face. Naturally one might expect that when these disputes escalated in 1791, the domestic faction with the most reasons to want a war would make the strongest case against Austria. Yet even still, there was a strong sentiment against the war in the assembly and the nation through the fall of 1791 and into 1792 – for the same reasons the majority had wanted to avoid a war over the previous two years. The antiwar leaders continued to make the case against a war through the fall of 1791 and into 1792 – and many in the assembly and the nation remained hesitant to launch a war despite the mounting legal case against Austria. The undecided majority was looking for a way to avoid war if possible, and only decided to favor a war when the case against Austria became undeniably strong.

The domestic motives school is incorrect for additional reasons. This school assumes that the domestic motives for war were different than the international motives for war that motivated most of the assembly and the nation. Yet this assumption is neither logically necessary nor empirically supported by the historical evidence. The standard definition of

"domestic motives" is that the *Brissotin* faction wanted war because it would provide an opportunity for them and the rest of the revolutionaries to defeat the domestic enemies of the revolution. From this perspective, the war was only an excuse or pretext for the *Brissotin* faction's domestic political agenda.

This is not the only way way to understand the connection between that faction's hostility to Austria and their hostility to the domestic enemies of the revolution, however. Another way to understand it is to assume the reverse – that the *Brissotin* faction was so hostile to the revolution's domestic enemies because they were pursuing the Austrians' political agenda of crushing the revolutionaries and restoring the absolute monarchy. Insofar as the revolution's domestic enemies were allied with Austria, a war against those domestic enemies *was* a war against Austria. The *Brissotin* faction assumed that a war against Austria would make it easier to defeat the revolution's domestic enemies. Yet they also believed the revolution's domestic enemies were in league with Austria, and as long as the domestic enemies went undefeated, Austria would continue to threaten a counterrevolutionary invasion. So the *Brissotin* faction assumed a war against the revolution's domestic enemies would make it easier to defeat the Austrian threat of a counterrevolutionary invasion.

The revolution's foreign and domestic enemies had the same goal – to crush the revolution and restore the absolute monarchy. Analytically, therefore, one might assume *either* that defeating the domestic enemies was a means to the end of defeating the foreign enemies *or* that defeating the foreign enemies was a means to the end of defeating that domestic enemies. If one chooses the first interpretation – that defeating the domestic enemies was a means to the end of defeating the foreign enemies – then the *Brissotin* faction's "domestic motives" were simply a part of their broader motive of defeating Austria. Even if one chooses the second interpretation – that defeating Austria was a means to the end of defeating the revolution's domestic enemies – there were still two ways to diffuse the Austrian threat. One way was through war: The *Brissotin* faction claimed that war with Austria was necessary as a means to the end of defeating the revolution's domestic enemies. Another way was by negotiating a peaceful settlement of the international legal disputes between France and Austria. Many French politicians and commentators argued that this approach was necessary to conserve resources for the internal fight against the revolution's domestic enemies. The

peace-with-Austria argument was more convincing to a majority of the assembly and the nation than the war-with-Austria argument for the first two years of the revolution (from 1789 to 1791). Even after the crisis with Austria began to escalate in the fall of 1791, a majority of the French assembly and French popular opinion found this argument convincing enough (along with other arguments) that they hesitated to lurch into a war with Austria.

Once the international legal disputes with Austria escalated in the fall of 1791, a majority of the French assembly continued to hope and work for peace with Austria, suggesting they still thought the peace-with-Austria strategy would be a better means of defeating the domestic enemies of the revolution than the war-with-Austria strategy. Why did a majority changed its mind over the next few months? Was it because the *Brissotin* faction came up with better logic and evidence to convince the assembly that war with Austria was a better means to defeat the revolution's domestic enemies – and a majority in the assembly decided to go to war against Austria for this reason? Or was it because a majority decided war with Austria was necessary as a means to the end of defeating the Austrian threat to the revolution?

### **1.3.5 Political science literature on revolution and war**

The political science literature on the relationship between revolution and war holds a prominent place for the French Revolutionary War. In a seminal article that spawned much of this literature, Walt (1992) presented a number of causal mechanisms by which a revolution might increase the likelihood of war and illustrated many of them with evidence from the French Revolutionary War. After reading so many different reasons why a revolution might cause a war, the reader of the article is tempted to throw up his hands and say, "Well, if one of these factors did not cause the French Revolutionary War, then certainly another one of these factors did." As Walt wrote, "War is over-determined." In statistical analyses and case studies over the next 15 years, political scientists seemed to confirm his general finding that revolutions cause wars. Eventually, however, Narang and Nelson (2009) refuted the conventional wisdom by presenting a statistical analysis showing that democratic revolutions like the one in France in 1789 *rarely* cause wars. Their study naturally raises the question whether the evidence that Walt culled from the history of the French Revolution between



1789 and 1792 really supports his claims about the reasons that revolutions cause wars in general and the reasons that this revolution caused this war. My model is well suited to answer this question because it disentangles Walt's three main claims about why revolution causes war, discards two of them, and turns the third into its opposite. If the numerous observable implications generated by my model are corroborated by the evidence presented in this chapter, it would disprove all of Walt's main claims, demonstrate what really caused the French Revolutionary war, and explain why this case is such an anomaly in a statistical universe of cases where democratic revolutions rarely cause wars. Therefore, it is worth clarifying how my explanation of this war differs from Walt's three explanations.

First, Walt assumes that the French Revolution caused a shift in the balance of relative military power between France and Austria and uncertainty about the balance between them. By contrast my model assumes that the revolution caused no shift in the balance of relative military power between France and Austria (or at least no shift large enough to give either state the incentive to initiate a war) and no uncertainty about the balance.

Second, Walt assumes that the *Brissotin* faction in France wanted a war for domestic political reasons and convinced the rest of Revolutionary France to initiate a war that it would not have initiated otherwise. By contrast my model assumes that while some domestic actors may have had domestic political reasons for wanting a war, no such actors or motives played any *causal* role in the escalation of the crisis or the initiation of war. In technical terms, such actors and motives were neither necessary nor sufficient to cause the war.

Third, Walt assumes that each state was uncertain about the *magnitude* of the political threat that the other state posed to its domestic interests. Vienna was uncertain whether the spread of radical French ideas would provoke only moderate unrest in Austria or a radical revolution in Austria. France was uncertain whether Austria would apply only mild pressure to moderate the revolution or launch a counterrevolutionary invasion to overturn the revolution entirely. By contrast, my model distinguishes between two different kinds of uncertainty that *conceivably* might have arisen between France and Austria. I allow the historical evidence to determine whether, when and why each kind of uncertainty *actually* arose.

First, each state might have been uncertain about the *magnitude* of the domestic political

shift taking place within the other state. Austria might have been uncertain about how moderate or radical the French revolution was at each stage of its unfolding between 1789 and 1792. On this point the evidence shows there was little or no uncertainty in Vienna. Austrian leaders made accurate estimates of how moderate or radical the revolution was at each stage of its development – and accurate updates to their estimates each time the revolution became more moderate or more radical. In turn, France might have been uncertain about whether the spread of its ideology of popular sovereignty would provoke only moderate unrest in Austria or a radical republican revolution. On this point the evidence shows there was little or no uncertainty in France. French leaders made accurate estimates that French ideas were not provoking widespread movements for a radical republican revolutionary in Austria – and factional leaders in France who claimed otherwise were not widely believed.

Second, each state might have been uncertain about the *intentionality* of the political threat that the other state posed to its domestic interests, e.g. whether the other state was posing the threat intentionally or unintentionally. On this point the evidence shows there was mutual uncertainty. Austria was uncertain whether French ideas were spreading into Europe intentionally or unintentionally on France's part. And France was uncertain whether Austria was honestly and unwittingly ignorant that French ideas were spreading into Europe unintentionally, or purposely ignoring evidence it was unintentional on France's part to be able to claim it was intentional and demand a moderation of the revolution. Because of this second kind of uncertainty, Austria suspected that France was intentionally trying to spread its ideas internationally to provoke a revolution in Austria for its own gain. And France suspected that Austria was intentionally trying to interfere in French domestic politics for its own gain.

One way for each state to deter the other state from engaging in such opportunism would be to fight a war that would impose more war costs on the other state than it might gain from such opportunism. This strategy would work as long as the revolution in France remained moderate. But if the revolution became too radical, then neither state would be able to deter such opportunism by the other state in this way. To see the logic consider a simple analogy. A fine of 100 is sufficient to deter a theft of 80, but not sufficient to deter a theft of 120. If the opportunist steals 120 but is only fined 100, then he gains 20 on net and he has an

incentive to steal more. By analogy, as long as the revolution in France remained moderate, Austria's existing military power would be sufficient to impose more war costs on France than it might gain from provoking an equally moderate revolution in Austria. But if the revolution in France were to become too radical, then Austria's power would not be sufficient to impose more war costs on France than it might gain from provoking an equally radical revolution in Austria. The same calculus applied in reverse. As long as the revolution in France remained moderate enough, France's existing military power would be sufficient to impose more war costs on Austria than it might gain from interfering in French domestic politics to reverse such a moderate revolution. But if the revolution were to become too radical, then France's power would not be sufficient to impose more war costs on Austria than it might gain from invading to reverse such a radical revolution.

When the French Revolution took a radical turn in 1791, it was accurately perceived in Vienna. But the Austrian leadership faced a difficult choice between two options. On one hand, it could allow the revolution to continue to become even more radical, in which case Austria's existing military power might become insufficient to deter France from intentionally provoking a similarly radical revolution in Austria. On the other hand, Austria could demand a moderation of the revolution and threaten to initiate a war if the demand were not met - in order to ensure the revolution did not become so radical as to eviscerate its deterrent posture. Austria chose the second option. In response the French revolutionaries faced a similar choice between two options.

In this escalation process the key causal factor was the mutual uncertainty about intentions: Austria's uncertainty about whether French ideas were spreading into Europe intentionally or unintentionally on France's part; and French uncertainty about whether Austria was honestly and unwittingly ignorant the ideas were spreading unintentionally on France's part, or purposely ignoring evidence they were spreading unintentionally to be able to demand a moderation of the revolution for its own gain. If it had not been for this mutual uncertainty, at least one state would have been better off making the concessions demanded by the other instead of initiating a war.

The key question, therefore, is what caused the mutual uncertainty about intentions. The historical record provides considerable evidence that the uncertainty was caused by

the legal contradictions between the two states' rights under international law. Aside from these legal contradictions, each state's leaders were inclined to believe that the other state, while not friendly, was restraining itself from intentional opportunism and keeping a lid on the domestic elements which harbored suspicions of the other state's intentions. The revolutionary legislature in Paris issued multiple declarations renouncing the right to engage in foreign conquest. These renunciations – and the complementary reports of Austrian diplomats – were taken at face value in Vienna *except* when the legal contradictions between the two states' rights under international law provoked disputes that cast doubt on such renunciations. In turn, the Austrian regime professed benign intentions toward the revolution were believed in France *except* when the legal contradictions provoked Austrian actions that cast doubt on such professions.

#### **1.4 Patterns of evidence to be demonstrated in this chapter**

I derived the observable implications of my theory in a previous chapter. The aim of this chapter is to corroborate the implications by demonstrating the following patterns of historical evidence.

1. Tensions between the upper and lower classes in Austria were high before the French Revolution broke out and remained high continuously until the war erupted (and beyond).
2. From the start of the French Revolution in 1789, each of the two opposing powers, France and Austria, feared that the other would interfere in its domestic politics. Elites in Austria were concerned about the spread of revolutionary ideology into Austria. And the revolutionaries in France were concerned that Austria would take military action to overturn the revolution.
3. As long as the Revolution remained moderate (1789-91), Austrian foreign policymakers were not in a rush to go to war against France; they rejected the idea of initiating war; and they maintained an accommodating posture toward France.
4. The mutual fears of interference in each other's domestic politics only turned into a *dispute* when the French revolution turned radical (1791).
5. Austria only changed its posture toward France from accommodating to demanding

when the French revolution turned radical (1791).

6. The purpose of Austria's demands and military threats was to induce France to moderate its domestic revolution.
7. The moderates in France wanted to respond with reassurances but the hardliners accused them of caving in to foreign threats.
8. The hardliners won the domestic debate and France responded to Austria's demands and threats with intransigence and confrontation.
9. France started to issue demands and threats to Austria within four or five months of the radical turn in France's revolution (October or November 1791).

## 2. Developments before the outbreak of the French Revolution

### 2.1 France before 1789

In the decades before 1789, French society and government faced a mounting crisis. The main causes were animosity toward the nobility's system of feudal privilege, mounting fiscal crises generated by that system, and frustration with the government's inability to reform the system. The *bourgeoisie*, composed of middle class professionals and merchants, resented the nobility for its exclusive control of social status, judicial office, political representation and tax exemptions.<sup>3</sup> The *bourgeoisie* also resented the government for its enforcement of guild restrictions and trade restrictions on agricultural markets.<sup>4</sup> The peasant class resented the nobility for its high rates of feudal taxation, abuse of judicial authority, control of natural resources, and immunity from taxation.<sup>5</sup> The absolute monarchy made several attempts to reform the feudal system and tax the nobility before 1789, but foundered on the nobility's unwillingness to give up its privileges and tax exemptions. When the monarchy tried to impose reforms on the noble assemblies in the provinces, they defied it and insisted that a general assembly, or *Estates-General*, be convened in Paris with representatives from the clergy, nobility and bourgeoisie of all provinces. In July 1788 King Louis XVI announced that

---

<sup>3</sup>McPhee 2002, pp. 24–6, 31–2, 41.

<sup>4</sup>McPhee 2002, p. 29.

<sup>5</sup>McPhee 2002, pp. 33–4, 41, 43–4.

a general assembly would convene in May 1789.<sup>6</sup> By this point a combination of agricultural depression, government fiscal crisis and previous failures to pass reforms were raising tensions across all levels of society and government to a critical point. By December 1788 peasants were refusing to pay feudal taxes to their lords. By the spring of 1789 popular militias and councils were seizing control of local governments and nobles were being removed from office in some of the provincial centers.<sup>7</sup> As the date of the assembly approached, all sectors of French society were required to elect representatives and submit their grievances and proposals for reform in writing. The ensuing public debate revealed that the main points of contention remained the same – the nobility’s refusal to give up its tax exemptions, control of resources and monopolization of power in the provincial assemblies.<sup>8</sup> Although the peasant and working classes would not be represented in the general assembly, they made clear that the bourgeoisie would speak and act in their name.<sup>9</sup>

### **3. Austria before 1789**

#### **3.1 Summary**

Relations between the upper and lower classes in Austria were tense in the decades before the French Revolution. But this was not because of failures to achieve reform, as in France. Rather the tensions were raised by the impact of fundamental reforms. These reforms made the cleavage between the upper and lower classes a major factor in domestic politics. Starting in the 1750s the Austrian rulers initiated a series of reforms designed to alleviate the feudal burdens of the peasants and prevent their exploitation by the lords. For the first few decades the central government in Vienna left the provincial nobility in control of some aspects of local government and taxation. But in 1780 Empress Maria Theresa was succeeded by her son, Joseph II, whose reform agenda was more ambitious. He instituted reforms to give himself absolute control over both the central and provincial governments, leaving the

---

<sup>6</sup>McPhee 2002, pp. 35–7.

<sup>7</sup>McPhee 2002, pp. 56–7.

<sup>8</sup>McPhee 2002, pp. 41, 46.

<sup>9</sup>McPhee 2002, p. 53.

nobility with almost no power. He also established greater protections for the peasants and more restrictions on the nobility's rights to exploit them. These reforms antagonized the nobility to the point that they mounted a conservative reaction to regain their rights and privileges from Joseph's absolutist regime. At the same time the peasant reforms raised the hopes of the lower classes for complete emancipation from the feudal system as well as the hopes of the middle classes for a transition to the rule of law.

When war broke out with the Turks in 1787, the tensions on all sides of the domestic debates over reform rose due to the added burdens of conscription, taxation, inflation and wartime shortages. Nonetheless, in early 1789, only months before the outbreak of the French Revolution, Joseph persisted in pushing through his most ambitious reform statute: an increase in the government's tax rate on the nobility and a reduction in the tax rate the nobility could assess on the peasantry. The measure received strong opposition not only from the nobility but even from Joseph's top officials. The Chancellor of Bohemia raised a range of objections including one that revealed the nobility's fear that it lacked the power to contain a restive peasantry in the event of class tensions.<sup>10</sup> "While the measure is being implemented, there are bound to be disorders, which there are not enough troops to suppress."<sup>11</sup> The nobles dug in their heels against both the regime and the peasants – demanding an end to the peasant reforms and abandonment of the detestable tax reform. At the same time many peasants refused to commit to the new, lower tax rates enacted by the reform, since they did not want to make any commitments that would hinder their campaign for even *lower* taxes and complete emancipation from the feudal system. Thus, on the eve of the French Revolution, relations between the upper and lower classes in Austria were tense and getting tenser. The regime of Emperor Joseph II was caught between two opposing forces, each of which was on the brink of rebelling against the regime to increase its share of the profits of the agricultural economy at the expense of the other.

---

<sup>10</sup>Beales 2009, pp. 593–7.

<sup>11</sup>Beales 2009, pp. 594–5.

### 3.2 Narrative

Austria nearly lost the War of the Austrian Succession (1740-48). To avoid another debacle its ruler initiated a series of reforms intended to rationalize the state bureaucracy, develop the economy, improve the condition of the peasantry, and increase tax revenues.<sup>12</sup> Empress Maria Teresa put limits on the nobility's power to extract labor services from the peasants, violate their legal rights, and dominate justice in local manor courts. While the nobility retained the rights to control some local offices and reject new tax requests, the monarchy took control of key aspects of provincial government.<sup>13</sup> When Joseph II succeeded as monarch in 1780, he continued the reform project and added a new dimension that transformed it entirely. He decreed further reductions in the nobility's power over peasants: tighter limits on the penalties lords could impose on peasants, protections against the worst forms of personal subjections, and extensions of the right to receive legal redress from abuses.<sup>14</sup> Contrary to his mother Maria Theresa, though, Joseph believed the monarch should hold absolute power without check or balance from the nobility. While he shared her belief that the monarch should be concerned about the welfare of all subjects, he believed that he had the *sole* right to decide what would serve their welfare. Accordingly he reduced the rights and privileges of the nobility and concentrated government power in the state bureaucracy under his absolute control.<sup>15</sup> As authority shifted from the noble assemblies in the provinces to the center in Vienna, the composition of the bureaucracy shifted away from the nobility to a new class of non-noble administrators. These civil servants were educated and trained to prioritize economic development and social equity between lords and peasants. Promoted by merit and unified by their ruler's goal of improving the welfare of all subjects, they developed a sense of social solidarity and purpose that made them a powerful pressure group in favor of reforms.<sup>16</sup>

---

<sup>12</sup>Macartney 1978, pp. 2016–17.

<sup>13</sup>Macartney 1978: 2-5, 14.

<sup>14</sup>Macartney 1978: 16-17.

<sup>15</sup>Macartney 1978: 1, 13-14.

<sup>16</sup>Macartney 1978: 14-15. Wangermann 1969: 10. Note that the organization of Wangermann's study is chronological in some respects and topical in other respects, so drawing evidence about the sequence of events in Austria requires cross-referencing between different chapters and sections.



From the start of the reforms, the nobility suspected that the aspirations of the lower classes would be difficult to contain once raised. They were right. The reforms raised the political consciousness and aspirations of the peasants, middle-class professionals and government administrators to the point that they conceived reform differently than Joseph did.<sup>17</sup> After the regime instituted reforms to reduce noble privileges, the middle class contended that the privileges should be eliminated entirely. One member of a task force commissioned to codify the country's laws issued a report calling for institution of the rule of law, so that subjects could hold the ruler and the nobles accountable to law, and the ruler would have to obtain the subjects' consent to issue new laws.<sup>18</sup> All of these developments antagonized the nobility and drove them into staunch opposition to Joseph's reforms. Even before the French revolution erupted, the nobles mounted a conservative reaction and pressed the regime to abandon its reforms.<sup>19</sup>

When war erupted between Austria and Turkey in 1787, the increased burdens of taxation, conscription, inflation and war-time shortages raised tensions on all sides of the domestic debates over reform. In response Joseph tightened censorship and increased the powers of the secret police to monitor discontent and prevent it from turning into disorder.<sup>20</sup> Joseph's need for new taxes to fund the war gave the nobles apparent leverage to oppose reforms, however. The Hungarian nobility forced Joseph to reverse most of the reforms while the Belgian noble and commercial classes, not used to a ruler having absolute power over them, mounted a revolution with some support from the lower and middle classes.<sup>21</sup> In late 1788, when the commercial classes in two Belgian provinces refused to grant war taxes to his regime, Joseph cracked down. He crushed their demonstrations with military force, dismissed their Estates, and annulled their provincial constitutions which had given the noble and commercial classes the right to oppose his measures – replacing it with an absolutist one that gave been no such

---

<sup>17</sup>Wangermann 1969: 3, 6, 11-12, 24.

<sup>18</sup>Wangermann 1969: 12, 14, 20-23.

<sup>19</sup>Wangermann 1969: 49-50. He cites the opposition of the nobility in Hungary and Belgium, both areas under the authority of the Austrian ruler.

<sup>20</sup>Beales 2009: 598-9. Macartney 1978: 19. Wangermann 1969: 27-30.

<sup>21</sup>Macartney 1978: 19-20.

right.<sup>22</sup>

Despite the noble opposition in Hungary and Belgium, Joseph pressed ahead with his most ambitious reform plan in February 1789 – increasing the land tax rate on the nobles and reducing the nobles' tax rate on the peasants throughout his dominions. The nobles in all provinces of the Monarchy dug in their heels against the regime and the peasants – demanding an end to the tax reform and the other peasant reforms. This noble opposition extended even to the provinces of central Austria and Bohemia where the nobility had been stripped of more rights, privileges and political power than in any of the other provinces.<sup>23</sup> Perceiving this reform quite differently, many peasants refused to accept the new *lower* tax rate because they expected that their obligations to the lords would be reduced even further in the future – so they did not want to be legally bound.<sup>24</sup>

In the spring of 1789, a few months before the French revolution erupted, Austrian government and society were in great turmoil. The realm was at war with Turkey. The peasantry was hoping that Joseph's regime would emancipate them from feudal oppression by the lords. In turn the lords were mounting a conservative reaction against the reforms that Joseph had pushed through to date.

## **4. The French Revolution erupts (July 1789 to February 1790)**

### **4.1 Outbreak and goals of the revolution**

In May of 1789 the general assembly announced by King Louis XVI convened in Versailles. The nobility insisted on the conventional structure for deliberation and voting in three separate chambers, one for each order of society: clergy, nobility and bourgeoisie. Since this would give the nobility disproportionate weight in the voting, the bourgeoisie insisted the voting be done jointly in a single chamber. When Louis sided with the nobility and blocked the bourgeoisie out of their own chamber, they simply met elsewhere on their own initiative. Louis tried to accommodate them by offering a minor tax reform that would leave the nobil-

---

<sup>22</sup>Beales 2009, p. 605.

<sup>23</sup>Beales 2009, pp. 623–4.

<sup>24</sup>Wangermann 1969: 30, 32, 68.

ity's feudal privileges intact and preserve its voting advantage. When they rejected it Louis sent 20,000 troops into Paris and moved to have the bourgeoisie dismissed from the general assembly. The working classes of Paris, fearing that the noble class was raising the price of bread and storing up arms to intimidate the other classes, went into the streets to search for hoarded grain, seize garrisons and confront the troops. On July 14 they seized the Bastille garrison and instituted a new city government for Paris, controlled by the bourgeoisie and protected by a bourgeois civil militia. Louis acquiesced and withdrew the troops.<sup>25</sup>

The monarchy's capitulation triggered a wave of local revolutions across the country in which peasant militias forced lords to surrender feudal tax contracts for shredding and burning. By August 1789 the nobles in the general assembly decided to renounce their feudal rights and privileges and help the assembly pass "Article 1" which abolished the feudal system of noble privilege and exploitative taxation. At the same time the assembly passed articles to abolish manorial courts, guarantee equality before the law, and establish free access to justice. When Louis rejected these articles – and the Declaration of the Rights of Man that accompanied them – the working women of Paris, followed by the National Guard, marched to the general assembly in Versailles to demand that Louis accept the articles and move to Paris.<sup>26</sup> Louis's compliance with these demands brought the revolution to a new phase defined by the search for agreement on a constitution. The royalist party's goal was to preserve as much power as possible for the monarchy and nobility through a combination of concessions and credible threats to mount a counterrevolution if their domestic opponents demanded too much. The revolutionary party's goal was to constrain royal and noble power as far as possible without demanding so much as to provoke a counterrevolution.

The final form of the constitution would not be clear for some time. The problem was not simply that there were many issues to deliberate in the assembly and settle with the monarchy, but also that the revolutionaries were divided among themselves. While most of the revolutionary politicians and bourgeoisie wanted to maintain the monarchy and add an elected parliament to restrain it, a vocal minority wanted to eliminate the monarchy and

---

<sup>25</sup>McPhee 2002, pp. 52–5.

<sup>26</sup>McPhee 2002, pp. 57–61.

establish a republic.<sup>27</sup> Both of these factions wanted to maintain the private property rights of the nobility. But many peasants wanted to expropriate the nobility's property without compensation by denying lords the right to collect feudal dues on peasant tenancies, seizing forest land that was customarily treated as a collective commons, and distributing wasteland freely. By the fall of 1789 the king had to declare that forest seizures would be punished – and the constituent assembly had to declare that forests were now owned by the nation, so peasants could only occupy them after obtaining a favorable court verdict. In the spring of 1790 the assembly had to pass legislation requiring peasants to prove that they had never agreed to pay feudal dues in the first place if they wanted to stop paying them now – and to continue paying them until they could prove it. Many peasant communities tried to prove this over the next few years – and refused to pay their dues while their cases were pending.<sup>28</sup> All in all, it took months for the disagreements among the revolutionaries to become clear and years for them to be resolved. From the very start of the revolution, the uncertainty about how radical it would turn out to be hampered efforts to reach a constitutional settlement and, consequently, left the other conservative monarchies of Europe wondering how seriously to view the threat of French radicalism and the possibility that it would spill over to their lands.

## **4.2 France's initial posture toward Austria**

France's posture toward Austria was driven by a number of conflicting factors, because attitudes toward Austria varied widely across the segments of French society and government. Many nobles wanted Austria to launch a military intervention to overturn the revolution in concert with Prussia and Russia. The nobility and the royalist party in Paris expected that the conservative powers would intervene if the revolution became too radical. They also knew that the Austrian leadership was more concerned about the fate of the French monarchy than the other powers, so that was where to put the pressure for intervention. Outside the regime and nobility, many in France expected a great power intervention to occur imminently but felt it would violate the fundamental norm prohibiting foreign interference in domestic politics.

---

<sup>27</sup>Doyle 2002, pp. 134–5; Forrest 1995, pp. 42, 48–9.

<sup>28</sup>McPhee 2002, pp. 69–71, 80–2.

Among the revolutionaries in the constituent assembly, opinion was divided. Most of them were vehemently anti-Austrian because Vienna represented the old feudal regime of Europe and posed a threat of military intervention against the revolution.. Vienna was not issuing any threats for the moment, however, so some of the revolutionaries figured that Austria did not want a war and would stay out of French domestic politics.

Although most of the French revolutionaries did not want to interfere in the domestic politics of other countries, some of them discussed the potential benefits of doing so and moved in that direction. A few months after the revolution erupted, one of the more prominent revolutionaries sent an envoy to Belgium – a province of the Austrian monarchy – to develop connections with democratic activists who were trying to broaden the base of the revolution transpiring there. The envoy promised that France would provide military support if the conservative powers intervened to crush the revolution. At the same time some French revolutionaries approached the Prussian government to inquire if it would tolerate this kind of interference in Belgium. When Prussia and Austria finally intervened to crush the Belgian revolution in 1790, however, France gave no support to the Belgians.

The revolutionaries in France believed their political principles to be universal. Most of them were focused on French domestic politics, though, and did not want to interfere in other countries' domestic politics. Not intending to go to war against anyone, they disavowed the right or the intention to promote revolutions abroad or make military threats. In May 1790 the Constituent assembly passed a constitutional article renouncing wars of conquest and the use of military force to interfere in other countries' domestic politics. By this point, however, Austria knew that revolutionary France had already engaged in this kind of interference by connecting with the Belgian democrats. This information alerted the Austrian leadership that France's posture toward foreign revolutionaries was multi-faceted at best, duplicitous at worst.

## 5. First effects of the revolution on Austria

### 5.1 Austrian domestic politics

The Austrian Emperor, Joseph II, and his first minister, Prince Kaunitz, had a strong preference for autocratic rule. When the French Revolution broke out in 1789 and the French constituent assembly provoked mob violence to extract concessions from the French king,<sup>29</sup> Joseph concluded that the assembly was an unruly threat to autocratic rule and it might tempt the lower classes to overthrow monarchical government altogether – not only in France but throughout Europe. Joseph expected the lower classes in Belgium to be among the first ones to act on that temptation.<sup>30</sup>

Joseph's expectations were confirmed quickly. The outbreak of the French Revolution triggered waves of unrest in many provinces of the Austrian Habsburg monarchy. The Austrian peasantry were inspired to new levels of opposition to the feudal system. After hearing reports of the French peasantry's actions, the Austrian peasants followed suit – increasing their demands and engaging in violent resistance to their feudal obligations and the nobility's authority to enforce them. Within a few months the regime had to deploy regular army troops to restore and maintain order in the countryside. The nobility saw these developments as threats to the feudal system that could lead to a revolution if not opposed.<sup>31</sup> The news from France also magnified other domestic threats to the Austrian regime. First, nobles throughout Austria interpreted the French revolution as evidence that Joseph's reforms would cause a peasant revolution in Austria if continued. The nobility intensified its pressure on the regime to abandon the reforms and restore its traditional rights and privileges – pressures that carried the implicit threat that the nobility would withdraw its cooperation with the regime if not satisfied.<sup>32</sup> Even in central Austria and Bohemia where the provincial nobilities were weakest, the estates met to protest the tax and peasant reforms scheduled to

---

<sup>29</sup>The violence was provoked unintentionally by the French constituent assembly.

<sup>30</sup>Hochedlinger 2003, p. 303.

<sup>31</sup>Wangermann 1969: 32-33, 73.

<sup>32</sup>Wangermann 1969: 51, 70-71.

go into effect in November.<sup>33</sup> Second, the middle-class campaign for constitutionalism and the rule of law took inspiration from the French constituent assembly's actions and hoped for similar measures to combat absolutism and noble privilege in Austria. Professionals and government administrators opposed the nobility's claim to a privileged place in society and supported the cause of the Austrian peasants.<sup>34</sup> Third, the nationalist movements in Hungary and Belgium, both provinces of the Austrian monarchy, took inspiration from the new French ideology that held that "the nation" was sovereign and the monarch was obligated to heed "the will" of the nation. Following the outbreak of the French Revolution in 1789, Belgium experienced a revolution of its own that led to a declaration of secession from Austria. The provinces of Hungary and Galicia threatened to do the same.

Each class of Austrian society wanted autonomy from Joseph's regime for a different reason. While the upper and lower classes intended to use their autonomy to oppose each other, the middle class intended to use its autonomy to promote the rule of law and support the lower class. Yet all classes were united in the short run by the goal of overthrowing Josephist absolutism.<sup>35</sup> In short, news of the French revolution impacted all levels of Austrian government and society and intensified the frictions between them dramatically. Because the absolutist regime held all the power to give each group what it wanted, all of the conflicting domestic pressures converged on Joseph's regime.

The regime was not uniform in its response to these pressures. Many of the rank-and-file administrators thought the main problem was still the oppressed condition of the peasantry and advocated more liberal reforms.<sup>36</sup> The senior ministers recognized this problem and expected the peasant grievances to intensify as news from France continued to spread. Yet the senior ministers also thought the nobility's claims had merit. On a practical level they knew the nobility could withhold the most basic resources the regime needed – tax revenue and local political authority – to win the war against the Turks, deter a war with Prussia over Poland, and maintain domestic order in the face of peasant disturbances and middle-class

---

<sup>33</sup>Beales 2009, p. 624.

<sup>34</sup>Wangermann 1969: 33-34, 70-73

<sup>35</sup>Wangermann 1969: 33-34.

<sup>36</sup>Wangermann 1969: 58-59[?mss].

demands for the rule of law. The ministers put enormous pressure on Joseph to join and lead the conservative reaction.<sup>3738</sup> As for the ruler himself, Joseph deeply wanted to continue the liberal reforms regardless of ministerial pressure.<sup>39</sup> The regime had a cohesive bureaucracy and a strong, well-financed military<sup>40</sup> that probably had enough power to suppress a mass *political* revolution of the lower and middle classes like the French Revolution.<sup>41</sup> At the same time, though, the regime lacked the military power to maintain *social and economic* order among peasants and lords in all provinces all the time.<sup>42</sup> Disorder in the agricultural economy could easily lower the monarchy's tax revenue and threaten its hold on authority just as seriously as a political revolution would. Joseph's decisions were determined not so much by material constraints, however, but by his commitments to maintain absolute power and pursue enlightened reform.<sup>43</sup> Accordingly he insisted that the tax and peasant reforms go into effect as scheduled in November 1789.<sup>44</sup>

## 5.2 Austria's relations with France

The French revolution had two international spillover effects that raised Austrian concerns about the threat of revolutionary contagion.

### 5.2.1 Avignon and Alsace

The enclave of Avignon in southern France was an independent state under papal sovereignty. It had had that status for centuries. When the French revolution broke out, it sparked a local

---

<sup>37</sup>Wangermann 1969: 35, 49-55, 58-59.

<sup>38</sup>Wangermann 1969: 35, 51-52.

<sup>39</sup>Wangermann 1969: 35, 51-52.

<sup>40</sup>Hochedlinger 2003.

<sup>41</sup>Beales 2009, p. 626.

<sup>42</sup>Wangermann 1969.

<sup>43</sup>Beales 2009, p. 593.

<sup>44</sup>"(Joseph) set forth on a collision course, reviving and exacerbating almost every possible grievance in every province. One can only wonder at the dedication and willpower of the dying Emperor in his desperate campaign to bring his policies to fruition. It was magnificent, but it was not politics." Beales 2009, pp. 593, 624.



civil war between the residents of Avignon who wanted to remain under papal sovereignty and the residents who wanted to be recognized as French citizens – with all the rights now attached to citizenship by the French revolutionaries. The French constituent assembly in Paris debated the option of annexing Avignon.<sup>45</sup>

The province of Alsace had been ceded to France in the Treaty of Westphalia (140 years earlier). The treaty stipulated that the French sovereign was prohibited from abridging the local rights and privileges of the Alsatian estate holders: landed nobles, churches and cities. The French and Germans had never agreed on how to interpret these clauses of the treaty on a practical level, however. Over the decades French kings took advantage of these contractual ambiguities to whittle away the rights and privileges of the Alsatian estate holders to the point that Alsace eventually became like any other French province. In 1789 the revolutionary constituent assembly followed these precedents to their logical conclusions – entirely eliminating the feudal rights and privileges of the Alsatian estate holders. A few months later the constituent assembly overran the jurisdictions of the Alsatian estate holders further by including Alsace in a new law which redrew the boundary lines between France's ancient counties and converted them into *departments*. Many Alsatian estates were held by noble tenants of German princes who were members of the Holy Roman Empire and, as such, had rights to legal redress from imperial institutions and military protection from the Holy Roman Emperor, Joseph II of Austria.<sup>46</sup>

Thus the revolutionary enactments in Alsace raised difficult legal questions about the Treaty of Westphalia that created a conflict of interest between France and Austria. How should the treaty be interpreted in light of the regime transition in France? Was revolutionary France exempted from the clauses of the treaty that conflicted with its new political principles – clauses it never would have accepted in the first place if it had been governed by these principles in 1648? If France was exempt from these clauses, then what would be the status of the treaty rights of the German princes in Alsace? In short the French revolution created a problem of *contradictory legal rights*. On one hand, the Treaty of Westphalia ceded the province of Alsace to France, and this gave the French government the right to protect the

---

<sup>45</sup>Clapham 1899: 18-19.

<sup>46</sup>Clapham 1899: 19-21.

Alsatian peasants from their lords. On the other hand, the treaty gave the Alsatian lords (and their German overlords) the right to immunity from acts of the French government that infringed their local rights and privileges. This contradiction of legal rights created a conflict of interest between France and Austria, because Joseph was obligated to protect the German overlords from any infringements of their rights and privileges in Alsace.

The French enactments in Alsace also raised difficult political questions about the right of the French revolutionaries to support their sympathizers beyond France's borders. How far did the revolutionaries have the right to go in supporting foreign sympathizers? What exactly did the revolutionaries have the right to do in support of foreign sympathizers? What exactly were the revolutionaries prohibited by international law from doing in support of foreign sympathizers? These questions about the revolutionaries' right to support foreign sympathizers were turned into conflicts of interest by the problem of contradictory legal rights. On one hand, as far as the revolutionaries were concerned, they had been awarded the right to govern Alsace by the Treaty of Westphalia, and *therefore* they had the right to support the victims of feudal oppression in Alsace just as anywhere else in France. On the other hand, as far as Austria was concerned French support for the Alsatian peasants was not simply a matter of the French revolutionaries supporting foreign sympathizers under German suzerainty, but also a matter of French violation of the rights of the German princes under the Treaty of Westphalia. Thus, by extending protection to the Alsatian peasants, the revolutionaries were claiming in effect that France's new political principles nullified the relevant clauses of the Treaty of Westphalia.

The French actions in Alsace thus raised a key question for the Austrian leadership. Were the French revolutionaries violating the relevant clauses of the Treaty of Westphalia intentionally and knowingly in order to export their revolution? Or were they violating these clauses unintentionally, simply because they had not bothered – in their zeal for emancipation and equality across all of the departments of the French nation – to consider whether their actions in Alsace contradicted France's existing treaty commitments? The answer was important because unintentional violations were not necessarily a sign of opportunism by the French revolutionaries, while intentional violations were a sign of French opportunism in supporting sympathizers abroad.

Austria's difficulty in discerning the difference was complicated by the French constituent assembly's proclamation that France did not intend or claim the right to export its revolution beyond its borders. The credibility of this reassurance would be thrown deeply into question if the revolutionaries had violated the Alsatian and German rights intentionally. Hence the Emperor had to screen the revolutionaries for their intentionality by bringing the legal issue to their attention and observing their response. The Emperor could afford to proceed slowly and judiciously in this screening process as long as the French revolution remained moderate in its domestic goals, because the risk of international revolutionary contagion aided by French opportunism was low. If the revolution turned radical, however, the Emperor would have to complete the screening process quickly.

The same problem arose in reverse when the conservative, monarchical governments of Germany, Prussia and Austria sought to support the French nobles who were working to mount a counter-revolution within France. This problem too was exacerbated by the existence of contradictory legal rights.

### 5.2.2 The French noble émigrés

After the revolution broke out, many French nobles emigrated to Germany, Prussia and Austria. While the largest numbers left in 1789, the emigrations continued periodically for years. The French noble émigrés tried to convince their hosts – the German princes, the Hohenzollern court in Berlin and the Habsburg court in Vienna – that the revolution in France had radical ambitions at home and abroad and they should intervene militarily to overturn it. This development posed four interrelated problems for Franco-Austrian relations.

First, the émigrés had an incentive to overestimate how radical the revolutionaries in Paris were, how easy it would be to overturn the revolution by force, and how much military force they could muster to support such a conservative intervention. Hence the Austrian leadership had to discount the émigrés' reports and advice, lest they provoke the Austrian regime into escalating against France without a true *raison d'état*.

Second, the French king and queen used the émigrés as back channel diplomats to communicate with the other conservative regimes of Europe. The messages they sent through this channel were multiple and mixed, however, because their own position was deeply conflicted.

After the revolutionaries forced the king and queen to move to Paris under a kind of house arrest, the king felt that any further constitutional concessions he made could be disavowed later on the grounds that they had been extracted by force. After he communicated this situation to the Bourbon court in Madrid, news of his conflicted position spread to the other rulers of Europe – and brought them to realize that the French revolution might become so radical they would have an interest in overturning it by force. This prospect gave them a keener interest in the news reports coming from the French noble émigrés. The Austrian ruler had the keenest interest because the French queen was his sister. Hence he had to listen to the noble émigrés' reports intently with one ear while discounting the information heavily with the other ear (for the reason stated above).

Third, because the French monarchs were in captivity to the French revolutionaries, they might be the first to die in a war of intervention initiated by the conservative powers. Hence the French monarchs implored the Austrian rulers *not* to yield to the émigrés' pressures to mount an invasion. This restriction placed the Austrian leadership in an even more difficult position with respect to the émigrés. For on one hand, if the revolution in France were to turn radical, the Austrian leadership would be able to assign greater credibility to the émigrés' reports about France. Yet on the other hand, the Austrian leadership would have to summon greater resistance to the émigrés' pressures for military action – simply to avoid getting the French king and queen killed in the action.

As if this situation were not complex enough, a fourth problem arose because the foreign activities of the noble émigrés had a significant impact on French domestic politics. The domestic support base of the French monarchy, the "royalist" party, was confident that if the revolution became too radical, the noble émigrés would convince the conservative powers to intervene militarily to overturn the revolution. The royalists' confidence on this point was clear to the French lower and middle classes, however, and simply intensified their drive to secure the gains of their revolution beyond attack and be rid of the royalists for good. Thus, the more intently the Austrian leadership might listen to the French noble émigrés for one reason or another, the more intently the French lower and middle classes would strive to end the feudal and monarchical system of government in France – and appear ever more radical to the Austrians.

In sum, the emigration of the French nobles from the country complicated relations between France and Austria tremendously. It was not just a matter of the intersection of domestic and international politics, however. It was a matter of contradictory legal rights. On one hand, the conservative, monarchical governments of Austria, Prussia and Germany had the right to host any foreigners of any ideological persuasion they chose. On the other hand, France had sovereign rights to its territory and political independence. These two legal rights would always be in contradiction as long as the French noble émigrés continued to pressure the conservative powers to intervene militarily to overturn the French Revolution. For the conservative powers could always claim their right to host foreigners of any ideological persuasion. But revolutionary France could always claim the right to protect its territorial sovereignty and political independence from the émigrés and the foreign governments conspiring to invade. As we will see, each side in this legal conflict of interest was torn between its resolve to defend its legal right and its recognition that the other side had something of a point too. Each side's tendency to be torn in these two directions provides *prima facie* evidence that the problem of contradictory legal rights drove each side's decisions and actions.

## 6. Austria's first response to the French Revolution

### 6.1 Summary

### 6.2 Domestic response: Conservative reaction

When the French Revolution broke out in July 1789, Joseph was already facing conservative noble reactions against his tax, peasant and absolutist reforms in all provinces.<sup>47</sup> Despite the revolutionary ferment coming from France he wanted to defeat these domestic movements and continue on the path of reform, either by making only minor concessions or by repressing the nobles forcefully. His top ministers made clear that neither policy would succeed and he had to offer major concessions to the nobility to save the regime from a *noble* revolution. So despite his desire for reform and perseverance against earlier noble opposition, the risk

---

<sup>47</sup>Beales 2009, pp. 623–4.

of a domestic revolution in Austria forced him to join forces with the nobles in a conservative reaction. He reluctantly agreed to their demands for restoration of their rights and privileges.<sup>48</sup>

After he passed away in February, 1790, his successor had to continue the policy of conservative reaction at least initially, because he lacked the political capital to change course against noble opposition. His top ministers recognized the threat posed by peasant discontent but thought the risk of a noble revolt against the regime to be even more threatening. In response to their advice Leopold abandoned the tax reform of 1789 (which had raised taxes on the nobles and reduced the tax rate they could charge peasants). This allowed the nobles to reimpose feudal labor obligations on the peasantry. He authorized measures to combat the threat of revolutionary contagion from France: strict censorship of news coming from France and surveillance of all foreigners and subjects suspected of revolutionary agitation in Austria.<sup>49</sup>

Leopold's concessions to the nobility only aggravated the domestic ferment over fundamental issues. From the moment he scrapped the tax reform, peasants throughout the monarchy resisted the reimposition of feudal labor duties so forcefully that regular army troops were needed to enforce the law and keep order. It took months to suppress the peasant rebellions. And even then the peasants resisted the old labor regime continually through legal and administrative appeals.<sup>50</sup> The middle-class advocates of constitutionalism and the rule of law supported the peasants even more strongly now, drawing up schemes for peasant emancipation and reform of the provincial noble assemblies to allow representation of all classes.<sup>51</sup>

---

<sup>48</sup>Beales 2009, pp. 627–30; Wangermann 1969, pp. 51–55.

<sup>49</sup>Wangermann 1969: 56–65.

<sup>50</sup>Wangermann 1969: 68–70.

<sup>51</sup>Wangermann 1969: 70–73, 75–76.

## 6.3 Foreign policy responses

### 6.3.1 Moderation toward France

From the start of the French Revolution in July 1789, the Austrian regime and nobility were concerned about the spread of revolutionary ideology into Austria.<sup>5253</sup> They believed the domestic goals of the French revolutionaries' were limited and the French monarchy would survive. So they saw the *immediate* threat of revolutionary contagion into Austria as being low. Yet they knew the Austrian lower and middle classes were taking lessons from France about how to oppose absolutism and feudalism. So the Austrian regime and nobility felt the need to watch developments in France closely for signs the revolution was radicalizing. Any radicalization in France would be concerning because the Austrian provinces of Belgium, Hungary and Galicia were already suffering discontent and unrest of the kind that had just caused the French revolution. The ruling circle in Vienna anticipated it might be necessary to pacify these provinces by military force.<sup>54</sup>

At the outset in 1789, the Austrian government considered the option of going to war to overturn the French revolution and restore the old regime in France. The Austrian ministers were divided on this question. The opponents of military action, led by first minister Kaunitz, held that military action would be either unnecessary or too costly. It would be unnecessary if the revolution collapsed. It would be too costly if the revolution survived. In any case there were cheaper means by which to insulate Austria from the ideology of the French Revolution. Under his leadership the Austrian ministers reached a consensus that as long as the revolution remained moderate the threat of revolutionary contagion was low, so it was not be necessary to take military action against the revolution (implication 2.1). They agreed to follow a strict policy of nonintervention.<sup>55</sup> Yet they also agreed to continue watching developments in France closely for signs of radicalization.

Thus the Austrian regime was torn between its *current* assessment that the threat of

---

<sup>52</sup>Hochedlinger 2003, p. 303.

<sup>53</sup>Wangermann 1969.

<sup>54</sup>Roider 1987, p. 76.

<sup>55</sup>Hochedlinger 2003, p. 304.

revolutionary contagion was low and its concern that the threat might escalate in the *future*. The regime struck a balance by adopting a policy of watchful non-interference. It resolved not to do anything that might provoke French fears of an invasion or trigger a radicalization of the revolution. It refused to support the French noble émigrés in Germany and Austria who were lobbying conservative rulers to make war on France to crush the revolution. On the other hand, however, the regime resolved to reevaluate the need for war in the event that the revolution in France turned radical. For the next two years the regime assessed the threat of revolutionary contagion from France cautiously but evenhandedly (implication 1).

Some elements in the Austrian regime were not content to watch the French Revolution idly, however. They preferred to see the French revolution produce a moderate constitutional monarchy – rather than either a republic or a restored absolutism – for both ideological and strategic reasons. Ideologically, French ideas about rational and constitutional government had some appeal to the Austrian rulers. While Joseph had spent years trying to rationalize his own government, his successor, Leopold, thought monarchs should be accountable to their subjects and looked for opportunities to resume the path of reform. Strategically, some constitutional limitations on the French monarchy would reduce France's power in the international arena and favor Austria. To enable the French Revolution to produce a moderate result, therefore, elements of the Austrian regime developed contacts inside the French king's regime to provide advice and support in dealings with the revolutionaries. This strategy of back channel engagement also held the prospect of averting a radicalization of the French Revolution that would increase the risk of revolutionary contagion and war.

Thus Austria's strategy of watchful non-interference had limits. It might offer France the reassurance that Austria was committed not to interfere in French domestic politics – and thereby avoid a radicalization of the revolution. But some behind the scenes engagement in French politics might also help to avoid a radicalization – even at the risk of provoking French fears of foreign interference!

All of this evidence supports implications 1 – 3 of the theory presented earlier. One might object that there is an alternative explanation for why Austria did not initiate war against France for the first three years after the revolution erupted. Namely, Austria was still involved in military action against the Turks for some time and faced the possibility of



a war over Poland too. This hypothesis is refuted by the evidence. Austria maintained its wait-and-see posture toward France long after these other conflicts ended. Austria continued to maintain that posture as long as the French Revolution remained moderate. Yet when the revolution turned radical Austria changed its posture and this triggered an action-reaction cycle that resulted in war. These patterns of evidence are consistent with the Contractual Realist theory but inconsistent with the alternative explanation just mentioned.

## **7. Leadership transition and strategy change in Austria (March 1790 to May 1791)**

Joseph II died in February of 1790. His brother and successor, Leopold II, changed the direction of Austrian strategy significantly.

### **7.1 Foreign policy**

Leopold believed that if the Austrian monarchy was to defuse the domestic and international threats it faced, peace was essential. So he adopted a grand strategy of "peace on all fronts". He also believed the stakes were higher on the domestic front than the international front. Hence his first priority was to make peace in the East with the Turks and in the north with Prussia. Leopold believed the Turks were continuing to fight only because Prussia encouraged them – to divert Austrian power to the East and increase Prussian leverage on the Polish question. Within a month of taking the throne (March 1790), Leopold offered Prussia a settlement of the outstanding disagreements on Poland. In mid-August 1790 military hostilities with the Turks ended.<sup>56</sup> Turning to the domestic threats, restoring Austrian rule in Belgium was the highest priority.<sup>57</sup> Not only was the loss of Belgian tax revenue significant, but as long as Belgian independence subsisted, it posed a threat of *secessionist* contagion to the other "national" communities of the Austrian monarchy in Italy, Hungary and Galicia. At the same time Leopold paid close attention to developments in France because the French queen was his sister and, more fundamentally, a radical turn of the revolution against the

---

<sup>56</sup>The formal truce did not take effect until September. Roider 1987, p. 79.

<sup>57</sup>Roider 1987, pp. 81–2.

French monarchy posed a threat of *ideological* contagion to the Austrian lands. The Austrian policy of noninterference in the French Revolution, inherited from Leopold's predecessor, continued to reign in Vienna.<sup>58</sup>

## 7.2 Domestic policy

Leopold continued the policy of conservative reaction at first, because he lacked the clout to change it. He repealed the previous reforms of peasant labor obligations. He responded favorably to the nobles' campaign to restore their rights and privileges and maintain their exclusive control of the provincial assemblies. Outside the nobility and its supporters in the regime, however, all segments of society throughout the Austrian monarchy opposed the conservative reaction. The peasantry resisted the reimposition of labor obligations and pressed for the right to be represented in the provincial assemblies. They were supported by the "Josephinian" civil servants in the regime's bureaucracy who were committed to liberal reforms.

Within a few months of his succession, Leopold gained enough confidence and authority to reverse course. Abandoning the conservative reaction of the nobility, he returned to the path of liberal reform. While his primary goal was to raise tax revenue by stimulating the rural economy, he also saw a political opportunity. By giving voice to lower and middle class interests, he thought he could force the nobility to moderate its demands and accept liberal reforms. He expected to be able to "ride the tiger" of the masses' demands to a stable class compromise.

Leopold spearheaded reforms in four areas. He ordered the police ministry to be redirected from the functions of surveillance and repression to those of public protection. He initiated efforts to reform the provincial assemblies to allow representation of the lower and middle classes. He authorized a comprehensive legal reform to reduce the privileges of the noble class and thereby insure the loyalty of the middle and lower classes. In this area he was caught between opposing ministerial factions, but managed to find a middle ground. On one hand he favored the right of the provincial assemblies to be consulted on major changes in the law. On the other hand his legal reform efforts stopped short of full emancipation for

---

<sup>58</sup>Roider 1987, p. 81.

the lower classes, full democracy and full subjection of the monarchy to law.

The most important area of liberal reform was the reduction of peasant labor obligations. The monarchy had gone back and forth on this issue over the years (see above). Reversing its position again, Leopold took a favorable view of peasant demands for the reduction or elimination of labor obligations to the lords. He ordered the provincial noble assemblies to submit proposals for the commutation of peasant labor obligations. Although resolution of the issue was postponed repeatedly by the nobles' foot dragging, Leopold persisted in demanding concessions from the nobility over the next two years until his death in 1792.

It is notable that Leopold persisted on the labor issue despite escalating tensions with France in the same period. After June of 1791 tensions with France escalated into a dispute over the right of either country to interfere in the domestic politics of the other country (see below). Yet Leopold continued to press the Austrian nobility to submit proposals for the commutation of peasant labor obligations. This pattern of evidence contradicts and refutes two other theories of why the war occurred: the "diversionary" theory and the "elite logrolling" theory. Consider each in turn.

The diversionary theory assumes that Austria's diplomatic escalations against France were driven by the goal of *diverting* the Austrian peasants' attention from their demands for commutation of labor obligations – so Leopold would not have to meet those demands. Yet the evidence shows that Leopold *wanted* to meet their demands. Hence Austria's diplomatic escalations against France could not have been driven by that diversionary motive. The elite logrolling theory assumes that the regime and the nobility *colluded* to start a war for their joint private gain at the *expense* of the lower and middle classes. Yet Austria's escalations against France occurred at a time when Leopold's domestic policy was to search for a *compromise* between the upper and lower classes, even though it might *antagonize* the nobility in the process. Hence Austria's escalations could not have been driven by the domestic logrolling mechanism. While some kind of domestic logrolling may have occurred after Leopold's death (see below), the mutual escalation to the point of deciding for war had already occurred in Leopold's reign. And it occurred in a manner consistent with the observable implications of the Contractual Realist theory.

## 8. The Belgian Revolution and its impact on Franco-Austrian relations

As part of the Austrian Monarchy, Belgium was ruled by Governors who were appointed and took orders from Vienna. Before 1786 Belgium had escaped the kind of centralizing reforms that Maria Teresa and Joseph implemented in the other provinces. So some of the Belgian provinces still had constitutions that placed tight limits on Joseph and the Austrian governors.<sup>59</sup> When Joseph started to implement reforms that reduced the traditional rights of the upper and middle classes, it did not take long for them to rebel. By the spring of 1787 the three Estates of the province of Brabant - clergy, nobility and merchants – were refusing to grant taxes to the regime.<sup>60</sup> All classes of Belgian society joined the resistance to Joseph's new laws and signaled a willingness to resort to arms to oppose them – leaving his governors with a stark choice between deploying military force to impose the laws or rescinding them. The governors rescinded the laws and Joseph's chief Minister, Prince Kaunitz, supported their decision because he knew that political realities in Belgium required it.<sup>61</sup>

The rebellion pitted the principle of absolutist government against the principle of representative government. Joseph maintained the right to impose his reforms over the representative Estates' opposition. They only represented the upper and middle classes (the clergy, nobility and commercial bourgeoisie). Yet the rest of Belgian society came to their aid because of a common interest in preventing the emergence of military autocracy in the country. At this point the lower classes simply supported the middle and upper classes' interest in restoring the traditional rights of the Estates to be consulted on new laws and taxes and to veto proposals. In this sense the rebellion was nothing like the French revolution that would erupt two years later.<sup>62</sup> Yet the development of broad popular support for the Belgian rebellion in 1787 indicated that a cross-class alliance was possible in the country.

Negotiations between Joseph and the Estates continued through the summer of 1787.

---

<sup>59</sup>Beales 2009, p. 501.

<sup>60</sup>Beales 2009, pp. 504–7.

<sup>61</sup>Beales 2009, pp. 512–3.

<sup>62</sup>Beales 2009, pp. 515–6.

Joseph stood firm in his insistence on pushing through the reforms that would reduce the Estates' traditional rights and privileges. Their representatives insisted that Joseph rescind all of those reforms. The two sides were still in deadlock when war erupted between the Turks and Austria's ally, Russia, in September 1787. Joseph's plan to send 300,000 troops to crush the Belgian rebellion had to be scrapped. Worse still, a separate rebellion in Holland provoked a Prussian invasion there which sent streams of dissidents with democratic and republican notions about government pouring into Belgium.<sup>63</sup> Now the possibility of a cross-class alliance in Belgium portended more than simply pressure to restore the traditional constitutions protecting upper and middle class rights. While the Turkish war continued, Joseph could only spare 18,000 troops from the Turkish front to send to Belgium. This was not enough to crush the Belgian resistance by overwhelming force.<sup>64</sup>

The Belgian resistance movement had two wings. The conservative nobility was pressing Joseph for a restoration of its traditional rights and privileges and, failing that, intended to declare independence from Austria and find another great power that would provide international and domestic security. The lower and middle class democracy activists, on the other hand, wanted to inspire a popular revolution in Belgium that would obviate the need for foreign support. The Austrian governments in Vienna and Brussels were well-informed about the conservative wing but scarcely informed about the radical wing,<sup>65</sup> a difference which may explain why the Austrian leaders underestimated the potential for a popularization of the resistance movement.

The outbreak of the French Revolution in July 1789 provided great inspiration to the Belgians who were resisting Joseph's absolutism.<sup>66</sup> The Belgian lower and middle classes joined the resistance in large numbers and broadened its goals. The middle class wanted representation in the provincial assemblies, movement toward the rule of law and constitutional restraints on the regime. The peasant class wanted emancipation from its feudal obligations as well as representation in the provincial assemblies. Over the summer of 1789

---

<sup>63</sup>Beales 2009, pp. 517–25.

<sup>64</sup>Beales 2009, p. 610.

<sup>65</sup>Beales 2009, p. 613.

<sup>66</sup>Beales 2009, p. 611.

small disturbances broke out and were crushed by Austrian troops. In August a popular revolution broke out in the small independent state of Liege (next door to Belgium) and took control of its government.<sup>67</sup> Belgian dissidents gathered there. The Belgian governor feared that the revolutionary movements of Paris and Liege would spread to Belgium.<sup>68</sup> The Belgian resistance movement began to muster troops across the borders in Holland and Liege in preparation to support a popular revolution.<sup>69</sup> The Belgian governors were divided over whether the rebel troops constituted a real threat.<sup>70</sup> In October the conservative and radical wings of the resistance movement reached an agreement on terms for their cooperation and the rebel army crossed into Belgium. Receiving much popular support, the rebels evicted all of the Austrian troops and government administrators from Belgium within two months.<sup>71</sup>

When the Belgian nobility initiated negotiations with the Liege revolutionaries, Prussia feared that a cross-class alliance would emerge in Belgium. To avert it, Prussia intervened to crush the popular revolution in Liege and position itself to crush any popular revolution in Belgium and turn it into a client state.<sup>72</sup> One of the leading French revolutionaries responded by sending an envoy to Belgium to make contacts with the Belgian democracy activists. The envoy promised them that if the Triple Alliance – Britain, Holland and Prussia – intervened to crush the Belgian popular movement, France would send military support.<sup>73</sup>

Joseph appointed a new governor for Belgium and empowered him to make whatever concessions were needed to restore Austrian rule. But the Belgian rebels wanted concessions that Joseph could not grant: the convocation of an Estates General, and a guarantee of their provincial constitutions by Britain, Holland, France and Prussia.<sup>74</sup> After the negotiations broke down in December, the conservative wing of the Belgian revolution created governments in all the Belgian provinces which were entirely aristocratic and denied the radical

---

<sup>67</sup>Illing 2009, p. 71.

<sup>68</sup>Beales 2009, pp. 611–2.

<sup>69</sup>Illing 2009, p. 73.

<sup>70</sup>Beales 2009, p. 614.

<sup>71</sup>Beales 2009, pp. 613–7.

<sup>72</sup>Illing 2009, p. 71.

<sup>73</sup>Clapham 1899, p. 41.

<sup>74</sup>Illing 2009, p. 73.

democratic revolutionaries any rights or role.<sup>75</sup> The next month the conservative Belgian nobles deposed Joseph, declared Belgium to be independent of Austria, and signed a domestic treaty to form the United Belgian States – a republic with no head of state and little central government.<sup>76</sup> The loss of Belgium reduced the Austrian monarchy's revenues by 10%.<sup>77</sup>

The Belgian nobility was divided in its attitude toward the lower and middle class democracy activists. The more liberal nobles were sympathetic to the Belgian democracy movement and wanted the country to enact reforms of the kind passed in Holland, Switzerland and the United States – all republics. But they formed a minority of the Belgian nobility.<sup>78</sup> The conservative majority that dominated the new Belgian government was unsympathetic to the domestic radicals and expelled many of them from the country. The conservative nobility quickly restored its rights and privileges at the expense of the lower and middle classes, thereby appropriating the Belgian independence movement to achieve its own class goals. Naturally, the lower and middle classes were aggrieved by the nobility's failure to reward their support by awarding them representation in the provincial assemblies. As a result popular support for the democracy movement continued to increase despite the expulsion of many of the democracy activists.<sup>79</sup> The conservative nobility's domestic victory thus did not eliminate the threat of a popular revolution in Belgium. It exacerbated the threat.

Worse, many of the expelled Belgian radicals emigrated into northern France. The French support for these émigrés was perceived in Austria as a threat to Austria's plan to restore its rule in Belgium.<sup>80</sup> The threat perception was fueled by the revolutionary rhetoric in Paris which claimed that France was actually obligated by international treaty to protect the lower classes beyond its borders from the German princes who were bent on crushing popular movements.<sup>81</sup> Thus Austria perceived its plan to restore its rule in Belgium to be under threat both from within Belgium – by the growing popularity of the domestic democracy

---

<sup>75</sup>Beales 2009, p. 621.

<sup>76</sup>Illing 2009, p. 72.

<sup>77</sup>Beales 2009, pp. 621–2.

<sup>78</sup>Illing 2009, p. 73.

<sup>79</sup>Wangermann 1969, pp. 71, 76.

<sup>80</sup>Hochedlinger 2003, p. 306.

<sup>81</sup>Savage 1998, p. 250.

movement – and from outside – by the radical émigrés in northern France and their French revolutionary supporters.

The Belgian nobility was divided not only in its attitude toward the Belgian democracy activists, but also in its views of whether Belgium should remain independent or fall under the suzerainty of another great power (beside Austria).<sup>82</sup> The liberal nobles wanted the country to remain independent and focus on broad-based domestic reforms.<sup>83</sup> The conservative nobles who controlled the new Belgian government wanted the country to fall under the suzerainty of another great power – to provide an internal security guarantee of the newly restored aristocratic constitution and international security guarantees against both revolutionary France and absolutist Austria. The conservative nobility thus faced a complex dilemma spanning the international and domestic spheres in its search for security. On one hand, it could turn to Prussia as the *foreign* power most interested to prevent a restoration of Austrian rule in Belgium. But that move might simply invite a new Prussian absolutism to take the place of the old Austrian one – as Prussia's heavy hand in Liege already demonstrated. On the other hand, the nobility could turn to the lower and middle classes as the *domestic* power most interested to prevent a restoration of Austrian rule in Belgium. But that move might simply invite a new democratic revolution to curtail noble rights in place of the old Austrian absolutism that curtailed noble rights.

The Belgian nobility's options were tightly constrained, moreover, by the preferences of the neighboring great powers: Britain, Holland, Prussia and France. These powers faced two questions with high stakes as a result of the Belgian declaration of independence. First, should Belgium remain an independent state or should it fall back under the suzerainty of a great power? Second, should the tension between Belgium's conservative nobility and liberal democracy activists be resolved in favor of the former or the latter? The great powers had connected and conflicting interests on these questions. Britain and Holland preferred to see Belgium fall under the suzerainty of any great power that would keep it out of French hands, first and foremost, and Prussian hands too if possible. The natural candidate for

---

<sup>82</sup>Illing 2009, p. 74.

<sup>83</sup>Illing 2009, p. 73.



the job was Austria in their view.<sup>84</sup> Prussian opinions on Belgium were divided. Some in Prussia preferred to see Belgium fall into the hands of a great power that would keep it out of Austrian hands, because this would reduce Austria's revenues and power to oppose Prussia's ambitions in Poland. One Prussian general even suggested to Lafayette, a leading French revolutionary, that Prussia was indifferent to the form of the Belgian constitution, and would welcome a democratic revolution in Belgium with French support – if only to keep the country out of Austrian hands. But others in Prussia preferred Belgium to fall back into Austrian hands *if* this would secure Austrian concessions on the Polish front!<sup>85</sup> All three powers of the Triple Alliance – Britain, Holland and Prussia – were agreed on one point though. Their alliance would only support the new Belgian government on the condition that it limit its concessions to the Belgian lower and middle classes and retain an aristocratic form of constitution. This condition placed a tight constraint on the conservative Belgian nobility running the new Belgian government.

In France, some of the revolutionaries claimed that France's greatest interest was to prevent another great power from dominating in Belgium, especially Austria or Prussia.<sup>86</sup> If Austria dominated in Belgium, it would stage its troops on France's northern border and threaten the French revolution with subversion and invasion. Such revolutionaries argued that France should enter an alliance with an independent Belgium to prevent Austria from reoccupying it.<sup>87</sup> Other French revolutionaries were reluctant to support Belgium's new government, however, as long as the tensions between its conservative nobility and liberal democracy activists remained unresolved. They argued that France should refuse to support Belgium until the ruling nobility abandoned the aristocratic constitution and recognized the popular sovereignty of the non-privileged classes; and in the meanwhile France should station its troops on the Belgian border to support the Belgian democracy activists and prevent an alliance between Belgium's conservative government and Prussia.<sup>88</sup> The French reluctance to support an aristocratic Belgian government also placed a tight constraint on the ruling

---

<sup>84</sup>Illing 2009: 71, 74. Savage 1998: 249.

<sup>85</sup>Illing 2009, pp. 72–3, 75.

<sup>86</sup>Illing 2009, p. 75.

<sup>87</sup>Savage 1998, p. 247.

<sup>88</sup>Illing 2009, pp. 74–6.

Belgian nobility.

Thus Belgium's conservative nobility faced a tough choice. It could receive French support against Austria by making concessions to the lower and middle classes, but at the risk of alienating the powers of the triple alliance or even provoking them to intervene to prevent French domination of Belgium. Or it could hold the Belgian lower and middle classes at arms length to retain the Triple Alliance's support in keeping Austria at bay, but at the risk of losing French support and provoking a domestic revolution! In the end the Belgian nobility chose the latter option. Yet the members of the Triple Alliance were bound not to intervene in Belgium at Austria's expense without prior consultation. And Britain refused to support Prussia's intervention plan. So independent Belgium was left without any great power help at all.

When Joseph, the Austrian ruler, died in February 1790, his successor Leopold seized the opportunity of Belgium's helplessness to drive a wedge between the Belgian upper and lower classes. Leopold's attitude toward rebellious nobles differed from that of his predecessor. While Joseph sought autocratic rule, Leopold detested it and favored constitutional constraints on the monarchy – as long as it was clear to all classes that the purpose of the constitution and the representative assemblies was to fortify the power of the monarchy, not to achieve popular sovereignty or perfect the rule of law. Leopold saw the helplessness of Belgium's conservative nobility as an opportunity to restore Austrian rule. He would offer the nobility a restoration of its traditional rights and privileges in return for acceptance of Austrian sovereignty. This offer would enable the nobility to receive the great power support it wanted without having to sacrifice its traditional position of dominance in Belgian government and society (either to the Belgian lower and middle classes or to another foreign power). His offer split the conservative Belgian nobles into two factions – those who wanted Belgium to remain independent and those who favored a restoration of Austrian rule. While the first faction continued to sustain Belgian independence for the time being, Leopold's offer narrowed the domestic base of resistance to an Austrian invasion. In this way his strategy weakened the *noble* base of the Belgian independence movement enough to enable a military invasion to restore Austrian rule.

This victory came at a stiff price, however. For the nobles who wanted Belgium to remain

independent were rendered more dependent on the lower and middle classes for support. This shift broadened the *popular* base of the Belgian independence movement and thereby raised the risk that it would forge tighter links with the radical Belgian émigrés in northern France and the French revolutionaries in Paris. The crushing of Belgian independence thus raised the risk of revolutionary contagion from France into Belgium, and thereby saddled Austria with the problem of how to assuage the Belgian democracy activists – whose insistence on fundamental reforms was only increased by the return of Austrian rule. As we will see, Austria's efforts to restore its rule in Belgium only increased domestic support for the Belgian democracy movement and drove it toward more violent tactics.

In preparation for a military campaign into Belgium, Leopold needed to settle tensions with Prussia and Poland to the north and the Turks in the east. After he offered Prussia a settlement of the Polish question, Berlin's main demand was that he settle with the Turks too. Since Leopold wanted to turn west to meet the Belgian and French threats, he readily agreed. In July 1790 the Convention of Reichenbach, signed by Austria and Prussia, settled the Polish question and freed Austrian troops to march on Belgium. Austria immediately asked the French Foreign Ministry for permission to allow Austrian troops to cross French soil on the way into Belgium. This request became public in France and provoked fears of an Austrian invasion to overturn the French Revolution – and fears that the French diplomats and the Foreign Ministry were colluding with Austria to mount a counterrevolution. In response the French constituent assembly established a diplomatic committee that would oversee the Foreign Ministry and ensure that its communications and decisions served the popular will. In September Austria signed a truce with the Turks that freed more troops for the Belgian campaign. By the end of the year Austria had crushed the Belgian revolution by force – and Austria used military force again to crush more popular revolutions in Liege and Basel in early 1791. These developments only exacerbated French fears of a counterrevolutionary invasion led by Austria. This threat perception was fueled by the revolutionary rhetoric in France which claimed that France was obligated by international treaty to protect the lower classes beyond its borders from the German princes who were bent on crushing popular movements.<sup>89</sup>

---

<sup>89</sup>Savage 1998, p. 250.

## **9. Developments in France in the second year of the revolution (summer 1790 to May 1791)**

### **9.1 The French King's attitude toward the French revolutionaries as an index of their radicality**

In the second year of the French Revolution, its domestic goals became more ambitious. It was not clear how radical the French revolutionaries' agenda would become. The French king adopted a wait-and-see attitude. As long as their agenda remained moderate, he was willing to cooperate with them. By December 1790 he began to suspect they might make a radical turn, so he formulated a contingency plan in case they did. He would escape from Paris and lead a counter-revolution. In the meanwhile, until his suspicions might be confirmed, he continued to wait and see – and to cooperate with the revolutionaries without making too many concessions. Austria's ruler, Emperor Leopold, followed the French king's judgment on this question. Leopold conditioned Austrian action on the French king's moving first.

### **9.2 The Belgian problem and Vienna's expectations of the French revolution on the eve of its radical turn**

By the summer of 1791 open revolt was breaking out in Belgium and some Belgians saw the French as potential liberators. This development made the Austrian leadership increasingly concerned about how the French Revolution would develop within France. For if the revolutionaries were to turn radical in their domestic agenda, that might increase French support for the Belgian émigrés. In turn the émigrés might increase their support for their fellows in Belgium. And any revolution in Belgium might spread to the other Austrian dominions of Bohemia, Hungary and Central Austria. In short, any radical turn within France portended an increased risk of revolution within Austria.

This compound threat focused the Austrian leadership on two related questions. Would the domestic agenda of the French revolutionaries remain moderate or turn radical? Would the French revolutionaries maintain the right to support the Belgian émigrés residing on French soil? As long as the French Revolution remained moderate, French support for the Belgian émigrés was not expected to be strong. So a French claim of the right to support

the émigrés would not be threatening to Austria. If the French Revolution turned radical, however, French support for the Belgian émigrés could become stronger. And a French claim of the right to support the émigrés would appear more threatening. Hence, as long as the French Revolution remained moderate, Austria could tolerate a French claim of the right to support the Belgian émigrés. If France claimed this right, however, then Austria could not tolerate a radical turn of the revolution within France. If it turned radical, then Austria would have to demand that France either moderate its revolution or stop supporting the Belgian émigrés.

As the French Revolution's domestic goals became more ambitious, Austrian fears of revolutionary contagion increased accordingly. Military action to overturn the revolution went from an option of last resort to an impulse of first resort (implication 2.1). Austria started responding to bad news from France by issuing demands for political moderation backed by threats of military action (implication 2.2).

## **10. Radicalization in France and escalation of the international dispute between France and Austria (June 1791 to September 1791)**

In June of 1791 the French Revolution took a radical turn. The French king fled from Paris with his family. The revolutionaries caught him, returned him to Paris in political captivity, and suspended his executive authority pending his agreement to their demands. Monarchs throughout Europe were dismayed at the rough treatment of the king and suspension of royal authority, Leopold the most of all. The radicalization in France triggered an escalation of the international legal dispute between France and Austria over the right of one country to interfere in the domestic politics of another country. The course of events followed the pattern laid out by the Contractual Realist theory.

### **10.1 the "Flight to Varennes"**

In June, 1791, the French king decided that the revolutionaries had gone too far in dismantling feudal institutions. They were undermining the foundations of law and order

so much that the authority of executive institutions now seemed empty. He decided that further cooperation with the revolutionaries to establish a constitutional monarchy was impossible. He attempted to escape Paris with the royal family, but was intercepted by the revolutionaries and returned to Paris in captivity. Suspecting that the King was in league with Austria to mount a counterrevolution through military invasion, the revolutionaries formally suspended the King's executive authority.<sup>90</sup> Thousands of French army officers deserted in protest against the mistreatment of the King and suspension of his authority. The French noble émigrés moved their headquarters to Koblenz to be closer to the French border. Throughout France there was fear that the émigrés and the Austrians would mount an invasion to rescue the king and overturn the revolution.<sup>9192</sup>

"The left assumed that the [king's] flight to Varennes had been part of an international conspiracy to crush the revolution; and that, indeed, was what the Queen and the émigrés had hoped to forge. When the flight failed, all sides assumed that the German powers would now redouble their efforts, and that the crumbling French army would not be able to stand up to them." (Doyle 2002: 153)

The French revolutionaries mobilized national guards, organized committees to uphold revolutionary discipline, and formed clubs to solicit oaths of loyalty to the nation.<sup>93</sup>

## **10.2 Austria's reaction to the radical turn in France**

### **10.2.1 Perception of an increased threat of income losses**

While monarchs throughout Europe were fear-stricken by the flight and capture of the French king, the Austrian ruler was more concerned than any of them. Leopold viewed the suspension of Louis' executive authority as a sign that the institution of monarchical government itself was under threat in France and therefore in Europe as a whole. He reasoned that Aus-

---

<sup>90</sup>Hochedlinger 2003, pp. 305–6.

<sup>91</sup>Clapham 1899, p. 110.

<sup>92</sup>Hochedlinger 2003, p. 305.

<sup>93</sup>Clapham 1899, p. 88.

tria must stand firm in defense of the pan-European law of the divine right of kings, or else France would continue to violate that right intentionally to inspire revolutions throughout Europe, which in turn would subvert the existing political order and impose heavy income losses on the Austrian monarchy and nobility (implication).

”Nothing, it seemed to [Leopold], could now prevent the ’democratic party’ from assuming full power in France and establishing unlimited democracy. And if this was to happen, then surely no European sovereign could ever again be confident ’that he was leaving an undamaged crown to his successors’.”<sup>94</sup>

Leopold’s fear of revolutionary contagion was shared by the Austrian Chancellor,<sup>95</sup> the Governor of Belgium<sup>96</sup> and Vienna’s most well-informed diplomats.<sup>97</sup> They too sensed the French revolution was turning radical and threatened to inspire a wave of revolutions against monarchical governments throughout Europe. One historian summarized the report of Austria’s envoy in France this way.

”The French Revolution had unleashed the irrational side of human nature and this made it impossible to predict accurately the course that it would take. The only certainty was that, whatever course it followed, it would be dangerous for the rest of Europe. . . . The revolution. . . was a serious danger that must be studied carefully before it could be confronted effectively.”<sup>98</sup>

The ruling circle in Vienna concluded that if the French Revolution promoted its ideology beyond French borders, then military action to overturn the revolution would be necessary.<sup>99</sup>

Austria did not need to initiate war immediately, however, because neither the French ideology nor the French military were threatening enough to warrant it yet. Baron Thugut,

---

<sup>94</sup>Here Wangermann is summarizing a report of Leopold’s view written by Prince Kaunitz, the Austrian Chancellor of State. The internal quotations are the exact words of Kaunitz’s report. Wangermann 1969, p. 66.

<sup>95</sup>Clapham 1899, p. 57.

<sup>96</sup>Clapham 1899, pp. 110–12.

<sup>97</sup>Roider 1987, pp. 86–8, 90.

<sup>98</sup>Roider 1987, p. 90.

<sup>99</sup>Roider 1987, p. 86.

a top Austrian advisor, recommended a policy of promoting the moderates within French domestic politics in the hope they would moderate the revolution from within. But even if Austria needed to initiate war eventually, the only permanent solution to French radicalism would be moderation from within.<sup>100</sup>

Austria's perception of a threat of revolutionary contagion was fueled by the international legal disputes between Austria and France. The two states were in deep dispute over the right of each country to harbor militant political refugees from the other state. Each state's exercise of its legal right to host foreigners was perceived by the other state as a violation of its legal right *not* to be threatened by émigré military forces harboring beyond its borders. Each state's *claim* to have the right to host foreigners was perceived as an *intention* to violate the other state's right not to be threatened by them, giving each state the suspicion that the other state was committing the violation intentionally for its own gain. The news from Belgium over the summer of 1791 justified Leopold's fear of revolutionary contagion fueled by international legal disputes. Belgian émigrés harboring in northern France were agitating for a democratic revolution in Belgium. Their leader sent a petition to the French constituent assembly in Paris to request assistance. The Governor of Belgium reported that many Belgians were looking to France as a liberating power – and some districts were already in open revolt against the Austrian regime in Belgium. Simultaneously, many French noble émigrés were sheltering in Belgium and recruiting fighters for a counterrevolutionary campaign into France. The Belgian governor sensed that their presence posed a threat to the effort to consolidate Austrian rule in Belgium, simply because the French revolutionaries might need to invade to stop the French emigres from launching a counterrevolutionary campaign into France.<sup>101</sup>

### 10.2.2 Policy responses

The Austrian regime responded with moderation on the domestic front but hostility on the foreign policy front. Domestically, Leopold maintained his flexible approach to the peasantry's aspirations and continued to push the nobles for proposals to commute peasant

---

<sup>100</sup>Roider 1987, pp. 86–8.

<sup>101</sup>Clapham 1899, pp. 110–12.



labor obligations and widen representation in the provincial assemblies. On the foreign policy front, however, Leopold and his ministers had a strong reaction that shifted the regime's policy toward France "radically".<sup>102</sup>

Leopold's ministers agreed with his assessment that the time had come to take action to force a moderation of the French Revolution. Abandoning the patient and judicious approach, Austria suspended diplomatic relations with France and adopted a strategy of issuing threats to use military force to intimidate the French revolutionaries into moderating their domestic political agenda.<sup>103</sup> On July 6, 1791 Leopold issued a public declaration that the captivity of the French king proved "the illegal character of the whole revolution and compromised directly 'the honor of all sovereigns and the safety of all governments'; [and therefore] concerted action on the part of the [great] powers to 'limit the dangerous extremes of the French Revolution' was most urgently required."<sup>104</sup> The great powers needed to intervene in French domestic politics to halt "the progress of a contagion of discontent, insubordination and revolt" fueled by "the unconscionable violence of a blind populous [acting] without any restraint."<sup>105</sup> Ten days later, Leopold's Chancellor issued another declaration emphasizing the risk of contagion and explaining that the goal of great power involvement was to guarantee that the French revolutionaries modified the country's constitution "in accordance with the freely expressed wishes of the [French] King."<sup>106</sup> While on the surface the goal of these declarations was to protect the French royal family, the more fundamental political goal was to ensure that monarchical government was maintained in France.<sup>107</sup>

Leopold took a number of steps to form a great power coalition that could coerce a moderation of the French Revolution by threats of force. Most importantly, he opened negotiations with Prussia to establish the coalition and convinced Berlin to join in issuing a demand and threat to France. Austria and Prussia demanded that France stop violating their legal rights and justified the demand by citing international law. They also claimed the

---

<sup>102</sup>Hochedlinger 2003, p. 306.

<sup>103</sup>Hochedlinger 2003, p. 306.

<sup>104</sup>Clapham 1899, pp. 55–6.

<sup>105</sup>Hochedlinger 2003, p. 306.

<sup>106</sup>Clapham 1899, p. 57.

<sup>107</sup>Roider 1987, p. 91.

right to use military force to defend their rights and threatened to use force if their demand to cease the violations was not met. In their joint "Declaration of Pillnitz", Austria and Prussia demanded that the French revolutionaries stop violating the French King's divine right to rule and justified the demand by reference to the pan-European law of monarchical government. They demanded that the French revolutionaries respect the French royal family and maintain a monarchical form of government; and if they did not, then a concert of great powers would take military action to restore both the royal family and the monarchical form of government to France.<sup>108</sup> Austria had no intention of taking military action without Britain's military support – which was unlikely. So the declaration was *not* a signal that Austria intended to go to war. (It was not even clear that Leopold wanted the declaration to be publicized in France.<sup>109</sup>) Rather the document was a joint decision by Austria and Prussia to declare their preference that moderation prevail in France, to lobby the other great powers to support this preference, and to watch French politics closely. To emphasize his pacific intentions, Leopold made public statements that he was not planning a war and the matter would be resolved as long as the French King and the constituent assembly reached agreement on the new constitution they were negotiating.

At the same time, however, Austria took a number of steps that intensified the international legal disputes between Austria and France. In each case Austria's exercise of its legal rights under international law was perceived by France as a violation of its legal rights under international law. First, Leopold permitted a rapprochement between his regime and the French noble émigrés, no longer viewing them as bitter losers in the inevitable emergence of a constitutional monarchy in France but as helpful advisors in the effort to tame the French radicals and moderate the French revolution. Second, Austria and the German princes began to take the legal disputes in Alsace and Avignon more seriously after the flight and capture of the French king in June.<sup>110</sup> On July 17 Prince Kaunitz sent a letter to the Austrian embassies to highlight the legal disputes over Alsace and Avignon and the problems raised by the French revolutionaries' disrespect for international treaties, wild rhetoric and

---

<sup>108</sup>August 27, 1791. Clapham 1899, pp. 76–7.

<sup>109</sup>Clapham 1899, p. 77.

<sup>110</sup>Hochedlinger 2003, p. 306.

threatening military build-ups.<sup>111</sup> On August 6, the Imperial diet of the Holy Roman Empire decreed that France's absorption of Alsace was illegitimate, null and void.<sup>112</sup> "For this decree Leopold was in part responsible."<sup>113</sup> Yet he refrained from ratifying it at first<sup>114</sup> to avoid provoking France further after the Declaration of Pillnitz. His primary goal in threatening France was to empower the French moderates to rein in the revolution. Yet, as we will see, when the radicals outmaneuvered the moderates and adopted a hard-line policy toward Austria, Leopold would ratify the decree.

### **10.3 French domestic politics from the "Flight to Varennes" to the Declaration of Pillnitz**

After the king fled Paris and was returned in captivity, one of the revolutionary factions called for deposing him. In response he authorized his brothers the Prince of Condé and the Count of Artois, heirs to the throne, to negotiate an agreement with the great powers "to check the revolution" on the condition they agree that "negotiations [with the revolutionaries] were to precede force." On July 17 the faction calling for his deposition was defeated and the revolutionary leaders in the constituent assembly swung back toward moderation, renewed loyalty to the king and establishment of a constitutional monarchy. For the next few months the constitution under negotiation in the constituent assembly was the central issue in domestic politics. The search for a moderate middle ground was difficult. The hard-line aristocratic royalists wanted no constitution at all, only a return to the absolute monarchy. The radical republican Jacobins wanted no monarchy at all, only a republic as in Switzerland, the Dutch Netherlands and the United States. The king was caught between these two extremes. While he wanted to make as few concessions as possible to the revolutionaries, he was resolved to accept whatever constitution the assembly produced – now that the moderates in the assembly were prevailing. Their proposed constitution might be more liberal and impose more limits on royal power than he wanted, but rejecting it would only

---

<sup>111</sup>Clapham 1899, 57n.

<sup>112</sup>Clapham 1899, p. 21.

<sup>113</sup>Clapham 1899, p. 130.

<sup>114</sup>Clapham 1899, p. 130.

make matters worse.<sup>115</sup> In sum, the king's willingness to compromise with the revolutionaries had limits. And he had a strong common interest with the great powers: To moderate the revolution even if this required a threat of military invasion.

With the help of an advisor, the king found a strategy for protecting his authority from the revolutionaries without opposing them so openly that he might appear disloyal to the nation. The adviser summarized the strategy neatly.

"A second revolution was inevitable, and... its results would be either (1) the modification of the constitution in the interests of the king or (2) the establishment of a republic. So the king must honestly carry out his promises [to the assembly and the nation] and thereby win popularity. Then, should the state of anarchy [that would be caused by the impending second revolution] continue, it would become evident that the fault lay not in him but in the laws [passed by the revolutionaries]. By this means the monarchy and not the republicans would benefit by the [second] revolution of the near future. If the king failed to win popularity [by contrast, then] the throne would fall, and France would eventually pass into the hands of a dictator."<sup>116</sup>

This was a strategy of "temporary concession"<sup>117</sup> to the revolutionaries in the hope of buying time for a restoration of monarchical power.

The king's strategy was based on his expectation – and hope – that the moderate center of French politics would not hold: The moderate royalists and moderate revolutionaries would not be able to reach a sustainable constitutional settlement which would retain the loyalty of the political nation and population at large. His strategy might work as long as the center did not hold *and* the people turned back to the monarchy. But if the center did not hold and the people turned to the radical Republicans, his only other option would be to call in the great powers to moderate the revolution. Thus the king would be caught in a bind by any event that moved the assembly and the revolution to the left, because everyone

---

<sup>115</sup>Clapham 1899, pp. 86, 87, 91.

<sup>116</sup>Clapham 1899, p. 92.

<sup>117</sup>Clapham 1899, p. 92.

knew that his willingness to make concessions to the revolutionaries was limited. So any movement to the left by the revolutionaries would raise popular suspicion that he preferred to call in the powers than to make more concessions to them. Yet of course such suspicion would undermine his effort to gain popularity with the people.

The moderate royalists and the moderate revolutionaries in the assembly did indeed form a coalition to push through the final agreements needed to formulate a constitution that would secure the king's assent. Approved by him on September 13, 1791, this constitution guaranteed the monarch's role as the head of the executive but distributed government authority accountably between the monarch and a new Legislative Assembly. The constitutional settlement appeared to resolve France's domestic and international problems. In Vienna, Emperor Leopold and Prince Kaunitz concluded that the problem of domestic radicalism within France was solved – and signaled as much to France.<sup>118</sup> King Louis himself was not so sure, because he saw domestic politics at closer range. The day before he approved the constitution, the constituent assembly issued a decree that formally annexed the papal enclaves of Avignon and Venaissin to France. While this act was legal in so far as the enclaves were fully within France's contiguous borders, the act also violated the papacy's sovereignty rights.

## **10.4 France's response to the Declaration of Pillnitz**

### **10.4.1 Domestic political response**

A few days after King Louis approved the constitution, the Declaration of Pillnitz was made public in France.<sup>119</sup> Despite Leopold's reassurances that Austria had pacific intentions, the Austrian threat issued in the Declaration caused immense damage in French domestic politics. Austria's claim to have the right to moderate the French revolution by force was perceived as an intention to violate France's right to political independence by interfering in its domestic lawmaking. France suspected that Austria was committing the violation intentionally for its own gain. The revolutionary leaders in the legislative assembly contended that

---

<sup>118</sup>Roider 1987, p. 92.

<sup>119</sup>Clapham 1899, pp. 77–8.

the Austrian threat to intervene constituted direct evidence that Austria asserted the right to interfere in French domestic politics with military force. The more radical revolutionaries, who distrusted the king and wanted to overthrow the constitution and the monarchy entirely, seized on Austria's threat as *prima facie* evidence that Austria was certain to lead a coalition of great power armies into France to crush the revolution. Rumors circulated throughout France that Austria was planning to invade. France reasoned that it must reject Austria's demand for a moderation of the revolution, for if it were to give in to the demand, then Austria would continue to demand changes in France's domestic lawmaking in violation of its right to domestic political independence – at increasing cost to its lower and middle classes who had just been emancipated by the revolution's new laws (implication). The Austrian demand for a moderation of the revolution was intolerable to the average Frenchman. He felt no obligation to sacrifice the Revolution's domestic goals simply because another country demanded it. Capitulation might lead to a restoration of royal absolutism and the feudal system.

France reasoned that it would be less costly to wage a war to prevent and deter further violations of its political independence than to allow Austria to continue such violations intentionally (implication). The revolutionary leaders argued to the assembly that sacrificing the Revolution's domestic goals would be so costly that it would be worse than a war – even a long and costly war to the death. The leaders' speeches received huge popular support, because a majority shared their view that even a costly war would be better than settling for a peace in which the revolution would be overturned and the feudal system restored.

#### **10.4.2 Foreign policy response**

When the new legislative assembly convened on October 1, it scheduled a debate on the problem of the French noble émigrés harboring in Germany. In anticipation of the debate the king pursued his strategy of trying to increase his popularity with the people. He publicly implored the émigrés to return to France and begged the princes among them – who were his brothers next in line to the throne – to abandon their planning for a military invasion. Yet privately he was vacillating between the tactic of continuing to make temporary concessions to the revolutionaries and the tactic of calling the great powers into action to moderate

the revolution.<sup>120</sup> So he told the princes privately that he still wanted to see a great power intervention. Hence they did not return to France and the assembly was not reassured by his public entreaties.<sup>121</sup>

In the legislative debate the *Brissotins* made a strong case against the émigrés. They perceived the German princes' exercise of the legal right to host the French noble émigrés as a violation of France's legal right *not* to be threatened by militant political émigrés harboring beyond its borders. The *Brissotins* also perceived the princes' rejection of previous demands to disperse the émigré forces as a threat that the émigrés, the princes and Austria intended to launch a counterrevolutionary invasion into France to overturn the revolution (implication). As *Brissotin* himself asked, "Why had there been a Congress at Pillnitz and why had the émigrés been admitted to it? Was it true that the Pillnitz Congress had agreed to destroy the French Constitution?"<sup>122</sup> He concluded that "The [German] neighbors of France must be taught to render sincere homage to her constitution and, as a natural consequence, to withdraw all patronage and assistance from those who were in arms against it."<sup>123</sup>

The legislative assembly and the king demanded that the German princes stop allowing the French noble émigrés to organize military forces on their territories and justified the demand by citing an international law which prohibited such activity (implication). The assembly claimed the right to use military force to defend all of these rights, deployed force in anticipation of the need to do so, and justified the deployment as necessary to do so (implication). On November 9 the assembly decreed that if the noble émigrés were still assembled on the French frontiers on January 1, the assembly would convict them of conspiracy, confiscate their property and issue sentences of execution. The diplomatic committee of the legislative assembly reported that the German princes were violating both imperial law and international law by allowing the French noble émigrés to organize military forces on their territories.<sup>124</sup> The decree also directed the assembly's diplomatic committee

---

<sup>120</sup>Clapham says the king vacillated between these tactics through September and October of 1791. Clapham 1899, p. 92.

<sup>121</sup>Clapham 1899, p. 114.

<sup>122</sup>Blanning 1986, p. 100.

<sup>123</sup>Clapham's paraphrase of *Brissotin's* words. Clapham 1899, pp. 114–5.

<sup>124</sup>The report was issued on November 22, two weeks after the decree was issued. Presumably the interna-

to provide the French King with options for combating foreign rulers who sheltered French noble émigrés. The king vetoed the decree but heeded its intent by asking the German princes to expel the émigrés from their territories.<sup>125</sup> He threatened one German prince, the Elector of Trier, with war if he did not evict the émigrés within a month.<sup>126</sup> On November 14 the French minister of the interior sent a message to Vienna to ask Emperor Leopold to pressure the German princes to expel the French émigrés.<sup>127</sup> The minister informed the legislative assembly a few days later that Leopold had evicted the French noble émigrés from Belgium and he had asked Rhenish princes to do the same.<sup>128</sup>

France also demanded that the German princes stop demanding compensation for their estates in Alsace and justified the demand by citing the Treaty of Westphalia in which Alsace was ceded to France (implication). On November 29 the assembly issued a resolution calling on the king to demand that (1) the Austrian Emperor and the Elector of Trier "disperse the fledgling armed forces of émigrés" and (2) the German princes stop insisting on compensation for their estates in Alsace. The assembly also authorized the deployment of French troops to the eastern frontier with Germany to enforce France's rights to a dispersal of the émigré forces and full sovereignty over Alsace.<sup>129</sup>

On the same day the assembly declared that any French priests who did not pledge allegiance to the new French constitution within a week would be dismissed from their positions. Upon seeing this decree the French king decided the revolutionaries had become too radical and sent requests to the great powers to convene an "armed congress" that could intervene in France to moderate the revolution.<sup>130</sup> At the same time, though, he made accommodating gestures to domestic audiences to demonstrate his fidelity to French national interests. He promised the assembly (1) he would threaten the rest of the German princes

---

tional law cited in the report had been on the books for more than two weeks and the spirit of the law was known intuitively by the assembly when it passed the decree. Clapham 1899, p. 117.

<sup>125</sup>Clapham 1899, pp. 116–7.

<sup>126</sup>Clapham 1899, p. 127.

<sup>127</sup>Clapham 1899, pp. 118, 131.

<sup>128</sup>November 16. Clapham 1899, p. 117.

<sup>129</sup>Roider 1987, p. 92.

<sup>130</sup>Clapham 1899, pp. 120, 122.



with war if they did not evict the émigrés harboring on their territories and (2) he would ask Leopold to press the German princes on the émigré issue (in his capacity as Holy Roman Emperor).<sup>131</sup> The French minister of war informed the assembly that the King was sending 150,000 troops to the German border to render his threats to initiate a war over the émigré issue credible.<sup>132</sup>

### **10.5 Austria's response to the French claims, demands and threats**

The French constituent assembly formally annexed the papal enclaves of Avignon and Venaissin to France in mid-September. Austria perceived the French assertion of sovereignty rights in Avignon and Venaissin as a violation of papal sovereignty, and suspected that France intended to violate Austrian sovereignty next by invading Belgium. A leading French royalist reported to Austria's top diplomat in Belgium that "in view of the recent theft of Avignon and Venaissin. . . , it would be in no way surprising should the new assembly order an attack on Belgium and Liege."<sup>133</sup> Leopold knew the Belgian population was already rebellious, so he feared that an attack might spread the revolution into the Belgian territory. He was of two minds on the legal issue however. He acknowledged that France's legal case had some merit, but maintained that Austria's legal case had merit nonetheless (implication). Leopold acknowledged France's legal right to a dispersal of the French noble émigré forces that were harboring in Belgium. On October 22 he ordered the dispersal of all armed French groups from Belgium and the prohibition of any military preparations by French citizens on Belgian soil. His goal in making the concession was not simply to eliminate French excuses for an attack on Belgium but to bolster the French king's effort to cooperate with the French revolutionaries.<sup>134</sup> He also pressed the German princes to disperse the armed French groups in their territories (see above). At the same time, however, Leopold perceived France's claim of the right to shelter Belgian émigrés in northern France as a threat to violate Belgian political independence, that is, a threat that the émigrés and the French would invade Belgium

---

<sup>131</sup>December 14. Clapham 1899, p. 127.

<sup>132</sup>December 14. Clapham 1899, p. 128.

<sup>133</sup>Clapham 1899, p. 112.

<sup>134</sup>Clapham 1899, pp. 111–2.

to foment a revolution there. Leopold hoped the new legislative assembly in Paris would reciprocate his concession by stopping all military support for the liberal Belgian émigrés harboring in northern France.<sup>135</sup> The assembly did reciprocate two months later by prohibiting the Belgian émigrés from convening military assemblies on French soil (see below). Neither the French nor the Austrian commitment to prohibit émigré military assemblies resolved the dispute over the right of foreign interference in domestic politics, however. For as we will see, each power maintained the right to host political émigrés on its soil as long as they did *not* convene military assemblies.

When the Austrian Council of Ministers heard the anti-Austrian rhetoric emanating from Paris, it advised Leopold to take a firm stance in defense of the German princes' rights.<sup>136</sup> Following this advice Leopold and Kaunitz switched from a policy of accommodation to one of intransigence in the disputes over German rights. On December 3 Leopold ratified the Imperial Diet's decree proclaiming the French seizure of Alsace as null and void, and he sent a strong diplomatic note to the French Foreign Ministry on the issue.<sup>137</sup> He demanded that France stop violating the rights of the German princes in Alsace and justified his demand by citing international law (implication). Specifically, he claimed that France had no right to confiscate the German princes' Alsatian estates and pay the princes compensation, because the Treaty of Westphalia protected the estates from encroachment by the French government.<sup>138</sup> He added that he had ratified the Diet's decree proclaiming such confiscations null and void.<sup>139</sup> And he suggested that Austria perceived French claims of the right to confiscate the Alsatian estates as evidence of a French intention to violate the princes' right to political independence more generally. In particular he claimed that France was trying to incite discontent and revolution in the territories of princes throughout Germany.<sup>140</sup>

A few weeks later, on December 21, the Austrian Chancellor sent another strong note

---

<sup>135</sup>Clapham 1899, p. 112.

<sup>136</sup>Clapham 1899, p. 131.

<sup>137</sup>Clapham 1899, p. 21.

<sup>138</sup>Clapham 1899, pp. 21, 131.

<sup>139</sup>Clapham 1899, pp. 131–2.

<sup>140</sup>Leopold wrote the note on December 3, 1791. The French legislative assembly referred the note to its diplomatic committee on December 24. Clapham 1899, p. 132.

to France on the émigré issue. He acknowledged that France's legal case had some merit, but maintained that Austria's legal case had merit nonetheless (implication). He offered some concessions in acknowledgment of the merits of France's legal case, but maintained Austria's demand for concessions as well (implication). On one hand, he acknowledged that France had a legal right to dispersal of the noble émigré forces from Germany and Belgium. And he offered assurances that Emperor Leopold and the Elector of Trier had already dispersed the forces from their territories.<sup>141</sup> On the hand, he claimed that Austria had the right to use military force to defend the German princes and the right to mobilize force in anticipation of the need to use it. And he justified the mobilization as necessary to prevent France from violating German rights (implication). Specifically, Leopold would provide military protection to the German princes (in his capacity as Holy Roman Emperor), because Austria was deeply uncertain about French intentions toward the princes. "The provinces and municipalities of France were so full of violence and misrule that raids into German territory were by no means unlikely."<sup>142</sup> Austria had therefore ordered its Belgian military force to defend the Electorate of Trier from any sort of French aggression.<sup>143</sup>

In sum, Leopold and Kaunitz reasoned that they must oppose the French force threatening the Elector of Trier with an Austrian force, for if the French force were allowed to go unopposed, then France would use its force to violate the rights of princes elsewhere in Germany - at increasing cost to the nobilities and monarchies of Germany and Austria (implication). Leopold believed that France was inciting discontent and revolution in the territories of the German princes and feared it would do so more widely if not opposed. Kaunitz believed that domestic order was breaking down in France and feared it would lead to "raids into German territory" if an Austrian force were not sent to protect the German princes.

Like Leopold, the French legislative assembly was of two minds on the émigré issue. It acknowledged that Austria's legal case had some merit, but maintained that its own legal case had merit nonetheless. The assembly issued a decree on December 21 which

---

<sup>141</sup>Clapham 1899, p. 132.

<sup>142</sup>Clapham 1899, p. 133.

<sup>143</sup>Clapham 1899, p. 133.

prohibited the liberal Belgian émigrés on French soil from convening military assemblies, thereby reciprocating Austria's concession on the same point (see above).<sup>144</sup> On the other hand, the assembly permitted the Belgian émigrés to remain in France and continue their planning for a new revolution in Belgium. They were encouraged by the more radical French revolutionaries who expected that a revolution in Belgium would make it more difficult for Austria to mount an invasion into France. The French and Belgian revolutionary clubs were in close contact for the purpose of promoting a Belgian revolution – a fact which the Belgian governor reported to Vienna on December 24.<sup>145</sup> Thus, while France reciprocated on the formal level by committing to prevent émigré military assemblies on its soil, France permitted the émigrés to remain and to organize a political threat to Belgium in other ways.

The German princes followed suit. The Elector of Trier, like Emperor Leopold and the French legislative assembly, was of two minds on the émigré issue. He acknowledged that France's legal case had some merit, but maintained that his own legal case had merit nonetheless. He conceded that he had an obligation to prohibit the French noble émigrés from organizing military forces on his territory. He assured the French foreign minister that he had dispersed all of the French noble "malcontents" and stopped all of their military preparations on his territory.<sup>146</sup> Like France and Austria, however, the Elector of Trier claimed he had the right to shelter French political émigrés as long as they did *not* muster military forces or preparations on his territory. He perceived France's deployment of force and claims of the right to use force to disperse the émigrés as a military threat to violate his rights to territorial sovereignty and political independence (implication). "The violent language of the new [French] assembly touching the protection [that the German princes] extended to the emigrants was more than discomfoting. It caused much alarm among the prince-bishops and other ruling personages. The Elector of Mainz...told Kaunitz... that it was an emperor's plain duty to protect the outlying circles of the empire against the aggression of foreign democrats."<sup>147</sup> The Elector of Trier appealed to Leopold, in his capacity as Holy

---

<sup>144</sup>Clapham 1899, p. 112.

<sup>145</sup>Clapham 1899, pp. 111–3.

<sup>146</sup>Clapham 1899, p. 132.

<sup>147</sup>Clapham 1899, pp. 130–1.

Roman Emperor, for military protection from French aggression.<sup>148</sup>

In sum, Austria responded to the French threats to Belgium and Germany in a mixed way. Austria acknowledged the legal merits of France's case and made an appropriate concession - which in turn induced France to acknowledge the legal merits of Austria's case and make a reciprocal concession. Yet Austria also stood firm on the remaining merits of its legal case - which induced France to stand firm on its remaining legal case. While the mutual concessions did little to defuse the dispute between the two states, the mutual intransigence did much to escalate it. As we will see, the cause of the next round of escalation was again news from Paris that the French Revolution was becoming more radical.

### 10.5.1 France's response to the Austrian claims, demands and threats

In his note of December 21, Kaunitz insisted that Austria had the right to protect the German princes from French aggression. France perceived Austria's claim of the right to protect the German princes as an intention to violate France's right not to be threatened by militant political émigrés harboring beyond its borders. And France suspected that Austria intended to mount a counterrevolutionary invasion of France in violation of its domestic political independence. Many people in France believed that Emperor Leopold was colluding with the French noble émigrés in Germany to mount a counterrevolutionary invasion of France. Hence Kaunitz's note appeared to confirm their suspicions about Leopold's intentions and provoked a new measure of ill-will toward Austria even among moderates.<sup>149</sup> The Jacobins, who were more radical, spread the idea that King Louis and his minister of war intended to take advantage of any war with Austria to strengthen the French monarchy at the expense of the revolutionaries.<sup>150</sup> In fact the royalists did hope to strengthen the French monarchy through a war, but they disagreed about how to do it. The moderate royalist party – the *Feuillants* – hoped that when Austria intervened with military force it would give the French monarchy assistance and protection in taming the revolutionaries. By contrast the minister of war hoped the monarchy would be strengthened by galvanizing a national effort to repel

---

<sup>148</sup>Clapham 1899, p. 132.

<sup>149</sup>Clapham 1899, pp. 136–7, 151–2.

<sup>150</sup>Clapham 1899, pp. 137–8.

Austrian interference.<sup>151</sup>

France reasoned that it must stand firm in defense of its legal right to domestic political independence or else Austria would continue to violate the right intentionally and impose increasing income losses on it (implication). The *Brissotins* presented this logic to the legislative assembly in the last week of December. One of them "spoke of the gathering of [foreign] troops on the frontiers, of rumored plots against liberty, [and]... the horrid results that a counter revolution and the reestablishment of the noblesse would entail."<sup>152</sup> Brissot himself "urged that [Emperor] Leopold and his fellows must be taught that the French constitution was no affair of theirs."<sup>153</sup> France acknowledged that Austria's legal case for the sovereignty rights of Germany and Belgium had some merit, but maintained that its own legal case for the right to political independence had merit nonetheless. The legislative assembly passed a decree declaring that "the French nation renounced all ideas of conquest [but]... war was being forced on it against its will."<sup>154</sup>

In January the assembly's diplomatic committee concluded that Austria was violating France's rights in three ways. First, Austria was protecting French rebels who were harboring on German territory. Second, Austria was forming a great power coalition with the intention to interfere in French domestic lawmaking. Third, Austria was deploying military force for the purpose of violating French rights.<sup>155</sup> In each case Austria's *exercise* of its own legal right was perceived by the diplomatic committee as a *violation* of France's overlapping legal right (implication). Austria's right to give military protection to the German princes was perceived as a violation of France's right to be free of rebel colonies beyond its border. Austria's right to protect European kings ruling by divine right from French revolutionaries was perceived as a violation of France's domestic political independence. Austria's deployment of military force to protect its rights was also perceived as a violation of France's right to domestic political independence.

---

<sup>151</sup>Clapham 1899, pp. 139–40.

<sup>152</sup>Clapham 1899, p. 134.

<sup>153</sup>Clapham 1899, p. 135.

<sup>154</sup>See also p. 147 on assembly debate of January 17–25, in which a similar proposal to express pacific intentions was voted down. Clapham 1899, p. 136.

<sup>155</sup>Clapham 1899, pp. 144–5.

The assembly debated the problem of how to respond to the Austrian violations for days. Some of the debaters acknowledged that Austria's legal case for the sovereignty rights of Germany and Belgium had some merit, but others maintained that France's own legal case for the right to political independence had merit nonetheless. The Foreign Minister contended that the assembly should refrain from initiating war against the Elector of Trier because he was willing to comply with its demands. To support the case the minister quoted multiple letters from the Elector to the assembly offering reassurances of his good faith in the émigré matter.<sup>156</sup> A law professor sympathetic to the revolution insisted that Emperor Leopold was willing to disperse the French émigré forces from Belgium and the Austrian lands in western Germany, and therefore the assembly should moderate its position and delay action.<sup>157</sup> However, others cited the diplomatic committee's report that Austria was violating French rights and contended that "France should strike suddenly and swiftly... unless a full and satisfactory explanation of her various unfriendly acts were given by Austria before February 10."<sup>158</sup>

France reasoned that it must oppose Austria's force with its own force, for if it were to allow Austria's forced to go unopposed, then Austria would use the force to modify France's constitution in violation of its right to domestic political independence (implication). On January 21, in the middle of the assembly debate, the French Foreign Minister sent a note to Austria to express France's main grievance, to signal its resolve to resist Austrian aggression, and to provide reassurance that the spread of French ideas into Europe was unintentional and harmless. He explained that many people in France interpreted Kaunitz's note of December 21 as a threat to mount an invasion to change the French constitution. But the French people would come together to oppose an invasion. And besides, it was beneath Austria's dignity "to go to war to avenge a few intemperate words or deeds, which were after all but the inevitable outcome of any revolution..."<sup>159</sup>

In the course of its debate the legislative assembly issued two decrees. The first decree,

---

<sup>156</sup>Clapham 1899, p. 144.

<sup>157</sup>Clapham 1899, p. 146.

<sup>158</sup>Clapham 1899, p. 145.

<sup>159</sup>Clapham 1899, pp. 154–5.

passed on January 14, prohibited French domestic actors from cooperating with the French noble émigrés, offering new concessions to the German princes over Alsace, or colluding with Austria to moderate the French Revolution.<sup>160</sup> The second decree, passed on January 25, demanded that Austria stop violating France's legal rights and justified the demand by citing the international law of sovereign rights (implication). The assembly also claimed the right to use military force to defend French rights (implication). Specifically, the decree demanded that Leopold renounce the intention to invade France and interfere in its domestic lawmaking – and renounce all international treaties that contradicted French sovereign rights. If he failed to respond satisfactorily by March 1, then the assembly would regard the failure as a declaration of war.<sup>161</sup> Upon seeing this decree King Louis warned the assembly that it was prohibited by the new constitution from discussing questions of war and peace without a royal request to do so. He also reassured the assembly that he had already sent Emperor Leopold the same basic message that the assembly intended to send by its decree.<sup>162</sup>

In response the legislative assembly passed a motion which demanded that Austria and her allies reveal their true intentions vis-à-vis the revolution. The purpose of the motion was to delineate and defend a legal norm: *You cannot interfere in our domestic politics*. France's demand that Austria clarify its intentions provoked an escalation in the fundamental dispute between the two countries over whether either country had the right to interfere in the other's domestic politics.

### 10.5.2 Austrian fears, demands, justifications and threats (January, February)

The French king had long expected that if the revolutionaries became too radical, he could rely on the European great powers to launch a military intervention to overturn the revolution. The assembly's decrees of November 1791 struck Louis as too radical indeed – proof that "the discontented [revolutionary] factions were ruining the state".<sup>163</sup> He sent letters to the other monarchical powers of Europe to call for an intervention. The Queen also sent a

---

<sup>160</sup>Clapham 1899, p. 145.

<sup>161</sup>Clapham 1899, p. 148.

<sup>162</sup>Clapham 1899, p. 154.

<sup>163</sup>Clapham 1899, p. 122.



letter to the top Austrian diplomat in Belgium to warn that the revolutionaries were likely to declare war against the German princes soon, and Austria should lead a great power invasion to save the institution of monarchical government in France. There was no further excuse for Austrian delay, she added, because the French royal family was not in any personal danger from the revolutionaries.<sup>164</sup> This last comment suggests that her motive in calling in the great powers was not to save her family but to save the institution of monarchy from the threat of republicanism.

After seeing these letters in late December, Leopold and Kaunitz reacted strongly and decided that firm action against France was required. They did not want a war with France but sensed it would be necessary if the more radical revolutionaries continued on their current path.<sup>165</sup> Accordingly Leopold resumed the effort to form a great power alliance to coerce the French revolutionaries into moderating their domestic political agenda.<sup>166</sup> Kaunitz sent a demarche to the French Foreign Minister on January 5 containing a concession, renewed demands and a threat to use force. Kaunitz offered a concession in acknowledgment of the merits of France's legal case, but maintained the prior demand for concessions from France as well (implication). He explained that the French noble émigrés had been dispersed from the Elector of Trier's territory – in response to French concerns – but maintained that France therefore had no right to threaten German territory.<sup>167</sup> He claimed Austria had the right to use military force to defend German sovereignty (implication). Austria would resist any violation of German territory with military force. And if the French revolutionaries continued to issue wild rhetoric and threaten monarchical government at home, then the great powers would intervene in France to prevent the spread of its ideology and revolution into Europe. In Kaunitz's words, the powers would become "united in concert for the maintenance of public tranquility and for the safety and honor of crowns."<sup>168</sup> Leopold and Kaunitz expected that by standing firm to the French radicals and cooperating with the French moderates,

---

<sup>164</sup>Clapham 1899, pp. 123–4.

<sup>165</sup>Clapham 1899, p. 150.

<sup>166</sup>Clapham 1899, p. 149.

<sup>167</sup>Clapham 1899, p. 154.

<sup>168</sup>Clapham 1899, pp. 154–5.

they could induce a moderation of the revolution.<sup>169</sup>

In a memorandum to the Austrian Council of Ministers, Kaunitz explained that it was necessary to stand firm against France for two reasons. The French revolution was becoming more radical at home and more likely to spread into Austrian territory in Belgium. On January 17 the council decided to issue demands to France. The revolutionaries must maintain a monarchical government at home, return the territories of Avignon and Venaissin to the papacy, give the German princes financial compensation for their estates in Alsace, and cease all military preparations on its border with Germany. The council also decided to mobilize an Austrian army to back the demands.<sup>170</sup>

At the same time, however, Leopold and Kaunitz were advised by their best diplomats and the French royalist party to acknowledge the merits of France's case for its rights. The Austrian diplomats with the best understanding of the revolution advised them that the domestic political balance between moderates and radicals was uncertain, and therefore Austrian policy must strike a delicate balance. It was necessary to be firm. Yet "Vienna must be careful about what it announced it was fighting for and against." The French people actually disliked the radical revolutionaries in the legislative assembly – yet they disliked the nobility, the old regime and absolute monarchy even more. Hence Austria should limit its demands and its communications to France accordingly. Moreover, the French leaders were divided in their war aims between those who wanted a limited war against the French noble émigrés in Germany and those who wanted a general war to liberate the lower and middle classes of Europe from feudal nobilities and monarchs.<sup>171</sup> A senior French royalist wrote to Leopold in much the same vein. The revolution's abolition of feudalism was appropriate but the radical revolutionaries were going too far and were not popular as a result. The majority of the French people wanted a constitutional monarchy and not a republic. Austria could help them secure the French monarchy and defeat the republican minority "by means of a firm and sustained policy" – but only if Leopold were to "repudiate all connection with the [French noble] emigrants [in Germany].... To avoid war he must disown publicly all ideas

---

<sup>169</sup>Clapham 1899, p. 150.

<sup>170</sup>Clapham 1899, pp. 149–50.

<sup>171</sup>Roider 1987, pp. 93–4.

of aggression and send no troops into the Electorate of Trier until the Elector had complied fully with the French demands.”<sup>172</sup>

A few weeks later Leopold wrote to the royalist to explain his motives and goals and to offer reassurances that the aims and methods of a great power intervention would be strictly limited. He acknowledged that France’s legal case for domestic political independence had some merit, but maintained that Austria’s legal case for the divine right of kings throughout Europe had merit nonetheless (implication). Specifically, he said he was strongly averse to a restoration of the absolute monarchy in France, but equally averse to the radical revolutionaries who were unruly at home and determined to provoke a war against Austria.<sup>173</sup> Then he offered some concessions in acknowledgment of the merits of France’s legal case for domestic political independence (implication). “The new concert [of great powers]... was based on three principles: no aid was to be given to the emigrants; there was to be no interference with the internal affairs of France, save to protect the king and his family; and no attempt would be made to overthrow the constitution.”<sup>174</sup>

### 10.5.3 French fears, demands, justifications and threats

In early February the assembly passed a decree accusing the French noble émigrés and their advisors of rebellion and treason.<sup>175</sup> The foreign minister tried to restrain war sentiment in the assembly numerous times by insisting that the Elector of Trier had dispersed the noble émigré forces completely.<sup>176</sup>

Leopold and Kaunitz believed that the French population “disliked mob rule and the methods of the extreme democrats” in the assembly pushing for war and preferred a monarchical government and peace with Austria.<sup>177</sup> The Austrian leaders expected they could coerce France into domestic political moderation by sending further diplomatic notes. In

---

<sup>172</sup>The letter was written in late December or early January. Clapham 1899, pp. 151–2.

<sup>173</sup>Clapham 1899, p. 152.

<sup>174</sup>January 31, 1791. Clapham 1899, p. 152.

<sup>175</sup>February 6. Clapham 1899, pp. 134, 145, 163–4.

<sup>176</sup>Clapham 1899, p. 164.

<sup>177</sup>Clapham 1899, pp. 164–5.

the notes of February 17, they acknowledged that France's legal case for a dispersal of the émigré forces had merit – and offered a concession on the issue – but maintained that Austria's legal cases had merit nonetheless – and maintained their demand for concessions from France (implication). As far as concessions, Austria had "discouraged [the French noble émigrés] without any advice from France."<sup>178</sup> But France was "still in a state of disease" of revolutionary ferment and war-mongering, and it was caused solely by the radical republican faction, not by any fear of invasion by the noble émigrés. The radical faction was planning to abolish the French monarchy, fomenting revolution in Belgium, claiming the French legislature had the right to renege on international treaty commitments, and threatening to use military force beyond French borders. The radicals were "a danger for France and a menace to all Europe." Therefore Austria and Prussia would protect "the peace of Europe" until the majority of the French population in favor of monarchical government and peace might defeat the radicals.<sup>179</sup>

While these notes were en route to France, an open political conflict erupted between the hard-line French royalists, who wanted to return to an absolute monarchy, and the moderate French royalists and revolutionaries, who wanted the constitutional monarchy to endure. In late February the minister of war, a moderate royalist, advised the king that his current approach would only weaken the unstable center of French politics and, therefore, the king should abandon the hardline royalists and adopt a more accommodating posture toward the revolutionaries. Then on March 1 the Austrian notes were read to the legislative assembly. The initial reaction was not hostile. "Even within the assembly there was still some hesitation [to initiate a war].... On the day after [the foreign minister's] reading of the notes, a renewal of the [assembly's] diplomatic committee gave the moderates a majority in that body."<sup>180</sup> Yet, because of the rift between absolutists and constitutionalists, the Austrian notes precipitated a dispute between the court and the assembly over the control of French foreign policy.

Up to that point authority over foreign policy had rested with the king, his royalist

---

<sup>178</sup>Clapham 1899, p. 165.

<sup>179</sup>Clapham 1899, pp. 165–6, 180.

<sup>180</sup>Clapham 1899, p. 176.

ministers and the moderate royalist faction in the assembly. The royalist ministers wanted peace with Austria – except for the minister of war (who had advised the king to be more accommodating to the revolutionaries). He thought a war was necessary not for domestic political reasons but to meet the Austrian threat. Since he believed the king's current posture to be counterproductive, he threatened to resign from office unless the king dismissed the one remaining hardline royalist within the ministry (the minister of marine). The king rejected that ultimatum and dismissed him instead. The dismissal of the war minister created an open rift between the remaining ministers – all moderate royalists who wanted peace with Austria – and the moderate revolutionaries in the assembly – who were divided on the question of war and peace with Austria. Which organ of the government would have ultimate authority over foreign policy: the royal court and its ministers or the assembly?

Into this breach stepped the *Girondin* party – the most pivotal of the revolutionary factions. The *Girondins* did not necessarily want to overthrow the king and establish a republic. They actually feared the Jacobins – who did – and wanted to prevent their rise to power. The *Girondins* believed, however, that the French royal court and the royalist faction were colluding with Austria to mount a counter-revolution. There was some truth to this, as the letter that one of the royalists sent to Leopold suggests (see above). The *Girondin* party's members were in favor of a war with Austria for different reasons. While its leaders shared *Brissot's* domestic motives for war – and in this sense were “*Brissotins*” – the rank-and-file did not.<sup>181</sup> In their speeches to the assembly, the *Girondins* attacked the Austrian Emperor, the French court, the royalist ministers and the royalist faction in the assembly. *Brissot*, a natural ally, proposed a motion to impeach the royalist foreign minister on the grounds of “weakness, duplicity and treachery,” and “communicating details of the internal troubles of France to a foreign and hostile minister.”<sup>182</sup> In response the assembly impeached the royalist foreign minister and deprived the royalists of their majority on the diplomatic committee. This outcome, made possible by the general sense in the assembly that the king and the royalists were colluding with the Austrians, deepened the rift between the assembly and the court considerably.

---

<sup>181</sup>Clapham 1899, p. 105.

<sup>182</sup>Note 2. Clapham 1899, p. 178.

The moderate royalists and the moderate revolutionaries tried to negotiate a compromise settlement to bridge the gap. But the effort got nowhere because "the disagreement as to the exact measure of constitutional reform to be aimed at, which had always existed, was now further complicated by differences of opinion with regard to the war policy. . . ." <sup>183</sup> The moderate royalists wanted peace with Austria while the moderate revolutionaries were divided on the question. The king decided that the only way to prove his court's loyalty to France and save his own popularity was to appoint a new ministry composed of *Girondins*. His decision gave the *Girondins* effective authority over French foreign policy. <sup>184</sup> Although the *Girondin* party did not want to overthrow the king and establish a republic, its members in the assembly were "bidding for popularity against the Jacobins" <sup>185</sup> – who did. The "unstable mass of open-minded men, which formed the center of the assembly," [was undecided on the question of war and peace with Austria]. <sup>186</sup> Now the *Girondins* convinced a majority of the assembly to join the pro-war camp. <sup>187</sup>

France demanded that Austria renounce the right and intention to use force (implication). On March 11 the French ambassador to Vienna met with Kaunitz and demanded that Austria repudiate the great power concert. <sup>188</sup> Austria still expected that it could settle its disputes with France peacefully if it could defeat the main opponent to a settlement: the radical republican faction in France. As the British ambassador in Vienna wrote, "It was still 'the general hope, as well as wish, to see matters amicably adjusted with France,' but the adjustment was sought by means of a suppression of 'the Jacobins'." <sup>189</sup> However, Austria reasoned that it must reject France's demand to renounce the use of force, because if it were to give in to the demand, then France would continue to violate its rights intentionally and impose increasing income losses on it (implication). And Austria rejected the demand

---

<sup>183</sup>Clapham 1899, p. 179.

<sup>184</sup>Clapham 1899, pp. 161–3, 177–80.

<sup>185</sup>Clapham 1899, p. 182.

<sup>186</sup>Clapham 1899, p. 105.

<sup>187</sup>Clapham 1899, pp. 179–82.

<sup>188</sup>Clapham 1899, p. 167.

<sup>189</sup>The quote is Clapham's summary of the ambassador's note to London on March 7. The internal quotes are from the note. Clapham 1899, p. 193.

(implication). On March 18 Kaunitz sent a note to Paris to say that Austria denied it had made any military preparations and refused to abandon the great power concert. Austria could not abandon the concert, he claimed, because France had not yet complied with the demands that the concert was organized to enforce. Its non-compliance was driven by the radical republican minority in France, he added, so if the majority that preferred a monarchical government could not suppress that minority, perhaps it "might welcome the friendly help of the powers."<sup>190</sup>

On March 27 the new *Girondin* foreign minister sent a note to Vienna to renew the demand that Austria repudiate the great power concert. He also demanded that Austria demobilize its military forces immediately and threatened that if Austria did not accept the demands by mid-April, then France would "resort to the sternest measures."<sup>191</sup> In response Austria offered a concession, in acknowledgment of the merits of France's legal case, but maintained its demand for concessions from France as well (implication). The Austrian vice-chancellor conceded that Austria had "no wish to interfere in the internal affairs of France." Yet Austria was resolved "to secure some guarantee of strong government in France, since that was essential for the safety of Europe." And Austria was equally resolved to defend the rights of the German princes in Alsace and the rights of the papacy in Avignon and Venaissin.<sup>192</sup>

On April 16 the French foreign minister reported to the king's ministerial council that Austria had failed to reply to the French demands, and this must be regarded as a declaration of war. He asserted that Austria's hostility to France was driven by two factors: the French people's unwillingness to remain slaves to an absolute monarchy and the French king's willingness to let his authority be limited by constitutional law.<sup>193</sup> In short it was the French Revolution which had provoked Austria to organize a great power coalition, mobilize a military force, and threaten to invade France to compel changes in its domestic legal regime. On April 20 the legislative assembly passed a declaration of war against Austria and

---

<sup>190</sup>The quote of Clapham is from page 194. Clapham 1899, pp. 167, 194–5.

<sup>191</sup>Clapham 1899, pp. 194–5.

<sup>192</sup>April 5. Clapham 1899, p. 195.

<sup>193</sup>Clapham 1899, pp. 197–8.

justified the declaration by reference to Austria's violation of French rights. Austria had given military protection to the French rebels harboring in Germany, supported the German princes in the Alsace dispute, tried to collude with the minority in France that wanted a counter-revolution, and organized a great power coalition to violate France's domestic political independence.<sup>194</sup> Notably, the French assembly acknowledged that Austria's legal case for the rights of Germany and Belgium had some merit, but maintained that France's legal case had merit nonetheless (implication). Addressing concerns that France intended to violate the sovereignty and political independence of Germany and Belgium, the assembly's decree "declared that France was forced to take up arms in self-defense. She remained true to the principle that wars of conquest and wars against liberty were iniquitous. [But] in the coming contest [waged in self-defense] she was prepared to wage war with all possibly humanity, and to welcome all who might seek to escape from tyranny by uniting themselves to a free people."<sup>195</sup>

---

<sup>194</sup>Clapham 1899, pp. 198–200.

<sup>195</sup>Clapham 1899, p. 200.



## CHAPTER 22

### Case study: The outbreak of World War II

The war was caused by the economic depression in Germany and the growing belief among Germans that its severity was due to the actions of France, Britain and the United States. This belief provoked an international dispute over whether these countries had the right to devastate Germany's economy so deeply – and whether Germany had the right to conquer foreign markets to the east to restore its economy. British leaders even acknowledged that some of Germany's grievances were legitimate. But it was not clear how to reconcile their differences. Once the dispute became irreconcilable, war was the only option for either side.

#### 1. Speed of threat assessment

Germany assessed the threat of economic spillovers from the Western powers slowly and patiently – with freedom to delay its assessment until the severity of the threat became clear (**implication 1**). There were a number of reasons for the gradual pace. First, the depression hit Germany in stages. It only became clear over time that the depression was so bad that it required drastic solutions. The idea of conquering foreign markets to solve the problem did not arise early or urgently in the crisis. The promoters of a military solution could only gain popular support and political power once the depression became bad enough and lasted long enough to make that solution seem like a good idea – or least not such a bad idea. The gradual onset of the depression would have been irrelevant, however, if it had not been for a more fundamental factor. The nature of the spillover problem itself meant both that the pace of threat assessment *could* be slow and that it *had* to be slow. As explained earlier, Germany's optimal army size was increasing as the economic crisis got worse. So the promoters of a military solution saw Germany's power as rising over time. This trend alleviated the time pressure to assess the threat and promote foreign conquest as a way to

address it. At the same time, the gap between Germany's current army size and its optimal army size was also increasing as the economic crisis worsened. So the size of the optimal rearmament program was increasing as the crisis worsened. This lengthened the rearmament schedule and thus created pressures to delay in assessing and responding to the threat.

First consider the stages by which the German economy sank into depression. The hyperinflation of 1923 had rendered the economy particularly vulnerable to economic downturns. This event destroyed vast amounts of domestic savings. The only way that commercial banks could restore their lending capacity in its aftermath was to attract deposits from abroad. This shift in capital sourcing rendered the German economy acutely vulnerable to currency market fluctuations caused by flights of capital owned by foreigners. For downward pressure on the German mark had to be relieved to avoid depreciations that would fuel domestic inflation (to avoid a repeat of the 1923 episode). Yet the only way to relieve these pressures was for the central bank to raise domestic interest rates. This response simply raised the price of capital to German firms, however, inhibiting investment and slowing the economy.<sup>1</sup> This depressionary mechanism would be triggered by all subsequent decreases in demand for the mark. The primary international causes of falling demand for the mark were the actions of governments and market actors in France, Britain and the United States.

In late 1926 Germany reversed its policy of exempting foreign investors from the capital gains tax. Foreign capital stopped flowing in and the German stock market declined significantly in the first half of 1927, reflecting the decline in business profits. The exemption was restored and the capital inflow resumed for a year.<sup>2</sup> By 1928, however, rising prices on the New York stock market led to a large outflow of capital from Germany that damaged the economy in two ways. First, the increased cost of capital to German firms provoked labor disputes and a strike in the steel industry in late 1928. This, in turn, caused an economic downturn that reduced further the inflow of new foreign capital.<sup>3</sup> At the same time, the flow of capital to New York put downward pressure on the mark that required the central bank to raise domestic interest rates. This move induced a drop in domestic lending that slowed

---

<sup>1</sup>Boyce 1989: 66; Boyce 2003: 257.

<sup>2</sup>Kindleberger 1973: 43-44.

<sup>3</sup>Boyce 2003: 257.

the economy further. The economy was already in depression by the summer of 1929. The stock market crash in New York a few months later triggered 10% declines in German stock prices and wholesale prices. Both developments reduced business profits again. The US demand for German goods then began a steady drop – even before the US Congress passed the Smoot-Hawley tariff – further depressing business profits.<sup>4</sup>

Unemployment rose so high in Germany that increased government payments to unemployed workers ballooned the government budget deficit. Tensions between the ministers who gave priority to reducing the budget deficit and the ministers who gave priority to maintaining the unemployment payments caused the collapse of the governing coalition in March 1930. The new coalition's leader, Heinrich Brüning, adopted deflationary policies that exacerbated the depression severely. These policies included tax increases and spending cuts to balance the government budget and wage and price controls to restore the economy. The budget balancing measures were driven by the conviction that the government must control public sector demand and interest rates to minimize the risk of inflation to avoid a repeat of the 1923 episode. The wage and price controls were driven by the conviction that business profitability could be restored by helping businesses to control costs. Yet the economy only fell deeper into recession and unemployment only increased. This was due in part to the deflationary policies. It was exacerbated, however, by further declines in German exports caused by the erection of tariff barriers by France, Britain and the United States.<sup>5</sup>

Brüning might have seen that the deflationary approach was a mistake and reversed course. But he had a deeper reason for maintaining this approach. He believed that if Germany's depression remained bad enough for long enough, then France, Britain and the United States would become convinced that Germany could no longer afford to make its reparations payments that were still due from the settlement of World War I. And then Germany could negotiate a final end to these payments. Gindely, 1884; Hiden 1993: 81; Kindleberger 1973: 172-174; Kolb 2005: 121; Lee and Michalka 1987: 115-119, 121-122. Driven by this goal, his insistence on balancing the government budget provoked such strong opposition by the German parliament (the Reichstag) that he had to dissolve parliament and call new elections.

---

<sup>4</sup>Kindleberger 1973: 101-104, 110, 114, 123-124, 131.

<sup>5</sup>Boyce 2003: 257-258; Kindleberger 1973: 131-132.

In the elections of September, 1930, the Nazi party increased its share of the vote from 3% to 18% and its seats in the Reichstag from 12 to 107. It was now the second largest party in the Reichstag behind the Socialists. Gindely, 1884 Boyce 2003: 259; Kindleberger 1973: 132; Kolb 2005: 113.

In the spring of 1931, a run on Austria's banks led to a run on Germany's banks in which vast amounts of capital owned by foreigners fled the country. The German central bank tried to defend the mark on international currency markets to avoid further depreciations that would fuel inflation. The effort failed and the bank lost millions in gold reserves in the process. This failure was fueled by the Brüning government's decision in the midst of the crisis to cut spending and raise taxes again. These measures immediately provoked a rebellion by the Reichstag that further sapped international confidence in the German economy just when it needed bolstering. The central banks of France, Britain and the United States loaned Germany's central bank \$100 million to help it defend the mark. The funds were exhausted in less than two weeks. Although France agreed with Germany that a new loan should be raised, France stipulated political conditions that Germany could not accept. These included cuts in military spending, prohibitions on paramilitary rallies and the renunciation of German plans to form a customs union with Austria.<sup>6</sup> The idea of providing a new loan was then vetoed by Britain's central bank governor and the United States' president. In any case, the damage to the German economy was already done. The lending capacity of German banks fell significantly because of the loss of foreign capital. The problem was exacerbated by the deflationary policy of Germany's central bank: It could not lower the interest rate it charged to private banks by too much since this would accelerate the flight of capital, the currency depreciations and the domestic inflation rate.<sup>7</sup> Brüning continued the deflationary policies through 1931.<sup>8</sup> Unemployment rose over the course of the year from 4,000,000 to 5,000,000 workers.<sup>9</sup>

In June, 1931 – in the heat of the German banking crisis – the United States proposed

---

<sup>6</sup>Boyce 1989: 71.

<sup>7</sup>Kindleberger 1973: 144-153.

<sup>8</sup>Kindleberger 1973: 169, 174.

<sup>9</sup>Boyce 2003: 258.

a temporary moratorium on Germany's war reparation payments. France objected strongly because she had lost more than the US in World War I and was owed the majority of the payments. News then slipped out that Germany intended to cease the reparation payments entirely (although she would wait five months to do so as long as all previous loans to Germany continued to be rolled over). France responded to this news by deploying troops to its eastern border with Germany.<sup>10</sup> This development touched German leaders in a sensitive area, because the size of Germany's army had been severely limited by the Versailles treaty imposed a decade earlier.

## **2. Goal of entry into military action**

Long before the depression had begun, Germany had planned to escape the Versailles restrictions and reassert Germany's previous international rights and interests. After the settlement of World War I in 1919, Germany's neighbors to the east and west had much larger armies and routinely use their military preponderance to assert their interests at Germany's expense. The leadership of the German army therefore saw rearmament as essential to defend the country's interests and restore its status as a great power.<sup>11</sup> Gustav Stresemann, the Chancellor and Foreign Minister in the mid-1920s, prioritized the goal of reclaiming the territory lost to Poland at Versailles.<sup>12</sup> Hitler too envisioned a path of military expansion for Germany as early as the 1920s. His plans were much more expansive than those of the other conservatives, however. He thought Germany should dominate Eastern Europe militarily, carve up and colonize Russia, and then defeat the Western powers and establish world dominance.<sup>13</sup> Before the depression the demand for Hitler's type of foreign policy remained limited to a narrow fringe of the political spectrum.<sup>14</sup> With the onset of the depression, however, Germany gained new and powerful reasons to rearm and assert its power beyond its borders – if not as far as Hitler ultimately wanted, then farther than anyone else had

---

<sup>10</sup>Boyce 1989: 71; Kindleberger 1973: 148-151; Rothwell 2001: 31.

<sup>11</sup>Hiden 1993: 51-52, 82; Lee and Michalka 1987: 58; Overy and Wheatcroft 1989: 96-98.

<sup>12</sup>Lee and Michalka 1987: 50, 76, 79; Overy and Wheatcroft 1989: 2-6.

<sup>13</sup>Evans 2008: 29-31, 39, 41, 45-46; Hiden 50-51; Rothwell 2001: 26.

<sup>14</sup>Boyce 2003: 258; Fritzsche 2008: 59; Kolb 2005: 101.

previously wanted. Other states seemed to be using a combination of economic coercion and military threats to resolve the international economic disputes arising from the depression in their favor. Why shouldn't Germany?

A broad range of conservatives in business, politics and the military felt that Germany should rearm and restore its economy and power by dominating the states to the east and southeast economically and politically.<sup>15</sup> Few of these conservatives shared Hitler's even more expansive goals. Yet they stood to benefit from the Nazi party's ability to attract voters, because this would restore popular support for a capitalist economy and a stable, conservative government. Anticipating this benefit, the leaders of big business tried to find out how conservative Hitler's economic policies would be<sup>16</sup> – at best to oppose him if his policies were not conservative enough, at worst to gauge the cost of his approach to stability and try to minimize it. Despite their differences, conservatives throughout business, politics and the military shared with the Nazis a common interest in restoring economic and political stability at home, rearming the German military, regaining the rights lost at Versailles, and establishing dominance over the states to the east and southeast.<sup>17</sup> Given this shift in the needs and aims of all German conservatives induced by the depression, it is not surprising that Germany's search for military solutions to the economic crisis was initiated by the Brüning government well before Hitler came to power.

Brüning's government took a number of steps in the direction of a military solution. His government shielded the army from budget cuts and expanded the production of battleships that were small enough to be in technical compliance with the Versailles treaty restrictions. His government extricated Germany from the reparations clauses of the Versailles treaty that had prevented Germany from rearming (because the reparations payments had sapped the tax revenues and borrowing capacity needed for rearmament). His government demanded

---

<sup>15</sup>Bell 2007, p. 160; Hiden 1993, p. 155; Lee and Michalka 1987, pp. 120–121, 123, 125–130, 137–141.

<sup>16</sup>Turner 1985: 129–142. Note that conservative economic policies were not an end in themselves for big business. They were only a means to the end of higher business profits. If the leaders of big business had known for certain that there was another kind of economic policy that could secure this end, they would not have been concerned about whether the Nazi party's economic policies were conservative enough. Later, when there proved to be another kind of economic policy that could secure this end, they *were* no longer concerned about it. Their conception of the policies required for business profits in 1930 was driven simply by conventional assumptions on the question.

<sup>17</sup>Hiden 1993: 49–50; Lee and Michalka 1987: 139–141.

revisions to the other clauses of the Versailles treaty that limited Germany's power and influence internationally. And his government explored ways to expand Germany's sphere of economic influence to the east and southeast.<sup>18</sup> Germany's economic and military tools in this expansion process were strategic complements, not substitutes.<sup>19</sup> Increasing the trade dependence of these countries would increase the marginal profitability of a given amount of military leverage over them. And increasing military leverage over them through rearmament would increase the marginal profitability of a given amount of trade dependence on their part. Similarly, the strategies of military coercion short of war (on one hand) and war itself (on the other) were also complements, not substitutes.

The Nazis, for their part, did not advertise to the public their plans for an aggressive foreign policy and a wide military expansion. They avoided all talk about their war plans in their public statements. They appealed to voters instead by promising to restore the German economy through job creation programs and to protect German society from the socialists and communists at home.<sup>20</sup> Their main promise to voters and elites alike was to provide domestic stability without having to make concessions to the labor unions or the left-wing political parties.<sup>21</sup> Yet Hitler was clear about his foreign policy plans when talking to the business elite.

As the depression deepened in 1931, leaders of the larger firms and the heavy industries – "big business" – became more frustrated with the intractability of the depression, its foreign structural sources, and the threats it posed to their positions in domestic politics and society. They were open to proposals for better solutions. In January, 1932, Hitler gave a speech to the leaders of big business in which he argued that it would be futile to renew trade and financial ties with France, England and United States. For this would only leave Germany vulnerable to a repetition of the current cycle of cut-throat price competition, tariff wars,

---

<sup>18</sup>Kindleberger 1973: 174; Lee and Michalka 1987: 128-132; Overy 1989: 98; Rothwell 2001: 30-31; Tooze 2006: 18.

<sup>19</sup>For an explanation of the difference between strategic complements and strategic substitutes, see Bulow, Klemperer and Geanakoplos (1983). Briefly, if increasing one policy variable *increases* the marginal profitability of a second policy variable, then the two dimensions of policy are strategic complements. Conversely, if increasing one policy variable *decreases* the marginal profitability of a second policy variable, then the two dimensions are strategic substitutes.

<sup>20</sup>Boyce 2003: 258-259.

<sup>21</sup>Fritzsche 2008, pp. 66-68; Rothwell 2001, p. 29.

currency pressures, credit crunches and depression. The only solution, in his view, was to cut trade ties with the Western powers permanently and pursue economic self-sufficiency by conquering foreign markets to the east and controlling them militarily.<sup>22</sup> Big business gave Hitler a mixed response.<sup>23</sup> On one hand, the business community shared with all of the right-wing parties – including the Nazi Party – a strong opposition to the current political equilibrium produced by parliamentary democracy under the Weimar Constitution. It required businesses to accommodate the demands of labor unions for higher wages and left-wing political parties for tighter regulation of business. The business community generally supported the right-wing parties because they favored free markets rather than negotiated concessions to labor unions and strong nationalist government rather than multi-party coalitions that included left-wing parties.<sup>24</sup> On the other hand, the larger firms and the heavy industries favored a return to international trade with France, Britain and the United States, whereas the Nazi party favored autarky and empire to the east.<sup>25</sup> The question for big business, therefore, was whether the other right-wing parties could deliver the popular support needed to maintain political stability in the face of rising unemployment, and if not, whether the Nazis could be utilized for that purpose alone.

Faced with this dilemma, some political and business leaders wanted to jettison the Weimar Constitution entirely and switch to an authoritarian regime even if at the risk of mass unrest. The army vetoed this option, however, on the grounds that it did not have the capacity to handle mass unrest or fight a civil war.<sup>26</sup> Thus the only way to transition to a more stable government was through constitutional means. The German President had the authority to dismiss the Chancellor and choose a new one who was capable of forming a cabinet that could sustain parliamentary support for his policies.<sup>27</sup> The president was therefore the focal point of pressures both to remove a Chancellor who had lost support and choose a particular candidate for the succession who might gather more support. If the

---

<sup>22</sup>Bell 2007: 159-160.

<sup>23</sup>Hayes 1991.

<sup>24</sup>Kolb 2005: 115, 220; Tooze 2006: 25, 103.

<sup>25</sup>Tooze 2006, pp. 29, 103, 105; H. A. Turner 1985, p. 282.

<sup>26</sup>Kolb 2005: 130.

<sup>27</sup>Kolb 2005: 126.



president could not be convinced to remove the current Chancellor, then he would stay. If the president could not be convinced to choose a particular candidate as the successor, then he would not be chosen. The man on whom President Hindenburg relied most heavily for advice in dismissing and choosing new chancellors was the leader of the army, General Kurt von Schleicher. His advice two years earlier had led to the choice of Heinrich Brüning.<sup>28</sup>

By 1932, then, Germany faced a unique combination of economic crises, political constraints and leadership opportunities. Unemployment was now above 6 million workers.<sup>29</sup> The voting masses were torn between the calls of the socialist and communist parties for redistributive policies, the calls of the liberal bourgeois parties for a return to the rising incomes of the mid-1920s, and the calls of the conservative nationalist parties for an entirely new political and economic order. "The social cement in the country crumbled. The Nazi party continued to gain."<sup>30</sup> Schleicher had the President's ear but lacked an effective strategy or popular support. He therefore reached an agreement with Hitler according to which President Hindenburg would dismiss Chancellor Brüning, choose a right-wing nationalist politician to replace him, dissolve the Reichstag and call new elections. In turn, Hitler agreed not to oppose the new nationalist government.<sup>31</sup> Brüning was duly dismissed and replaced by Franz von Papen, a right-wing member of the Center Party. This solution might have worked had the unemployment rate not been so high and the voters not been so drawn to the Nazi party's platform of job creation programs and vigorous action to restore social stability in the face of the communist and socialist threat.

In the Reichstag elections of July, 1932, the Nazi party increased its share of the vote from 18% to 37% and its seats from 107 to 230. It was now the largest party in the Reichstag by a large margin – the Socialist Party coming a distant second with 133 seats garnered from 22% of the vote.<sup>32</sup> Schleicher invited Hitler to join the new cabinet or appoint other Nazi party members to join it, hoping to control the Nazis' influence on policy in this

---

<sup>28</sup>Kolb 2005: 118.

<sup>29</sup>Boyce 2003: 258.

<sup>30</sup>Kindleberger 1973: 174.

<sup>31</sup>Kolb 2005: 127.

<sup>32</sup>Kolb 2005: 225.

way. Hitler demanded instead that, in light of the election results, he should be appointed Chancellor himself and the government be reorganized under his control. Neither Schleicher nor President Hindenburg acceded to this demand. Hitler then turned the Nazi Party against the Papen government and passed a vote of no-confidence against it in the Reichstag in September. Hindenburg and Papen responded by dissolving the Reichstag and calling new elections for November. Hitler again demanded to be given the Chancellorship himself with the same powers of decree that the previous two chancellors had enjoyed. Hindenburg refused and said that Hitler could only have the position if he could form a parliamentary majority – a condition which Hitler rejected.<sup>33</sup>

Hitler's approach to the economic crisis may not have been the best approach from the business community's perspective. But as the other right wing parties appeared incapable of forming stable coalition governments, it began to appear that political instability in itself was the problem. Continued cycling between right-wing and center-left governments posed the threat that the next swing to the right would induce labor unions to call a general strike and the army to come out and suppress it forcefully.<sup>34</sup> The next swing to the left, however, could threaten businesses with such labor-friendly policies as to make capital investment a losing proposition.<sup>35</sup> Business profits required economic and political stability. If the price of stability was to support the right-wing party that enjoyed the widest popular support, then the price was worth paying even if that party advocated public works projects and overly aggressive foreign policies that big business opposed.<sup>36</sup> The best solution, indeed, was to support that party as far as necessary to restore political stability and popular support for a market economy, yet constrain the party from enacting its least attractive policy proposals.<sup>37</sup>

"By the late autumn of 1932 and early winter of 1932-33, a good many [business leaders] were abandoning their strong opposition to a Nazi-led government. Some were vainly trying to cultivate relationships with seemingly more reasonable Nazis

---

<sup>33</sup>Kolb 2005: 127-129.

<sup>34</sup>Patch 1998, p. 291.

<sup>35</sup>Turner 1985: 310-311.

<sup>36</sup>Turner 1985: 302.

<sup>37</sup>Hiden 1993: 50; Kolb 2005: 131-132; Patch 1998: 279, 285-291.

like Gregor Strasser.”<sup>38</sup>

As the depression worsened, this path appeared to be the only one open to big business.

Simultaneously, German diplomats were scoring foreign policy successes that seemed to demonstrate that a strategy of reasserting German power could bring tangible rewards. In July, 1932, they convinced France, Britain and the United States that Germany could no longer afford to make war reparations payments; and the governments of those countries effectively absolved her of them.<sup>39</sup> With this news the expected tax rate on German businesses fell significantly. In September, German diplomats threatened to withdraw completely from the great power disarmament conference in Geneva if the Western powers did not allow Germany to breach the armament limits imposed at Versailles and rearm to their level. The Western powers gave in to this demand in early December.<sup>40</sup> Thus the policy of national strength and assertiveness advocated by all of the conservative nationalist parties – including the Nazi Party – seemed to offer benefits as well as liabilities.

”By 1932-33, even those industrialists and entrepreneurs who were most committed to international trade and the world market were compelled to acknowledge that markets were being closed everywhere and to turn to a German sphere [in eastern and central Europe] which seemed marked out for her by history and geography.”<sup>41</sup>

The only question for the business community was how high the domestic political price of economic recovery through foreign conquest would be. The longer the domestic economic and political crisis continued unresolved, and the greater the threat of a breakdown in civil order, the higher was the price worth paying.

In early December, Hindenburg dismissed Papen and appointed Schleicher himself to be the chancellor. Schleicher in turn offered the vice chancellorship to Hitler’s first lieutenant in the Nazi party, Gregor Strasser. He wanted to take it, but Hitler remained insistent

---

<sup>38</sup>Kobrak and Schneider 2004: 147.

<sup>39</sup>Kindleberger 1973: 175-176.

<sup>40</sup>Hiden 1993: 83-84; Lee and Michalka 1987: 132.

<sup>41</sup>Bell 2007: 160.

that he be made Chancellor himself. This was the last effort by any of the key players to enroll the Nazis in the cabinet yet try to limit the Nazis' influence on policy by keeping Hitler out of the Chancellor's seat. Schleicher, now in that seat himself, attempted to draw working class support away from the Nazi party to his own government by advocating redistributive policies appearing similar to those of the left-wing parties – concessions to labor unions and the breakup of large, bankrupt estates for redistribution and resettlement by peasant smallholders.<sup>42</sup> These plans provoked strong opposition from the groups that were threatened by them. The large landholders lobbied President Hindenburg to dismiss Schleicher and appoint a new chancellor who could enlist Nazi Party members in his cabinet. The leaders of heavy industry were split: half continued to support the Schleicher government while the other half pressed for a return to the conservative economic policies of the previous chancellor, or failing that, a change of government back to Papen.<sup>43</sup> One newspaper owned by leaders of big business concluded that the political deadlock could only be resolved by suspending the Reichstag and the constitution itself.<sup>44</sup>

Schleicher resigned on January 28 and Hitler took office two days later as the new chancellor. His cabinet included two Nazi party members and nine other conservatives from the other right-wing parties – including Papen himself as the vice chancellor. The aim of the nine conservatives was to limit the Nazis' influence on policy. Over the next six years the conservatives would try to control Hitler's behavior to reap the benefits of his chancellorship without having to pay the costs. And they had some success. For they held many interests in common with the Nazis. "A continuity of interests and a similarity of goals [was] shared by the National Socialists and the representatives of almost every other political and economic group."<sup>45</sup> As one reviewer of the historical literature on this issue put it:

"The explanation for the regime's initial relative success lies perhaps in the fact that it could create an overarching consensus by mobilizing German people behind a program of national revival identified with Hitler's leadership. In this re-

---

<sup>42</sup>Kolb 2005: 131-132; Patch 1998: 287-288; Turner 1985: 305-309.

<sup>43</sup>Kolb 2005: 221; Turner 1985: 309-311.

<sup>44</sup>Turner 1985: 311.

<sup>45</sup>Lee and Michalka 1987: 141.

spect, too, Hitler's leadership performed an integrative function. Above all, the traditional German elites – the armed forces, the civil service, business, and the professions – all highly experienced and effective, perceived in National Socialism some aims and values which they shared, notably a strong German nationalism. Moreover, they were encouraged to pursue their own individual agendas, freed from the restraints hitherto imposed by constitutional procedures or democratic forces or by adherence to liberal values. For example, the army could rearm [and] business could 'be master in its own house'..."<sup>46</sup>

The foreign policy goals of the Nazis were not identical to those of the other conservatives. But the overlap was broad enough to sustain cooperation between the two groups for many years.<sup>47</sup> Non-Nazi diplomats retained the highest offices in the Foreign Ministry in the expectation that the conservative agenda they shared with the Nazis would continue to be pursued. The non-Nazi military leaders did the same with the same expectation.

For the first five years of the regime, moreover, the balance of power between the Nazis and other conservatives was not equal. The latter could set firm boundaries with Hitler when he crossed their red-lines – and win – because Hitler needed their support.

"Not the least of his political concerns was that, as an interloper, he found it at the outset impossible to do without the conservative elites, either in government administration, foreign policy or the economy."<sup>48</sup>

The army conditioned its continued loyalty to the Nazi regime on the disbandment of the Nazi paramilitary groups. And it was done.<sup>49</sup> The foreign policy establishment conditioned its loyalty on adherence to its goals. And it was done.<sup>50</sup> Big business conditioned its loyalty to the Nazi regime on economic recovery and restored business profits (or at least the Nazis thought so). And it was done.<sup>51</sup> From this perspective, it is only clear in hindsight which

---

<sup>46</sup>Noakes 2008: 93.

<sup>47</sup>Lee and Michalka 1987: 146.

<sup>48</sup>Lee and Michalka 1987: 140.

<sup>49</sup>Wachsmann 2008: 129-130..

<sup>50</sup>Hiden 1993: 87. Lee and Michalka 1987: 140-141.

<sup>51</sup>Hiden 1993: 58; Tooze 2006: 108-109.

side of each partnership was "collaborating" with the other to safeguard its own goals. At the time one might equally have thought that the core agenda-setters were the army and big business while the Nazis were the collaborators.<sup>52</sup> The domestic balance of power eventually shifted in favor of the Nazis, of course. By the late 1930s Hitler was taking actions that both groups would have liked to prevent if they could have. And they tried (which in itself is evidence that prior to this point, their goals overlapped with his). But in the early and mid-1930s, it was not obvious that things would turn out that way.

### **3. Capability and credibility: The rational roots of obfuscated threats by Germany and appeasement by Britain**

The depression also caused a significant drop in Britain's civilian economic productivity. In contrast to Germany, however, Britain had the opportunity to restore its civilian economy by erecting tariff barriers and trading exclusively with its own imperial markets. Because British leaders knew the civilian economy could be restored in this way, they expected the country's optimal army size to decrease again once the economy hit bottom rather than

---

<sup>52</sup>This is not to say that Hitler was merely a figurehead. He was that and more. It is simply to distinguish between the question of authority and subordination, on one hand, and the question of agenda setting and collaboration, on the other. The two questions are distinct. There have been plenty of regimes through history in which the authority holder was only a figurehead while the real agenda-setters were key interest groups apparently subordinated to the authority holder. The distribution of power that emerged in Nazi Germany was admittedly more complex than that. Yet the point remains that the core agendas of big business, the army and the traditional foreign policy establishment were well served by the Nazi regime. If it had departed from their agendas in its first 4 or 5 years in power, it would not have lasted. Hitler was constrained to pursue the goals that he shared with them. In this sense, he represented their interests and advanced their agendas.

Historians have noted how the Nazi regime served the core agendas of big business. See Hiden (1993: 57-58) and citations therein, Kobrak and Schneider (2004) and Tooze (2006). "After [Hitler came to power], business leaders adapted quickly enough. Big business began to view the Nazis more opportunistically as a possible means of ridding Germany of the vestiges of the near-universally hated Versailles Treaty, reducing international financial and competitive pressures and quelling labor unrest." (Kobrak and Schneider 2004: 147)

"[W]hat was clear was that legitimate authority in the Third Reich proceeded from the top down, ideally from the very top down. And what was also clear was that many leaders of German business thrived in this authoritarian atmosphere. In the sphere of their own firms they were now the undisputed leaders, empowered as such by the national labor law of 1934. . . . In material terms, the consequences of [political and social] demobilization [of the masses] made themselves felt in a shift in bargaining power in the workplace. . . . When the wage freeze in 1933 was combined with the destruction of the trade unions and the highly permissive attitude towards business cartelization, the outlook for profits was certainly very favorable. . . . And, perhaps most importantly, Hitler's regime promised to free German firms to manage their own internal affairs, releasing them from the oversight of independent trade unions. In future, it seemed, wages would be determined by the productivity objectives of employers, not the dictates of collective bargaining." (Tooze 2006: 102)

remain high. Hence Britain's optimal strategy was to appease Germany (**implication 4.1d**). There is much historical evidence that Britain appeased Germany in the 1930s. Historians have attributed this policy to many different factors but done little analysis to determine which factor was the root cause or how the various factors operated together. The model presented earlier implies that the root cause was the drop in Britain's optimal army size – in the sense that without this factor, the other factors alone would not have produced a policy of appeasement. After considering the other factors, the analysis shows how the optimal army size drove them to lengths they would not have gone to otherwise.

Even before the depression, British foreign policy makers focused on seeking diplomatic solutions and went to great lengths to secure them before resorting to military action. The priority of diplomacy was driven by two motives: a long-run decline in the perceived benefits of using military force and a general decline in the perceived costs of reaching diplomatic agreements. After World War I, the people were tired of war and there were no military threats on the horizon. Britain reduced its military spending significantly as a result.<sup>53</sup> At the same time, the advent of the League of Nations and its system of treaties and protocols produced a drop in the expected costs of resolving international disputes diplomatically.<sup>54</sup> More fundamentally, the British diplomatic establishment had an age-old tradition of presuming that international disputes could be resolved diplomatically if only one could find the right combination of concessions and threats to hold forth to the opponent.<sup>55</sup> At the least it was worth trying this approach before concluding that the opponent was implacable and war was necessary.

Britain's reductions in military spending in the 1920s meant that if a threat were ever to arise, it would take some time to ramp up the military capacity needed to meet it. This factor alone implied that an interim strategy of appeasement would be needed during the ramping up process. Yet the decision whether to begin ramping up – and the speed of ramping up – would always depend on two more fundamental factors: the clarity of the threat and the expected power balance after both sides had fully ramped up. If the threat were

---

<sup>53</sup>Murray 2003: 112; Overy and Wheatcroft 1989: 65, 68, 71.

<sup>54</sup>Overy and Wheatcroft 1989, p. 64.

<sup>55</sup>Boyce 2003: 262; Overy: 73-74.

clear or the expected power balance favored Britain, it would be rational to begin ramping up immediately and do it speedily. For the costs of war would clearly be worth paying. Conversely, if the threat were unclear or the expected power balance favored the opponent, it would be rational to delay the ramping up process to see if the threat might be defused diplomatically by material concessions that would be less costly than fighting such a strong opponent. This is exactly the situation that Britain faced in the 1930s. The decrease in its optimal army size – and the increase in Germany’s optimal army size – shifted the expected balance of power so dramatically that it was rational to delay the ramping up process and pursue appeasement first. The larger was the shift in the power balance, the greater would be the amount of concessions worth offering to avoid a war. Domestic constituents in Britain understood this calculus. This is why domestic political pressures delayed the tax increases that would have been needed to fund rearmament for so long in the 1930s. It took that long for the constituents to realize that Germany’s demands exceeded the maximum amount of concessions that were worth offering to avoid a war. Now consider how this interpretive framework organizes the historical evidence from this period.

The depression hit Britain the hardest between 1929 and 1932. In 1931, the run on the German mark led to a run on the pound that forced Britain off of the gold standard.<sup>56</sup> Other countries followed suit and international trade collapsed.<sup>57</sup> By 1932, British trade had fallen by 40% and the unemployment rate stood at 20%.<sup>58</sup> The voting public demanded economic recovery. And politicians on both the left and the right understood that if the demand were not met, the political system itself could collapse in social disorder.<sup>59</sup> Moderates in the Labor Party joined the Conservative Party in a coalition government pledged to make economic recovery the first priority. Military spending was reduced significantly to leave money in the civilian economy to restore consumption and investment as well as to allow government spending on social programs to pump-prime the economy. This order of priorities had always been advocated by the Labor Party. Even the conservatives favored these reductions in military spending, however, because they knew that if they failed to restore the economy

---

<sup>56</sup>Kindleberger 1973: 154-164.

<sup>57</sup>Boyce 2003: 261; Kindleberger 1973: 170.

<sup>58</sup>Overy: 69.

<sup>59</sup>Boyce 2003: 261-262; Overy: 68.



then they would be outvoted at the polls and replaced by a left-wing majority government that might dismantle the more fundamental foundations of British capitalism, or worse, trigger political unrest.<sup>60</sup> This domestic constraint on military spending operated at least until 1935 and in some sections of public and elite opinion even later.<sup>61</sup>

Historians usually assume that this domestic constraint *interfered* with Britain's ability to pursue an optimal foreign policy, and that if this constraint had not been there, then Britain would have rearmed sooner, faster and farther. In fact, however, the voter constraint *accurately reflected* Britain's international position and the optimal foreign policy response to it. Voters knew that the economy could be restored by leaving money in the civilian economy rather than raising taxes to fund military rearmament, because Britain's imperial markets offered an opportunity to restore trade quickly even without multilateral cooperation from the other great powers. Voters knew their incomes would be higher with this approach than with any other – especially a policy of war. Political leaders simply heeded this voter-income constraint when they erected tariffs barriers to protect Britain's imperial markets in 1932. They also heeded this constraint in keeping military spending low and looking for diplomatic solutions to the rise of German fascism and militarism. The only question was whether the German government could be conciliated into remaining peaceful by a set of concessions that would be cheaper than going to war (against Germany).

Because the depression had increased Germany's optimal army size while decreasing Britain's, the expected power balance in such a war was strongly in Germany's favor. So even considerable concessions would be cheaper than having to fight a war. More fundamentally, many British leaders felt that Germany's grievances were legitimate up to a point.<sup>62</sup> They felt that France was to blame for Germany's grievances on two counts. First, France had imposed an overly vindictive peace settlement on Germany at Versailles. So it was legitimate for Germany to demand revisions to the treaty. Second, at the height of the Depression between 1929 and 1933, France had been exceedingly demanding in its terms for economic cooperation with Germany to mitigate the depression. British leaders thus had two reasons

---

<sup>60</sup>Overy: 69.

<sup>61</sup>Overy: 76, 78.

<sup>62</sup>Overy: 74-75.

(the power imbalance with Germany and the legitimacy factor) to believe that by offering concessions to Germany, they could attain a mutually satisfactory distribution of rights between the two countries and maintain the peace. There were limits to the extent of concessions that Britain could make and still get off more cheaply than fighting a war, of course. Yet the limits were broad as long as Hitler's government proved to be satiable in the end. And as long as Hitler pursued a policy of obfuscating the threat that Germany posed, he seemed to be satiable. Until as late as 1937, Hitler himself actually believed that an accommodation could be reached with Britain. His belief was driven in part by the sense that Britain was on the weak end of the power balance between the two countries *and knew it*.<sup>63</sup> Thus from his end, too, it seemed that accepting Britain's offer of concessions and giving at least something in return could produce a mutually satisfactory solution short of war.

This set of mutual expectations explains Britain's actions to appease Germany. It also explains why Britain shifted from appeasement at a low level of armament in the early to mid-1930s to appeasement at a high level of armament in the late 1930s. In the first period appeasement was driven by Britain's expectation that a mutually satisfactory settlement could be reached short of war. In the second period Hitler's credibility started to evaporate, however. So now appeasement was driven by the need to buy time for ramping up to the optimal army size and negotiating an alliance with France.

By 1934 British leaders knew that Hitler's regime was firmly in power. It was clear that Hitler sought revisions to the Versailles Treaty and an expansion of German power for some broader goal. A British defense committee concluded that Germany was potentially the biggest threat to British security in the long run.<sup>64</sup> Yet, in the near term, the threat was not seen as big enough to warrant large-scale rearmament or a diversion of existing military assets from imperial defense.<sup>65</sup> It was still not clear what Hitler's ultimate goals were or the degree to which he was constrained by other regime elements in pursuing them. There was some evidence to suggest that the constraints on him were real. Hence rearmament was only initiated on a small scale in 1934. The army's requirements for a ground war against

---

<sup>63</sup>Hiden 1993, pp. 95–97; Leitz 2004, p. 53.

<sup>64</sup>Overy: 72.

<sup>65</sup>Overy: 74–75.

Germany were given third priority behind the navy's requirements for a war against Japan and the air force's requirements for punitive air strikes against Germany.<sup>66</sup>

Even after Germany reinstated army conscription and warplane construction in 1935 – both in violation of the Versailles Treaty – Britain continued to pursue agreements in search of a peaceful solution to the German problem. Britain proposed to restore some colonies that Germany had lost in World War I and signed a naval arms limitation treaty in the spring of 1935 (see above). When Germany remilitarized the Rhineland in 1936 – again in violation of Versailles – the uncertainty about Hitler's true intentions rose. Yet many in the British government still saw reasons why the move might be justified. First, Germany could not be kept from occupying its own territory indefinitely.<sup>67</sup> Second, the demilitarization of the Rhineland in the Locarno Treaty was part and parcel of France's overbearing settlement at Versailles. Although the German move into the Rhineland inspired more caution on Britain's part, it did not destroy all trust that Germany was satiable.

The domestic voter-income constraint still constrained British rearmament in 1936 and 1937. The British economy had recovered significantly by mid-decade. Treasury officials were emphasizing, however, that continued recovery depended on maintaining the policy of low taxes. Members of parliament on both the left and the right knew that a major rearmament drive would require large tax hikes, cuts in social programs and increases in government control of the economy – and these measures would slow the economy.<sup>68</sup> They also knew that their constituents would revolt at the polls if the economy collapsed again. The chancellor of the exchequer, Neville Chamberlain, did his best to increase military spending within this constraint. A defense coordination agency was created in 1936 and quickly concluded that the nation should rearm by enough within three years to deter Germany from starting a war, and if deterrence failed, to stave off defeat in a war.<sup>69</sup> In 1937 Chamberlain entered the prime minister's office with a three-fold mandate. He was to reach a diplomatic Grand Settlement with Germany to insure the peace while simultaneously increasing the rearmament rate by

---

<sup>66</sup>Overy and Wheatcroft 1989, p. 75; Rothwell 2001, p. 58.

<sup>67</sup>Overy: 72.

<sup>68</sup>Overy: 78.

<sup>69</sup>Overy: 81-82.

enough to create the bargaining leverage needed to reach such a settlement, yet without letting the rate get so high as to damage the civilian economy!<sup>70</sup> These complex trade-offs drove British policy for the next two years.

The key question facing everyone in Britain – voters and leaders alike – was when to pull out all the stops on the rearmament process. At what point would it be clear that the Germans were truly insatiable – even at the highest price worth paying? It is tempting to claim that the British should have known this about Hitler long before 1937. Yet even *Hitler's own generals* did not learn that he reserved the right to override their preferences until 1938 when he fired the top generals and installed others who shared his own preferences.<sup>71</sup> If the Germans themselves were unclear who would control policy in the final analysis, how could the British be any more clear? In this environment the optimal British strategy was to continue screening Germany for its satiability by continuing to offer concessions that would be cheaper than a war until it became clear that the concessions needed to satiate Germany would be costlier than a war. Only then would it be rational to pull out all of the stops on rearmament. If this actually were the strategic dilemma that Britain faced, then one would expect to observe British policy moving *simultaneously* in three directions: (1) making offers to Germany, (2) estimating the likelihood that Germany's true aims were irreconcilable with British security, and (3) looking for ways to reduce this likelihood by intervening in regime politics in Germany. The offers would screen Germany for its satiability while the estimates of its aims would determine whether to make the next offer – and the interventions would raise the chances that the next offer would be accepted. This is in fact the strategy that Britain pursued. It is only in hindsight that the strategy seems too optimistic.

---

<sup>70</sup>Overly: 79-82.

<sup>71</sup>Hidden 1993: 102.

Part VI

**The path of power unleashed: Grand  
strategy in hegemonic wars**

## CHAPTER 23

### Theories of grand strategy in hegemonic wars: A review and critique

Grand strategy is the art and science of waging war in a way that will produce an acceptable peace at an affordable cost. In each hegemonic war, the statesmen on each side expected their grand strategy to produce an acceptable peace at an affordable cost. Only one side was right. This fact presents a puzzle. How could the statesmen on the other side have been so wrong in their expectation of achieving success at an affordable cost? Why did their domestic constituents allow them to continue their grand strategy after its failure became clear and its cost continued to grow astronomically?

The conventional answers to this question are of two types. The first type of answer is to assume that the statesmen and their constituents on the losing side had "revisionist preferences". (The conventional theories that adopt this assumption are called offensive realism and defensive realism.) The second type of answer is to assume that rational cost-saving decisions were confounded by relative power factors or domestic political factors. (The conventional theories that adopt these assumptions are called structural realism and neoclassical realism.) Each type of answer is unsatisfactory, however, for reasons I will explain in this chapter.

Contractual Realism offers a different answer. Each hegemonic war was caused by a legal incompatibility dispute. Once this dispute became intractable, each side realized that it could not "split the difference" in the dispute, because that would allow the other side to receive a net gain from the dispute – and then it would have an incentive to create more disputes like it in the future. So each side decided that it needed to impose costs on the other side until the other side had suffered more losses than it expected to gain from a final settlement of splitting the difference. This calculus required each side to continue waging

war until the other side had suffered that much losses, despite the astronomical cost of waging war for that long. This answer explains more of the facts of hegemonic war more parsimoniously than the conventional answers explain.

In the study of grand strategy, the leading explanatory paradigm is Conventional Realism. This paradigm is composed of different theories whose causal assumptions contradict each other (e.g. the four theories mentioned above). The conventional wisdom is that there is no need to develop a research agenda that would resolve the contradictions between these theories and no need to develop a new theory that would outperform these theories in explanatory power. According to this view, it is valuable and even scientifically progressive to have different theories that explain different state behaviors. It would be scientifically counterproductive to try to develop a single theory that would explain such variations in state behavior as a consequence of variation in one causal factor. As a leading contributor to the conventional realist paradigm wrote:

”No single subschool or theory [within conventional realism] is always right or always the source of the master explanation to which others are subservient. Different strands of [conventional] realism are more or less relevant to different problems and cases. The question for contemporary researchers is which subschools or specific theories to apply to a given problem or case?”<sup>1</sup>

I argue that this approach is scientifically regressive for several reasons. First, it prevents theories that are refuted by empirical evidence from being discarded. Second, it prevents critical tests between existing theories and new theories that explain more of the historical facts more parsimiously than the existing theories. Third, it prevents progress from a pre-scientific field composed of disparate theories with contradictory assumptions and no method for reconciling them to a scientific field where the contradictions have been resolved by a more general theory that explains a variety of state behaviors based on a few mutually consistent assumptions.

In this chapter I will summarize the Conventional Realist theories of grand strategy, identify their logical contradictions and empirical failings, and outline how Contractual Realism

---

<sup>1</sup>Wohlforth 2008, p. 143.

resolve those contradictions and outperforms those theories in explanatory power.

## 1. "Challengers" and "balancers": Debunking the conventional wisdom

According to the conventional wisdom, the best way to explain grand strategies during hegemonic wars is to assume that in each war there was a "Challenger" and a coalition of "Balancers" that opposed it. The challenger sought to revise territorial boundaries by military force and the balancers sought to restore the original territorial boundaries. This approach has given rise to two distinct theories of grand strategy – one theory to explain the behavior of the challengers and another theory to explain the behavior of the balancers. The main difference between the two theories is in their assumptions about the ultimate goal that states pursue. *Offensive realism* assumes that states pursue power.<sup>2</sup> *Defensive realism* assumes that states pursue security.<sup>3</sup> This difference in assumptions seems natural, because offensive realism is intended to explain the behavior of the "challengers", while defensive realism is intended to explain the behavior of the "balancers".

### 1.1 Logical problems with offensive realism and defensive realism

Each of these theories locates the cause of state behavior in the utility function of the state. According to offensive realism, the challenger pursues power because it has an innate "taste" for power in its utility function. According to defensive realism, the balancer state pursues a security because it has an innate "taste" for security in its utility function. However, it is fallacious to locate the cause of an actor's behavior in its utility function, for a number of reasons.

First, such a theory is unrefutable. Utility functions are unobservable, so there is no way to verify empirically whether the causal variable is present or absent in a particular historical case. Second, such a theory cannot explain changes in behavior. For an actor's

---

<sup>2</sup>Representative works in the offensive realist school include Labs 1997; Zakaria 1998; Mearsheimer 2001; Elman 2004.

<sup>3</sup>Representative works in the defensive realist school include S. M. C. Walt 1989; Posen 1993; Glaser 1994; Kydd 1997; Van Evera 1999.



utility function does not change (by definition). So the theory cannot explain why the actor engages in the behavior at some times but not at other times. For example, a theory which assumes that states have a "taste" for war cannot explain why a particular war broke out at a particular time rather than a year earlier or a decade earlier.

Third, such a theory is tautological. It *assumes* that an actor behaves a certain way because it gains utility from that behavior. Using this approach, the theorist can explain any behavior simply by assuming that the behavior is embedded in the actor's utility function. This kind of theory can only explain changes in the actor's behavior by assuming that the old behavior dropped out of the utility function and the new behavior fell into the utility function. This way of explaining behavioral change simply begs the question, however. *Why* did one behavior drop out of the utility function and another behavior fall into the utility function?

Fourth, this approach produces unresolvable debates. When two competing theories attribute state behavior to two different utility functions that supposedly motivate statesmen, the two theories cannot be tested against each other. For again, the utility functions are unobservable. For example, if both theories purport to explain the same historical event, then there is no way to *observe in the historical evidence* which of the two utility functions motivated the historical actors involved. So there is no way to test one theory against the other.

Fifth, this approach prevents generalization in theory building. When each behavior is given a different theory to explain it, the result is a growing set of specific theories but no progress toward the development and testing of a more general theory – a single theory that can explain a range of behaviors by identifying the observable conditions that cause actors to behave in one way or another and change from one behavior to the other.

## **1.2 Empirical problems with offensive and defensive realism**

Some theorists try to solve these logical problems by attributing a state's taste for security or power to its regime type. Regime types are observable, so such a theory might be tested. And regime types change, so such a theory might explain the switch from security-seeking to power-seeking behavior. The historical evidence on hegemonic wars refutes such a theory,

however. Of the seven cases of hegemonic war between 1550 and 1950, there was a change in regime type in only two cases. In one of the two cases, the regime type changed from absolutism to democracy and the new regime apparently switched from security-seeking to power-seeking. Some theorists claim that new democracies are more prone to aggression than other regime types – and cite this case as a prime example.<sup>4</sup> The most recent statistical study shows they are not,<sup>5</sup> however, and therefore it casts doubt on this explanation of this case. In the other case, the regime type changed from democracy to authoritarianism and the new regime apparently switched from security-seeking to power-seeking. This switch seems consistent with the conventional hypothesis that non-democratic regimes are more prone to power-seeking behavior.

Yet this still leaves the switch from security-seeking to power-seeking unexplained in the other five cases where there was no change in regime type during the international dispute or the ensuing war. Some theorists try to explain these cases by assuming that absolutist regimes were more prone to bouts of power-seeking than constitutional monarchies or democratic states. Yet this hypothesis is also refuted by the historical evidence. In two of the five cases, both of the opposing states in the international dispute and the ensuing war were absolutist regimes. Yet only one of the two opposing regimes switched from security-seeking to power-seeking. This evidence is clearly an anomaly for a theory that attributes the behavioral switch to the absolutist regime type.

### **1.3 State power and its role in grand strategies: A Contractual Realist approach**

State power is composed of a number of elements. State power changes in response to a number of causal factors. When state power changes, it has a number of consequences. Offensive and defensive realism conflate these aspects of state power in a way that oversimplifies the realities of state power and obscures the role of state power in grand strategies. Other variants of conventional realist theory make the same mistake (as we will see below). To provide a better explanation of grand strategies during hegemonic wars, it is necessary to disentangle the key aspects of state power and demonstrate the causal relationships between

---

<sup>4</sup>Snyder..

<sup>5</sup>CT..

them.

Let us adopt the simplest possible definition of state power: The number of men that a state puts into military use. The more men the state puts into military use, the larger its army size will be, and the greater will be its power in international relations. Thus power is defined as military manpower, and a state's power is measured by its army size. This definition of state power provides the basis for a simple theory of changes in state power. Each state puts men into military use up to the point that the marginal benefit from doing so just equals the marginal cost of doing so. By implication, when a state's marginal *benefit* from putting men into military use increases, the state *increases* its army size. Conversely, when a state's marginal *cost* of putting men into military use increases, the state *decreases* its army size.

In order to generate observable implications from this theory, it is necessary to assume that each state's ultimate goal is simply to maximize its income. That is to say, each state's utility function contains one element: income. All states have the same utility function. This utility function never changes. Thus different sources of income have the same relative value at all times (e.g. relative to each other). This rules out a situation where some state values security more than power this year – when it is a "security seeker" – but then values power more than security next year – after it becomes a "power seeker".

Under these assumptions, a state's army size will vary over time and space in response to shifts in the marginal benefit and marginal cost curves for putting men into military use. Changes in *material* costs and benefits have the greatest impact on national income. So they are the most important drivers of army size and all state strategies that depend on army size.

From this perspective there are three main causes of change in a state's power. By distinguishing between these three causes we can disentangle the elements of state power and resolve misunderstandings about the role of state power in grand strategies.

Cause #1. Long run changes in population size, territory, technology and institutions that shift the marginal benefit or cost of putting men into military use. For example, when a state's population grows over the long run, this reduces the marginal cost of putting men into military use and thereby induces an increase in the state's army size.

Cause #2. Short run changes in the marginal benefit or cost of putting men into military use. For example, if one state suffers an economic depression and it spills over into another state, this lowers the second state's marginal cost of putting men into military use and thereby induces an increase in its army size. Conversely, if one state experiences a domestic reform movement and it spills over into another state and triggers a secession crisis there, this lowers the second state's marginal benefit from putting men into military use and thereby induces a decrease in its army size.

Cause #3. An international spillover problem that causes a dispute over income between two states. If the dispute becomes irreconcilable, then *each* of the opposing states in the dispute gains an incentive to put men into military use simply to secure its rights in the dispute (as shown in Part I of this study). This induces each state to increase its army size from the peacetime level to the wartime level. The wartime level may be high or low depending on the impact of the spillover problem on the state's marginal labor productivity (e.g. Cause #2). Yet in either case the wartime level will be higher than the peacetime level.

Conventional Realism is correct to assume that power flows from the material capacity to wield force. The greater is a state's material capacity to wield force, the greater is its power. However, in defining power by state characteristics that do not change in the short term, conventional realists ignore the most important material factors that enable and motivate states to use force (Causes #2 and #3). As a result they fail to see that these factors are a source of power. For if a state is materially motivated to use force, this gives it power ipso facto – regardless of what the motivation is. By excluding a whole range of material motivations, conventional realists are left with no choice but to assume that states' underlying motivations are located in the "utility functions" of statesmen. That is to say, they assume that statesmen are motivated to use force by their "tastes" for power, influence, status, glory or what have you.

#### **1.4 A Contractual Realist theory of security-seeking and power-seeking**

When an international spillover problem causes an irreconcilable dispute over income between two states, each state gains an incentive to transfer civilian workers into the military sector

to secure its income rights (Cause #3). Each state's incentive to put men into military use rises by enough to turn it from a "security seeker" into a "power seeker". Thus the shift from security seeker to power seeker does not require two competing theories – offensive and defensive realism – based on two different assumptions about a state's ultimate goal (power or security). The shift is attributable to the emergence of an irreconcilable dispute over income in a world where each state's ultimate goal is simply to maximize its income.

Once the two states are at war, each state faces a further question about how it will use its military manpower in the war. Will it use its military manpower *only* to combat the other state on the battlefield? Or will it *also* use its military manpower to conquer some third state and establish territorial hegemony over that state in order to increase its *territorial power*? This decision is determined by the impact of the international spillover problem on each state's labor productivity (Cause #2 above).

If the spillover problem *decreases* the state's military labor productivity relative to its civilian labor productivity, then it has an incentive to reduce the number of men it puts into military use (e.g. its military manpower). Hence it will lose the war unless it conquers other states and establishes territorial hegemony over them in order to increase its territorial power. Conversely, if the spillover problem *increases* the state's military labor productivity relative to its civilian labor productivity, then it has an incentive to increase the number of men it puts in to military use. So it does not need to conquer other states to avoid losing the war.

In the first case, the state seeking hegemony over others will appear to have the ultimate goal of *territorial power-seeking*. Yet the goal of territorial expansion is simply an induced preference – induced by the the spillover problem which reduced its military labor productivity. The territorial expansion is not driven by an underlying preference or taste for territorial power (as offensive realism would claim). In the second case, the state with no need to seek hegemony will appear to have the ultimate goal of *territorial security-seeking* – in so far as it simply wants to overturn the first state's hegemony and return to the pre-war territorial boundaries (once the spillover problem is resolved). Yet here, too, the goal of territorial security is simply an induced preference – induced by the spillover problem which increased its

military labor productivity. The security agenda is not driven by an underlying preference or taste for security (as defensive realism would claim). Thus each state's decision whether to seek power through territorial expansion or to seek security within existing territorial boundaries is motivated by the impact of the spillover problem on its labor productivity.

Historically, each type of spillover problem had a *different* impact on labor productivity in the two opposing states. For example compare the impact of secession spillovers and revolution spillovers.

1. Secession wars. The spillover problem decreased the productivity of military labor relative to civilian labor in *the state suffering income losses from the spillovers*. So it was this state that needed to conquer other states and establish territorial hegemony over them to avoid losing the war. Examples include: Austria in the Thirty Years War, Spain in the War of the Spanish Succession, and Austria in World War I.)
2. Revolution wars. The spillover problem decreased the productivity of military labor relative to civilian labor in *the state producing the spillovers*. So it was this state that needed to conquer other states and establish territorial hegemony over them to avoid losing the war. The main example is France in the French Revolutionary wars.

### **1.5 Resolving debates between variants of conventional realist theory**

From this perspective, the root cause of a state's grand strategy is not a taste for power or a taste for security embedded in its utility function. Each state's grand strategy is driven by a deeper cause: the international spillover problem and the dispute over income that it provokes. This problem drives each state to oppose the other state by military force simply to secure its income rights. This problem also drives each state to choose whether to pursue territorial expansion and hegemony or counter-hegemonic balancing and security within the pre-war territorial boundaries. Each state's choice depends on whether the spillover problem lowers its own military labor productivity or that of its opponent. Thus factor thus explains regime intentions (e.g. whether to pursue power or security). By explaining stability and change in regime intentions, Contractual Realism resolves the debate between offensive realism and defensive realism. Contractual Realism also explains shifts in territorial power between one state and another during a hegemonic war – first in

favor of the state that is pursuing hegemonic expansion and then in favor of the counterbalancing coalition of states. By explaining both changes in regime intentions and changes in the international distribution of power, Contractual Realism also resolves another debate between two other variants of conventional realist theory (as we will see in the next section).

The advantage of the Contractual Realist approach is that it explains state behavior as a consequence of changes in the material costs and benefits encountered by statesmen in the world around them rather than by differences in statesmen's unobservable utility functions. In the Contractual Realist approach, the fulcrum of explanatory power is the rationality of statesmen. They are assumed to respond to changes in the material world around them in a rationally consistent manner. The theory shows how statesmen respond to specific changes in *their* observable world, and thereby identifies causes and effects that are clearly observable to *us* in the historical record. Both the variation in material conditions caused by the spillover problem and the logical implications for changes in state behavior are directly observable in the historical record. Hence the lines of causation from shifts in the causal variables to shifts in the effect variables are directly observable from beginning to end. This makes it possible for scholars to test the theory against other theories and resolve the debates between the theories.

## **2. "International system structure" and "domestic politics" as independent causal factors: Debunking the conventional wisdom**

As we just saw, Contractual Realism explains both changes in the international distribution of power and changes in state intentions. It thereby resolves the debate between two other variants of conventional realist theory: structural realism and neo-classical realism. In this section, I summarize the claims of each theory, explain how it misunderstands the role of state power in grand strategy, and show how Contractual Realism resolves the misunderstanding.

### **2.1 Structural realism**

This theory assumes that the strongest driving force in international relations is the international distribution of power, or as some theorists call it, "the structure of the international

system". According to this theory, a state's strategy is constrained more tightly by the international distribution of power than by any other factor. When the international distribution of power changes, each state is driven to change its strategy accordingly. In particular, when the international distribution of power shifts in a state's favor, it is driven to pursue "preponderance" over other states. Conversely, when the international distribution of power shifts against a state, it is driven to form a "balancing" coalition in order to oppose the state that is seeking a "preponderance of power". To use the standard language, "states balance against power".<sup>6</sup> According to structural realism, state strategies are driven more forcefully by shifts in the international distribution of power than by any other factor.<sup>7</sup>

From this perspective, the history of hegemonic wars in Europe is a simple story of cause and effect. In each case the cause was a shift in the international distribution of power in favor of one state. The ensuing effects were inevitable. That state sought a preponderance of power over the other states. They formed a counter-balancing coalition in response to oppose it and put it back in its place.

This theory suffers from several logical and empirical problems. First of all, it focuses only on the long-run determinants of state power and the international distribution of power: each state's population size, territorial holdings, technology and institutional efficiency (e.g. Cause #1 above). By reifying the long-run determinants of state power as the only determinants that matter from a cause-and-effect perspective, the theory is blind to the ways in which state power is increased or decreased by other factors that impose even tighter constraints on a state's choice of army size and strategies that depend on army size (e.g. Causes #2 and #3 above). So the theory is blind to the ways that these other factors shift the international distribution of power by even wider margins than the long-run factors do.

In each hegemonic war, the international dispute over income drove both of the opposing states to increase their army sizes from peacetime levels to wartime levels – and keep them there for longer than any other factor in modern European history. Thus these income disputes (Cause #3) were a more fundamental determinant of state power than any other

---

<sup>6</sup>For reviews of the long history of thinking about the balance of power, see M. Sheehan 1996; Haslam 2002.

<sup>7</sup>Representative works in the structural realist school include Layne 1993; Layne 2006; Mearsheimer 2001; Waltz 1979; Waltz 2000.



factor. During each war, each state's choice of army size was constrained tightly by the relative productivity of its military and civilian labor. This factor constrained each state's choice more tightly than the long-run determinants of army size did. In turn, each state's choice of grand strategies was constrained more tightly by its army size than by any other factor. As we will see, each state's grand strategies were constrained not only in the choice whether to pursue hegemony or to balance against the hegemon, but also in a number of other respects. Thus military labor productivity (Cause #2) was a more fundamental determinant of each state's power each state's grand strategies, and the international distribution of power than the long-run determinants were. By focusing only on the long-run determinants, so-called "structural" realism is blind to the real causes of changes in state power, grand strategy and shifts in the international distribution power.

Contractual Realism identifies these root causes and demonstrates how they shaped several aspects of grand strategy that shifted the international distribution of power shifted during each war. Briefly, each state was driven to war simply by the need to secure its income rights in the dispute over the international spillover problem. Once it was at war, each state needed to increase its military power simply to avoid being forced out of the war by the opposing state. Each state had three options for increasing its power: mobilizing domestic manpower, contracting voluntary allies, or conquering other states and establishing hegemony over them. Each state's choice among these three options was determined primarily by the impact of the spillover problem on its military labor productivity. This factor also determined whether each state could overpower the opposing state simply by increasing its own coalition size or only by a combined strategy of increasing its coalition size and decreasing the opponent's coalition size. By extension, this factor also determined whether each state could stay in the war after an increase in its opponent's coalition size or had to drop out of the war for a time and look for more allies to join its own coalition. All of these variations in grand strategy determined the way in which the international distribution of power shifted back and forth during each war. All of these variations in grand strategy were driven, at root, by the international spillover problem and its impact on each state's labor productivity.

Thus the modern history of grand strategies and international outcomes is a simple story

of cause and effect. The root cause of each hegemonic war was an international spillover problem that caused an irreconcilable dispute over income and changes in each state's labor productivity. Once this cause arose, all of the effects just mentioned were inevitable. From this perspective, shifts in the international distribution of power were simply an intervening variable. They played an intermediary role between this root cause and its various effects.

By explaining how these international spillover problems caused these effects, contractual realism redeems the notion that each states fights hard to create favorable shifts in the international distribution of power and avoid unfavorable shifts in it. In this sense the "structural realists" are correct to assume that shifts in the international distribution of power can motivate the formation of counter-balancing coalitions. They simply focus on the wrong cause of shifts in the international distribution of power and treat such shifts as root causes when they are only intervening variables.

The "structural realists" also make another mistake by defining the international distribution of power strictly by its long-run determinants. By ignoring the short-run factors that boost a state's incentive to put men into military use, structural realists are left with no explanation of the rise of the revisionist state. They simply claim that it is beyond the capacity of a "structural" theory of international politics to *explain* the rise of a revisionist state. In practice, though, they import the *phenomenon* of the revisionist state into the theory by making implicit assumptions about the intentions, motivations or dysfunctions of particular states. By contrast, Contractual Realism provides definitions of relative power and system structure that can explain the rise of the revisionist state. It explains variations in revisionist goals as a function of the type of spillover problem that caused the war and the shift in labor productivity due to the spillover problem.

As mentioned, structural realism makes implicit assumptions about the intentions, motivations or dysfunctions of particular states. In reaction against such extra-theoretic assumptions, conventional critics of structural realism developed a rival theory that explicitly acknowledges the need to make assumptions about state's intentions, motivations or dysfunctions. There variant of conventional realist theory is called *neo-classical realism*. As we will see, the debate between structural realism and neo-classical realism is a false one.

## 2.2 Neo-classical realism

This theory focuses on the fact that in each hegemonic war there was one state that pursued military expansion to establish hegemony over other states by eliminating their political independence and, in some cases, revising territorial boundaries. The theory assumes that one state's adoption of this revisionist goal was the root cause of the war and the root driver of other states' decision to form a "balancing" coalition to oppose it. From this perspective, the tightest constraint on state strategy is the intentions of those who exercise power abroad. If their intentions are benign, then strategy toward them can be accommodating. But if their intentions are aggressive or malevolent, then strategy toward them must be confrontational. This view of state behavior and international outcomes has come to be called "balance of threat" theory.<sup>8</sup> Balancing against threats is the main theme of *neoclassical realism*. It labels the threatening states of European history as "revisionist" states to clarify that their goal was to revise territorial boundaries and sovereignty norms – and this was the root cause of the "balancing" responses of other states that generated hegemonic wars.<sup>9</sup>

This theory suffers from several empirical and logical problems. First, it defines "revision" solely in terms of territorial goals and assumes, therefore, that in each war there was a single revisionist state and the other states that opposed it had no revisionist goals. This assumption is refuted by the evidence. In each war *each* of the opposing states claimed the other was trying to revise the existing distribution of rights and obligations across states, and *each* claimed it therefore had the right to oppose the other by force. Each state therefore saw itself as "balancing" against the "challenge" of "revision" posed by the other state. This was not simply a smokescreen generated by rhetoric or propaganda from the one "true" revisionist state in each war. In fact each state's claim that the other state sought revisions to the status quo was backed by a legitimate source in existing laws or norms. It was exactly this "legal incompatibility" that caused the dispute over the international spillover problem to become irreconcilable (as shown in Part I of this study).

This disparity between the theory and the evidence points to a deeper logical problem

---

<sup>8</sup>Representative works in the balance of threat school include S. M. Walt 1987; Glaser 1994; Evera 1998.

<sup>9</sup>Representative works in the neoclassical realist school include Wohlforth 1994; R. L. C. Schweller 1994; R. L. Schweller 1996; R. L. Schweller 2003; R. L. C. Schweller 2004; Gideon 1998; Zakaria 1992.

with the theory. The only way it can assign revisionist goals to just one state in each war is to assume that the other states in each war pursued the goal of "security" (only). This assumption drives the main causal mechanism of the theory: A security-seeking state chooses its strategy based on the intentions of the opposing state. If its intentions are benign because it is another security-seeking state, then strategy toward it can be accommodating. But if its intentions are aggressive because it is a "revisionist" state, then strategy toward it must be confrontational. This is supposedly how the rise of a revisionist state causes war: It must be opposed.

The problem with this approach is that it embeds each state's behavior in its root goal structure or utility function (in the language of the previous section). This is problematic not only for the reasons outlined above, but also for another reason. It leaves the theory incapable of explaining which of the two opposing states pursued hegemony and territorial revision in each war – the state producing the international spillovers or the state suffering from them. For example, in the French Revolutionary Wars, it was the state *producing* the spillovers, Revolutionary France, that sought hegemony and revision. By contrast, in the Thirty Years War, it was the state *suffering* from the spillovers of the Protestant Reformation, Austria, that sought hegemony and revision. Contractual Realism explains this difference as a consequence of the impact of the spillovers on military labor productivity, under the assumption that each state's goal is simply to maximize its national income (and minimize its national income loss in any international dispute). The causal role of labor productivity is obscured by neo-classical realism because it attributes a state's goal of hegemony or counter-balancing to its utility function as either a "revisionist" state or a "security-seeking" state.

That approach also obscures the causal role of labor productivity in driving patterns of alliance formation and dissolution before and during each war which, in turn, drove repeated shifts in the balance of territorial power. In each war the spillover problem caused changes in labor productivity that determined which state could drive its opponent out of the war simply by increasing its own coalition size and which state could only drive its opponent out of the war by a combined strategy of increasing its coalition size and reducing the opponent's coalition size. This difference explains several mechanisms that drove shifts in the balance of territorial power during each war. It explains why one state entered the war as soon as its

coalition was large enough to enter, even though was not large enough to win. It explains why this state expanded its coalition size by a large margin (by hegemonic conquest in some cases and by voluntary alliance in other cases), yet without winning. This increased the number of allies that its opponent would have to gain to win in a quick route. In turn, it explains why that state waited to enter the war until it had formed a coalition large enough to win in a quick rout (by hegemonic conquest in some cases and by voluntary alliance in other cases). And it explains why that state was able to force its opponent to relinquish previous territorial gains and drop out of the war in a quick rout.

All of these patterns of cause and effect are obscured by neo-classical realism but revealed by Contractual Realism. Contractual Realism also provides a unified explanation of alliance formation and dissolution before, during and after hegemonic wars. By contrast previous theories only explain alliance behavior each of these three periods in isolation.

### **3. Alliance strategies before, during and after a war: Conventional theories *versus* Contractual Realism**

#### **3.1 "Challenging" and "balancing": A new approach**

The conventional theories of alliance strategy ignore a simple fact borne out by the evidence on actual hegemonic wars. In each war the "Balancer" also sought a revision of the status quo (although not a territorial revision). And it was therefore acting wrongly in the eyes of the "Challenger". Once this fact is recognized, it is clear that each of the opposing states in the war faced a "challenge" from the other state that needed to be met by aggregating power. Each of the opposing states in the war was trying to restore the status quo ante and needed to aggregate power to do so. And each state's increase in power posed a "challenge" to the other state which it needed to meet by aggregating power. From this angle it is impossible to distinguish whether each state's increase in power was driven by a "challenging" motive or a "balancing" motive. It was both at the same time.

This conclusion refutes the conventional assumptions about the cause of "balancing". The conventional wisdom assumes that the "revisionist" state (which was trying to "revise" the territorial status quo through military expansion) was not "balancing" against any other

state. But the "balancer" was "balancing" against the growing territorial power of the "revisionist" state. In fact, however, each state was "challenging" the other state's claim to the rights in dispute. And each state was "balancing" against the other state's challenge to its rights. With this point clarified, it becomes possible to explain shifts in the territorial holdings of each state during each war as a consequence of a deeper cause: the dispute over income rights provoked by the international spillover problem. It also becomes evident that the competition for a relative power advantage and the resulting shifts in territorial holdings and power were all simply means to an end – rational strategies driven by a deeper motive cause – the dispute over rights and the drive of each state to secure its rights in the dispute in order to deter further opportunistic claims on its rights by the other state.

Once grand strategies are interpreted in this way – as strategies of contract enforcement – each state's fluctuations in power can be understood and explained as a rational response to changing material opportunities and constraints using the "economics of contract enforcement". (This includes fluctuations in power through resource mobilization, territorial expansion and contraction, and alliance formation and dissolution). Moreover, instead of reifying power as the root cause of all of these strategies and outcomes, power becomes visible as a composite of three elements that each state combines in variable proportions – ones that vary with the evolution of the material opportunities and constraints the state faces.

#### **4. Conclusion: A critical test between Conventional Realism and Contractual Realism**

Conventional Realism is a theoretical paradigm composed of different theories whose causal assumptions contradict each other. The paradigm provides no research agenda to resolve the contradictions. By contrast, Contractual Realism provides a research agenda to resolve the contradictions based on a new theory whose assumptions differ from those of the Conventional Realist theories. In the next chapter, I derive a range of observable implications for differences in grand strategy across states and across wars. In the following chapter, I test these implications against the historical evidence on hegemonic wars. A major goal of the analysis is to conduct a critical test between the Contractual Realist theory and the Con-

ventional Realist theories. If the Contractual Realist theory outperforms the others, then it will renew and deepen the realist tradition of power-centric analysis with greater parsimony and explanatory power.

## CHAPTER 24

### The Contractual Realist theory of grand strategy in hegemonic wars

In the previous Part of this study, I assumed that there are only two states in the international system – the two states engaged in the dispute over the spillover problem. Each state was assumed to have only two options: to fight alone or to drop out of the war. Under this assumption the first round of the war would produce a clear winner and a clear loser.

In this Part of the study, I consider the situation where each state has a third option: to find an ally. Each state chooses among the three options by comparing its expected income from each option and choosing the option that is expected to yield the highest income (given the other state's choice of option). Under this assumption, the winner of the first round only continues to enjoy its victory as long as the loser does not find an ally and reenter the war. If it does, then the winner must decide whether to continue fighting alone, to drop out, or to find an ally in response.

The third option changes the nature of the war completely. In the main conflict between the state causing the spillover effects (state 1) and the state threatened by the spillover effects (state 2), there is no longer a clear winner and a clear loser. If one side would lose fighting alone, then it can find an ally to raise its expected war income above its expected peace income, and thereby stay in the war (or reenter the war). Thus it may happen that *both* of the main antagonists are better off fighting on than dropping out – as long as each one can retain its current allies and find new allies whenever its opponent does. In this kind of war, each state has an incentive to continue finding more allies – and continue picking off its opponent's allies – to force the opponent to drop out and settle for peace. Yet even if one side succeeds in this aim, it might only be able to *keep* the opponent out of the war by retaining its own allies. This is the fundamental motive for establishing hegemony: to retain



allies, by force if necessary. Only one of the two main antagonists in each type of war has this incentive. Which one it is differs from one type of war to the next (e.g. secession wars, revolution wars, and depression wars). These differences are explained by the impact of the spillovers on the relative productivity of military and civilian labor in each state.

In fact the three types of war differ in many aspects of coalition formation, grand strategy and international outcomes. All of these differences are explained by the impact of the spillovers on the relative productivity of military and civilian labor in each state. The key to this analysis is the concept of credibility developed in the previous Part of the study. Credibility is defined as a state's expectation that its income from war is greater than its income from peace. If a state has credibility, then it has an incentive to continue fighting and it can stay in the war. If it lacks credibility, then it has an incentive to stop fighting and it must drop out of the war. If one state has credibility but the other does not, then the first state is the clear winner and the second state is the clear loser. If both states have credibility, then there is total war.

This chapter develops a theory in which each state's credibility is determined by the relative productivity of its military and civilian labor. This factor is a structural cause in the sense that it arises from the economic structure of the spillover problem: Whether the spillovers shift the productivity of the state's military labor, civilian labor, or both. Shifts in labor productivity caused by a spillover problem are structural shifts in two senses. First, labor productivity is one of the most fundamental determinants of a state's income. The higher a state's labor productivity is, the higher its income will be. The lower its labor productivity is – due to a spillover problem for example – the lower its income will be. Second, the relative productivity of labor in military use versus civilian use is the most fundamental determinant of a state's army size and hence its power in competition against other states (as explained in the previous Part of this study). Each state is driven to allocate its labor efficiently between military and civilian use simply to maximize its income – and to minimize its income loss in an adverse contingency. Hence shifts in relative labor productivity are the most fundamental drivers of state strategy and international security competition.

The theory presented in this chapter aims to show that this structural cause lies at the root of state strategies and international outcomes in hegemonic wars. The theory charac-

terizes differences between the three types of spillover problem in terms of labor productivity and derives the observable implications for how the three types of war are expected to differ in the grand strategies of the main opponents and the resulting international outcomes. If these implications are corroborated by the case studies presented in the following chapter, this will prove that it is the relative productivity structure of the spillover problem that causes hegemonic war as we know it in the historical record, rather than some other aspect of the spillover problem or some other cause entirely.

## 1. The origins of hegemony-seeking, balancing and bandwagoning: Alliance incentives in competitive equilibrium

The analysis starts with the basic model presented in the previous Part of this study and extends it by allowing each of the two opposing states to obtain allies. The option to obtain allies is captured simply by loosening the population and territory constraints. Each state's population is assumed to be 100, as before, and each state's territory is assumed to be 100, as before. Each new ally increases the total population of that coalition by 100 and the total amount of territory to be divided between the two coalitions through war by 100. Each *coalition* is assumed to adopt its optimal army size given its total population, just as previously each *state* was assumed to adopt its optimal army size given its total population.

A coalition of two states has a larger optimal army size than a coalition of one state. So a state that had *no* incentive to fight alone – when it had no allies – can *gain* an incentive to fight by obtaining an ally. The opposing state can find an ally of its own in response, however. The resulting dynamics of security competition vary from one type of spillover problem to another due to differences in the impact of the spillover problem on the relative productivity of military and civilian labor within each state.

This logic generates four key points that are used to derive observable implications.

1. If a state's coalition is too small, then it will be better off backing down than standing firm and fighting. So it will have to exit the war (or not enter in the first place).
2. If the state can increase its coalition size by enough, then it will be better off standing firm and fighting than backing down. So it can enter the war (or reenter the war).

3. If its *opponent's* coalition becomes too large, however, then it will be better off backing down than standing firm and fighting. So it will have to exit the war (or not enter in the first place).
4. If it can reduce the opponent's coalition size by enough, however, then it will be better off standing firm and fighting than backing down. So it can enter the war (or reenter).

Given these constraints, each of the two main opponents has an incentive to gain allies simply to ensure that it will be better off standing firm and going to war than backing down in the dispute. Recall that for each state, the stakes in the dispute are whether it will be vulnerable to further income losses from similar actions by the opposing state in the future. So if it fails to gain enough allies, then it will have to back down in the current dispute and it will lose these stakes. From this perspective, alliance formation is simply a consequence of each state's need to fight over the dispute to secure its national income. And the ensuing security competition between alliances is simply a consequence of each state's need to fight over the dispute to secure its national income. This view of alliance rivalry differs fundamentally from previous theories of hegemony-seeking, balancing and bandwagoning. As discussed in the previous chapter, those theories assume that at least one state in each war was driven at root by the goal of power maximization. By contrast, this theory assumes that all states in each war were driven by the goal of income maximization and, in an international spillover crisis, the goal of income loss minimization.

### 1.1 The impact of labor productivity on entry, exit and alliance incentives

The relevant simulation results are shown in figure 24.1. The left column shows the results for a secession spillover problem. The middle column shows the results for a revolution spillover problem. The right column shows the results for a depression spillover problem. In each column, the top panel shows *state 1's* marginal products of military and civilian labor as a function of state 1's allocation of labor between the military and civilian sectors. State 1's optimal army size is the allocation point at which the two marginal products are equal – where the two lines intersect – as before. The middle panel shows *state 2's* marginal products of labor and optimal army size in the same way.

The bottom panel of each column displays at several points. First, it displays the com-

parison between state 1's income from war and its income from peace as a function of both states' coalition sizes. The left side of the panel shows the combination of coalition sizes for which state 1's war income is *less than* its peace income, because its coalition is too small. So state 1 is better off backing down and dropping out of the fighting than standing firm and fighting. This is state 1's "capitulation zone". The right side of the panel shows the combination of coalition sizes for which state 1's war income is *greater than* its peace income, because its coalition was large enough. So state 1 is better off standing firm and making war over the dispute than backing down and dropping out of the fighting. This is state 1's "credibility zone".

Second, the bottom panel of each column displays displays the comparison between state 2's income from war and its income from peace as a function of both states' coalition sizes. The bottom part of the panel shows the combination of coalition sizes for which state 2's war income is *less than* its peace income, because its coalition is too small. So state 2 is better off backing down and dropping out of the fighting than standing firm and fighting. This is state 2's "capitulation zone". The top part of the panel shows the combination of coalition sizes for which state 2's war income is *greater than* its peace income, because its coalition was large enough. So state 2 is better off standing firm and making war over the dispute than backing down and dropping out of the fighting. This is state 2's "credibility zone".

To understand the bottom panels, consider two questions for each panel.

1. How many allies does that state need to gain to move from its capitulation zone into its credibility zone? This number indicates how many allies the state needs to gain to be better off standing firm and fighting than backing down and dropping out of the fighting.
2. How many allies does the opposing state need to gain to force the state from its credibility zone into its capitulation zone? This number indicates how many allies the opposing state needs to gain to force the state to back down and drop out of the fighting.

Seen from this perspective, the bottom panels of the figure reveal exactly the pattern that is expected, given the differences in labor productivity across the three types of spillover

problem. Moving across the bottom row from left to right, state 1's credibility zone gets smaller from one type of spillover problem to the next. Thus state 1 needs more allies to get into its credibility zone in a revolution crisis than in a secession crisis, and even more in a depression crisis than in a revolution crisis. At the same time, state 2 needs fewer allies to force state 1 into its capitulation zone in a revolution crisis than in a secession crisis, and even fewer in a depression crisis than in a secession crisis.

The cause of these differences is shown in the first and second rows of the figure. In the secession case, state 1's new domestic activity that is generating the international spillovers *increases* its optimal army size and *decrease* state 2's optimal army size. So state 1's income from making war over the international dispute is *increased* relative to its income from backing down and remaining at peace, as its new domestic activity intensifies. Hence state 1 needs to obtain comparatively few allies to get into its credibility zone. In the revolution case, the spillovers leave the optimal army sizes of the two states *unchanged*. So state 1's income from making war over the dispute is *left unchanged* relative to its income from backing down and remaining at peace. Hence state 1 needs comparatively more allies to get into its credibility zone. In the depression case, the spillovers *decrease* state 1's optimal army size and *increase* state 2's optimal army size. So state 1's income from making war over the dispute is *decreased* relative to its income from backing down and remaining at peace. Hence state 1 needs even more allies to get into its credibility zone.

The figure also reveals the expected pattern for state 2. Moving across the bottom row from left to right, state 2's credibility zone gets larger from one type of spillover problem to the next. Thus state 2 needs fewer allies to get into its credibility zone in a revolution crisis than in a secession crisis, and even fewer in a depression crisis than in a revolution crisis. The cause of these differences is shown in the first and second rows of the figure. In the secession case, the spillovers *increase* state 1's optimal army size and *decrease* state 2's optimal army size. So state 2's income from making war over the spillover problem is *decreased* relative to its income from backing down and remaining at peace as the spillovers worsen. Hence state 2 needs to obtain comparatively many allies to get into its credibility zone. In the revolution case, the spillovers leave the optimal army sizes of the two states *unchanged*. So state 2's income from making war over the spillover problem is *left unchanged* relative to

its income from backing down and remaining at peace as the spillovers worsen. Hence state 2 needs comparatively fewer allies to get into its credibility zone. In the depression case, the spillovers *decrease* state 1's optimal army size and *increase* state 2's optimal army size. So state 2's income from making war over the spillover problem is *increased* relative to its income from backing down and remaining at peace as the spillovers worsen. Hence state 2 needs even fewer allies to get into its credibility zone.

Some general conclusions emerge from these figures. First, each type of spillover problem produces a different size credibility zone for each state, because each type of spillover problem has a different effect on each state's labor productivity.

**When the spillover problem *increases* a state's productivity of military labor relative to civilian labor, the state's credibility zone is *large*. Conversely, when the spillover problem *decreases* a state's productivity of military labor relative to civilian labor, its credibility zone is *small*.**

Second, the size of a state's credibility zone determines *both* states' incentives for entry, exit and alliance formation.

**When state 1's credibility zone is *large*, as in a secession crisis, the minimum coalition size that state 1 needs to enter the war is *small*. And the minimum coalition size that state 2 needs to force state 1 to exit the war is *large*. Conversely, when state 1's credibility zone is *small*, as in a depression crisis, the minimum coalition size that state 1 needs to enter the war is *large*. And the minimum coalition size that state 2 needs to force state 1 to exit the war is *small*.**

Third, these differences across the three types of spillover problem are robust across all coalition sizes. For example, state 1 needs more allies to get into its credibility zone in a depression conflict than in a secession conflict *regardless of how many allies state 2 has*. To take another example, state 2 needs fewer allies to force state 1 into its capitulation zone in a revolution conflict than in a secession conflict *regardless of how many allies state 1 has*.

## 1.2 What happens when state 2 gains military options to stop the international spillovers by force?

figure 24.1 showed the results when state 2 does not have military options to stop the international spillovers by force. These results revealed the impact of the spillover problem itself on labor productivity and the incentives for entry, exit and alliance formation. Now consider the additional effects that arise when state 2 gains military options to stop the international spillovers by force. The relevant simulation results are shown in figure 24.2. The top row in this figure is the same as the top row in the previous figures. The middle row is the same as the middle row in the previous figures, except for the addition of the dotted lines. These dotted lines show state 2's capacity to use military force to stop the international spillovers and thereby restore its marginal labor productivity (which had been reduced by the spillovers). The bottom row in this figure displays the impact of such military options on both states' credibility and capitulation zones.

The main patterns that were revealed in the previous figures are also evident in this figure:

1. State 1 needs more allies to get into its credibility zone in a revolution crisis than in a secession crisis, and even more in a depression crisis than in a revolution crisis.
2. State 2 needs fewer allies to force state 1 into its capitulation zone in a revolution crisis than in a secession crisis, and even fewer in a depression crisis than in a secession crisis.
3. State 2 needs fewer allies to get into its credibility zone in a revolution crisis than in a secession crisis, and even fewer in a depression crisis than in a revolution crisis.

The main novel impact of state 2's military options is to increase the size of its credibility zones and decrease the size of state 1's credibility zones. Clearly state 2's credibility zones are larger after it gains such military options, while state 1's credibility zones are smaller. This impact only occurs in the secession and revolution cases, however. In each of these cases, state 2's military options produce two income effects that are mutually reinforcing. They increase state 2's income directly by restoring its marginal labor productivity (as evident in the dashed lines in figure 24.2). And they increase its optimal army size (as evident in figure 24.2), thereby giving it a greater share of the system's territory in the war and increasing its income in that way too. In the depression case, by contrast, the two income effects of its

military options are mutually canceling. On one hand, the military options increase state 2's income directly by restoring its marginal labor productivity. On the other hand, though, they decrease state 2's optimal army size, thereby giving it a smaller share of the system's territory in a war and reducing its income in that way.

A second impact of state 2's military options is to reduce the number of allies that it needs to gain to force state 1 to exit the war. Again this impact only occurs in the secession and revolution cases. In each case state 2's military options increase its optimal army size (as evident in figure 24.2), thereby reducing state 1's war income and reducing the size of its credibility zone commensurately.

A third impact of state 2's military options is to increase its opportunities to force state 1 out of the war. As noted, state 2 can drive state 1 out of the war simply by increasing its own coalition size (e.g. moving vertically up the graph). State 2 has this opportunity both before and after it gains military options to stop the international spillovers by force. The opportunity is relatively narrow before it gains such options, however. In the secession case, for example, the opportunity only exists when state 1's coalition size is less than 1.2. In the revolution case, it only exists when state 1's coalition size is less than 1.5. By contrast, after state 2 gains military options to stop the spillovers by force, it has a much greater opportunity to drive state 1 of the war simply by increasing its own coalition size. In the secession case, it can do so whenever state 1's coalition size is less than 2.7. In the revolution case, it can do so whenever state 1's coalition size is less than 3.5.

**Once state 2 gains military options to stop the international spillovers by force, it gains a strong incentive to win the war simply by increasing its own coalition size.**

## **2. Minimum coalition size needed for entry and sustainment of military action**

Let us define a *hegemonic* victory coalition as a binding alliance supported by a unified military command. The hegemon retains its allies either through positive inducements or



through threats to impose costs of some kind.<sup>1</sup> By contrast, a *non-hegemonic* victory coalition is defined as a voluntary alliance in which each ally maintains its own separate military command. Each ally may “come and go” from the alliance by its own unilateral decisions.

With these definitions in mind, consider the bottom row panels of figure 24.1. If a state is in its credibility zone even when fighting alone, then it can enter the war and remain in the war even without allies. If it wins the war, its victory coalition can be non-hegemonic. That is to say, it does not need to bind its allies to the victory coalition through hegemonic control in order to maintain the credibility to deter the losing state from trying to restart the war. By contrast, if a state is in its capitulation zone when fighting alone, then it cannot enter the war or remain in the war unless it has allies. If it wins the war (with the help of allies), its victory coalition must be hegemonic. That is, its victory coalition requires a binding alliance under a unified military command to have the credibility to deter the losing state from trying to restart the war.<sup>2</sup>

### **Result 1.**

In some cases, a state is in its credibility zone when it has no allies. In other cases, a state is in its capitulation zone when it has no allies. In yet other cases, whether it is in its credibility zone or its capitulation zone without allies depends on the opponent’s coalition size.

### **Secession War.**

1. When state 1, the state producing the spillover effects, has no allies:
  - (a) it is in its credibility zone when its opponent has no allies, but
  - (b) it is in its capitulation zone when its opponent has one or more allies.
2. When state 2, the state suffering the spillover effects, has no allies, it is in its capitulation zone.

### **Revolution War.**

1. When state 1, producing the spillover effects, has no allies, it is in its capitulation zone.

---

<sup>1</sup>A “benevolent” hegemon would retain its allies by offering positive inducements (carrots). A “coercive” hegemon would preclude its allies from leaving the alliance through threats to impose costs (sticks). See Snidal 1985 for discussion of the distinction between benevolent and coercive hegemony.

<sup>2</sup>If its victory coalition is composed of natural allies, then its coalition can take the form of a benevolent hegemony. If the victorious state has no natural allies in its winning coalition – only conquered satellite states – then its victory coalition must take the form of a coercive hegemony.

2. When state 2, suffering the spillover effects, has no allies, it is in its credibility zone.<sup>3</sup>

#### **Depression War.**

1. When state 1, producing the spillover effects, has no allies, it is in its capitulation zone.
2. When state 2, suffering the spillover effects, has no allies, it is in its credibility zone.<sup>4</sup>

These results demonstrate that the spillover problem gives the two opposing states different structural incentives for avoiding defeat. These structural differences generate a number of observable implications for differences in their grand strategies.

#### **Implication 1. Searching for allies before entering military action.**

##### **Secession War.**

1. State 1, producing the spillover effects,
  - (a) will not search for allies before entering military action as long as its opponent has no allies.
  - (b) will search for allies before entering military action once its opponent gains at least one ally.
2. State 2, suffering the spillover effects, will search for allies before entering military action.

##### **Revolution War.**

1. State 1, producing the spillover effects, will search for allies before entering military action.
2. State 2, suffering the spillover effects, will not search for allies before entering military action.

##### **Depression War.**

1. State 1, producing the spillover effects, will search for allies before entering military action.
2. State 2, suffering the spillover effects, will not search for allies before entering military action.

---

<sup>3</sup>As long as its opponent has less than four allies.

<sup>4</sup>As long as its opponent has less than four allies.

## **Implication 2. Sustainability of military action without allies.**

### **Secession War.**

1. State 1, producing the spillover effects,
  - (a) will find its military action sustainable without allies as long as its opponent has no allies.
  - (b) will find its military action unsustainable without allies once its opponent gains at least one ally.
2. State 2, suffering the spillover effects, will find its military action unsustainable without allies.

### **Revolution War.**

1. State 1, producing the spillover effects, will find its military action unsustainable without allies.
2. State 2, suffering the spillover effects, will find its military action sustainable without allies.

### **Depression War.**

1. State 1, producing the spillover effects, will find its military action unsustainable without allies.
2. State 2, suffering the spillover effects, will find its military action sustainable without allies.

## **Implication 3. Type of victory coalition.**

### **Secession War.**

1. If state 1 wins a victory in the war (by forcing state 2 to stop fighting and stripping it of allies), then state 1 can sustain its victory without allies. Hence state 1 will secure its victory with a voluntary coalition.
2. If state 2 wins a victory in the war (by forcing state 1 to drop out of the fighting), then state 2 can only sustain its victory with allies. Hence state 2 will secure its victory with a hegemonic coalition.

### **Revolution War.**

1. If state 1 wins a victory in the war (by forcing state 2 to drop out of the fighting), then state 1 can only sustain its victory with allies. Hence state 1 will secure its victory with a hegemonic coalition.
2. If state 2 wins a victory in the war (by forcing state 1 to drop out of the fighting), then

state 2 can sustain its victory without allies. Hence state 2 will secure its victory with a voluntary coalition.

### **Depression War.**

1. If state 1 wins a victory in the war (by forcing state 2 to drop out of the fighting), then state 1 can only sustain its victory with allies. Hence state 1 will secure its victory with a hegemonic coalition.
2. If state 2 wins a victory in the war (by forcing state 1 to drop out of the fighting), then state 2 can sustain its victory without allies. Hence state 2 will secure its victory with a voluntary coalition.

### **3. Requirements for victory: Increasing own coalition size *versus* decreasing opponent's coalition size**

In secession and revolution wars, the state suffering the spillover effects (state 2) can win simply by increasing its own coalition size. This pattern is evident in the panels of figure 24.1 for secession and revolution wars. In each panel, if state 2 simply increases its own coalition size by enough (by moving upward vertically in the graph), then it can force state 1 into its capitulation zone, and thereby force state 1 to exit the war. However, the number of allies that state 2 needs to gain to force state 1 to exit the war depends on state 1's coalition size. The larger state 1's coalition is, the more allies state 2 needs to gain to win simply by increasing its own coalition size.

By contrast, the state producing the spillover effects, state 1, cannot win simply by increasing its own coalition size. It can only win through a combined strategy of increasing its coalition size and decreasing state 2's coalition size. This is also evident in figure 24.1. State 1 *cannot* win simply by increasing its own coalition size (by moving horizontally to the right). State 1 must also reduce state 2's coalition size to win (by moving downward vertically in the graph).

### **Result 2. Requirements for victory.**

**2a.** In each type of war, the state producing the spillover effects, state 1, cannot win simply by increasing its own coalition size. It can only win through a combined strategy of increasing its coalition size and decreasing state 2's coalition size.

**2b.** By contrast, the state suffering the spillover effects, state 2, can win simply by increasing its own coalition size (at least in secession revolution wars).

### **Result 3.**

The number of allies that state 2 needs to win in this way depends on state 1's coalition size.

**3a.** The larger state 1's coalition is, the more allies state 2 needs to win in this way.

**3b.** The smaller state 1's coalition is, the fewer allies state 2 needs to win in this way.

### **Result 4.**

As mentioned, state 1 can only win through a combined strategy of increasing its coalition size and decreasing state 2's coalition size (Result 2 above). More precisely, state 1 can only win by reducing state 2's coalition size to a singleton, e.g. stripping state 2 of all of its allies, including its largest and most committed allies. This requirement for victory is independent of state 1's coalition size. No matter how large state 1's coalition might be, it can only win by stripping state 2 of its largest and most committed allies.

These results demonstrate that the spillover problem gives the two opposing states different structural incentives for the pursuit of victory. These structural differences generate further observable implications for differences in their grand strategies.

## **3.1 Security margins and deprivational allies**

While state 2 can win simply by increasing its own coalition size (Result 2), the larger state 1's coalition is, the more allies state 2 needs to win in this way (Result 3). Therefore state 1 has an incentive to increase its coalition size to make it more difficult for state 2 to win in this way.

### **Implication 4.**

**4a.** State 1 will increase its coalition size in order to make it more difficult for state 2 to win in this way. The more allies state 1 gains for this purpose, the larger its "security margin" will be.

**4b.** State 1 will acquire any allies that increase its coalition size for this purpose, including allies that are small or less committed to its cause. For even this kind of ally serves the purpose, as it deprives state 2 of the opportunity to win simply by gaining  $x$  allies, forcing it to gain  $x + 1$  allies to win. And it deprives state 2 of potential allies, thereby forcing state 2 to look elsewhere for the  $x + 1$  allies that it needs to win.

In so far as the aim of state 1's alliances is to deprive state 2 of these opportunities, state 1's allies can be called deprivational allies.

### **3.2 Quick routs, attrition wars and aggregational allies**

While state 2 can win simply by increasing its own coalition size, state 1 cannot win simply by increasing its own coalition size (Result 2). Therefore:

#### **Implication 5. Impact of gaining an ally.**

**5a.** When state 2 gains an ally, it will force state 1 either to gain an ally in response or to drop out of the fighting quickly.

**5b.** When state 1 gains an ally, it will *not* force state 2 to gain an ally in response or drop out of the fighting quickly. State 2 will be able to stay in the fighting even if it does *not* gain an ally in response.

#### **Implication 6. Speed of victory after gaining an ally.**

**5a.** When state 2 gains an ally, its victories will come in the form of quick routs.

**5b.** When state 1 gains an ally, its victories will come only after long campaigns of attrition.

In so far as the aim of state 2's alliances is to compel state 1 to drop out of the fighting quickly, state 2's allies can be called aggregational allies.

### 3.3 Size of victory coalition

The larger state 1's coalition is, the more allies state 2 needs to gain to win simply by increasing its own coalition size (Result 3). Hence after state 1 wins the war, the larger its coalition is, the more allies state 2 would need to gain to "turn the tables" and transform its loss into a win simply by increasing its own coalition size. After state 1 wins the war, therefore, it will increase its coalition size in order to make it more difficult for state 2 to "turn the tables" in this way. Again state 1 will acquire any allies that increase its coalition size for this purpose, including allies that are small or less committed to its cause.

By contrast, after state 1 loses the war, it cannot "turn the tables" and transform its loss into a win simply by increasing its own coalition size. Hence after state 2 wins the war, it will not increase its coalition size to avert such a risk.

#### **Implication 7. Size of victory coalition.**

##### **Secession War.**

State 1's victory coalition will be all-inclusive.

State 2's victory coalition will be selective.

##### **Revolution War.**

State 1's victory coalition will be all-inclusive.

State 2's victory coalition will be selective.

##### **Depression War.**

State 1's victory coalition will be selective.

State 2's victory coalition will be selective.

The next three chapters present case studies that test these implications against the historical evidence.

## CHAPTER 25

### Case study: Grand strategy in the Thirty Years War

#### 1. Pre-war alliance formation

*State 2, the state suffering the spillover effects, will search for allies before entering military action (implication 1b.).*

When the Bohemian crisis erupted in May, 1618, the Austrian government in Vienna immediately looked to Spain for a commitment of military support. However, the Spanish government was already overstretched militarily and financially. So while the king and his ministers wanted to provide a military commitment to Austria, they simply could not.

*State 1, the state producing the spillover effects, will not search for allies before entering military action (implication 1a.).*

The Kingdom of Bohemia was composed of five provinces: Bohemia, Moravia, Silesia, Upper Lusatia and Lower Lusatia. The Protestants who launched the crisis of May, 1618, were centered in the province of Bohemia. Although they did ask the other provinces for military support before entering military hostilities in August, 1618,, they did not look beyond the Kingdom of Bohemia for military support before this point.

#### 2. Sustainability of military action

*State 1, the state producing the spillover effects, will find its military action sustainable without allies (implication 2a.).*

Even though the Bohemian Protestants were fighting without allies, their military action



was sustainable.

*State 2, the state suffering the spillover effects, will find its military action unsustainable without allies (implication 2b.).*

The Austrian government could not sustain its military action without allies. In the fall of 1618, Austrian forces were too weak to stand up to an invasion by the Bohemian Protestant forces which drove deep into Austrian territory. Again in the spring of 1619, Austrian forces were too weak to prevent a Bohemian invasion that drove all the way to Vienna.

### 3. Alliance formation during the war

#### 3.1 Impact of state 2's alliance formations

*When state 2 gains an ally, it will force state 1 either to gain an ally in response or to drop out of the fighting quickly (implication 5a.). Therefore, when state 2 gains an ally, its victories will come in the form of quick routs (implication 6a.).*

This pattern occurred four times between 1620 and 1634. The first time was in 1620 when Austria gained Spain, the Catholic League and Saxony as allies. Early in the year the Spanish Council of State decided to commit a force of 20,000 to the Austrian cause. In July the Catholic League committed a force of 30,000. The Austrian coalition quickly defeated the Protestants in upper and lower Austria, Lusatia and the lower Palatinate. A few months later the Austrian coalition scored a decisive military victory in the battle of the White Mountain, forcing the Bohemian Protestants out of the war.<sup>1</sup> In the summer of 1621 the Austrian coalition forces overran and occupied the upper Palatinate in a quick campaign, sending Frederick V, its ruler and the erstwhile King of Bohemia, into exile. In the summer of 1622, the Austrian coalition scored a string of quick military victories in the Rhineland that cemented its position of dominance in southern Germany.<sup>2</sup> The Protestant coalition tried to make a comeback the following summer, but was defeated decisively in the

---

<sup>1</sup>Parker and Adams 1997, pp. 52–5.

<sup>2</sup>Parker and Adams 1997, pp. 57–8.

battle of Statlohn. "Frederick...gave up all further military ambitions and surrendered his cause without reservation to the mediation of James I [King of England]." <sup>3</sup> His capitulation brought the "Palatine" phase of the Thirty Years War to a close.

The second time was in 1625 when Austria gained the Duchy of Friedland as an ally. This increased the size of the Austrian coalition by 30,000 troops. By this time Denmark had joined the Protestant coalition and taken the leading role. In June the Danish army, numbering 20,000, marched into northern Germany to save the Protestants principalities from Austrian domination. Yet the Danes were "forced to fall back, only avoiding disaster because [the Austrian generals] Tilly and Wallenstein quarreled about the limits of their authority." <sup>4</sup> The Protestant forces tried to advance again in the summer of 1626, but their Danish contingent was beaten decisively at the battle of Lutter. Without Danish protection the Protestant principalities of northern Germany fell quickly to the Austrian coalition forces. Denmark, defeated and financially drained, ended its intervention in Germany and promised not to intervene again. <sup>5</sup> This ended the Danish phase of the war.

The third time was in the spring of 1632. Protestant coalition forces, led by Sweden, had overrun much of Germany and were headed toward Bavaria and Austria, the heartlands of Austria's coalition. General Wallenstein had been fired by the Austrian Emperor in 1630 under pressure from the Catholic German princes. In April of 1632 he was rehired and commissioned to raise a fresh army to reinvigorate the Austrian coalition. Splitting his forces, one division held down the Swedish army at Nuremberg while the other division defeated the Saxon army occupying Bohemia and Silesia. After the Swedish army retreated into north-western Germany, Wallenstein captured Saxony and forced it out of the Protestant coalition. This halted the long Swedish campaign of 1630 to 1632 to protect German Protestants from the Austrian Catholic forces. <sup>6</sup>

The fourth time was in September of 1634. Austria's army numbered 18,000. Its ally

---

<sup>3</sup>Parker and Adams 1997, p. 61.

<sup>4</sup>Parker and Adams 1997, pp. 67–8.

<sup>5</sup>Denmark fought on within its own boundaries, simply to defend the homeland from an invasion by Wallenstein's forces. While successful in this defensive goal, the Danish king committed in the final peace agreement not to intervene in Germany again. Parker and Adams 1997, pp. 69–71.

<sup>6</sup>Parker and Adams 1997, pp. 100, 117–8.

Spain sent an army of 15,000. The coalition force of 33,000 delivered a decisive defeat to the Protestant coalition force, numbering 25,000, in the battle of Nordlingen on September 2. This battle marked the end of the Swedish phase of the war. The Protestant forces pulled out of southern Germany and Sweden left the Heilbronn League (a military alliance that had been formed to protect the Protestant principalities in southern Germany). Saxony switched sides and joined Austria and Bavaria to secure a Habsburg hegemony in southern Germany. The Peace of Prague, signed in May of 1635, formalized the Austrian coalition's victory.<sup>7</sup>

### 3.2 State 1's formation of security margins

*State 2 can win simply by increasing its own coalition size. But the more allies state 1 has, the more allies state 2 needs to gain to win in this way. Hence state 1 will increase its coalition size, thereby creating a security margin, to make it more difficult for state 2 to win in this way (implication 4.1s).*

This pattern occurred four times between 1620 and 1635. The first time was in 1622. Two years earlier, Austria had increased its coalition size by gaining Spain, the Catholic League and Saxony as allies. After their coalition defeated the Bohemian Protestants and overran the Palatinate in the fall of 1620, its ruler, Frederick V, began searching for allies to broaden the Protestant coalition. England and the Dutch Republic refused to join. But a large force of more than 50,000 troops coalesced from the territories of two Protestant principalities in Germany, Brunswick and Baden-Durlach, and the recruitment of Protestant soldiers from Switzerland and France.<sup>8</sup>

The second time was in the spring of 1625 when Denmark joined the Protestant coalition. As the Danish forces marched into northern Germany, the local duchies of Mecklenburg supported their campaign.<sup>9</sup>

The third time was in 1632. After Swedish forces won the battle of Breitenfeld, their general and king, Gustavus Adolphus, forced a number of the German principalities to join

---

<sup>7</sup>Parker and Adams 1997, pp. 126–9.

<sup>8</sup>Parker and Adams 1997, pp. 56–8.

<sup>9</sup>Parker and Adams 1997, p. 70.

the Protestant coalition. While their force contributions were small, their territories provided quarters and supplies for the Swedish forces.<sup>10</sup> In 1633 his successor formed an alliance of principalities in western Germany – the Heilbronn League – that was intended to enlarge the Protestant coalition even further.<sup>11</sup>

The fourth time was in 1635. After Swedish forces lost the battle of Nordlingen and the Heilbronn League fell apart, the Treaty of Prague gave Austria control over much of Germany. In response, the Protestant coalition expanded to include France.<sup>12</sup>

### 3.3 Impact of state 1's alliance formations

*When state 1 gains an ally, it will not force state 2 to gain an ally in response or drop out of the fighting quickly. State 2 will be able to stay in the fighting even if it does not gain an ally in response (indeed even if it loses allies) (implication 5.2s). When state 1 gains an ally, therefore, its victories will come only after long campaigns of attrition (implication 6.1s).*

This pattern occurred three times between 1620 and 1642. The first time was in 1622 when a force of more than 50,000 troops joined the Protestant coalition from Germany, Switzerland and France (as described above). This growth of the Protestant coalition did not force the Austrian coalition to gain an ally or drop out of the fighting, however. On the contrary, the Austrian coalition prevailed over the Protestant armies in the key battles of Wimpfen and Höchst in the spring of 1622, thereby ending the Palatine phase of the war (as described earlier).<sup>13</sup>

The second time was in 1625 when Denmark invaded northern Germany and the duchies of Mecklenburg supported their campaign (as mentioned above).<sup>14</sup> This did not force the opposing Austrian coalition to gain an ally or drop out of the fighting. On the contrary, it defeated the Danes easily (as described earlier).<sup>15</sup> In December of 1625 England and the

---

<sup>10</sup>Parker and Adams 1997, pp. 114–5.

<sup>11</sup>Parker and Adams 1997, pp. 120–22.

<sup>12</sup>Parker and Adams 1997, pp. 131–4.

<sup>13</sup>Parker and Adams 1997, pp. 56–8.

<sup>14</sup>Parker and Adams 1997, p. 70.

<sup>15</sup>Parker and Adams 1997, p. 68.

Dutch Republic joined the Protestant coalition by committing to provide financial subsidies to Denmark.<sup>16</sup> Yet again it did not force the Austrian coalition to gain an ally or drop out of the fighting. It defeated the Danes at Lutter and ended the Danish phase of the war (as described above).

The third time was in 1635 when France joined the Protestant coalition.<sup>17</sup>

### 3.4 The exception that proves the rule: Austria *loses* allies and drops out

After the Danish phase ended, the Austrian coalition consolidated its hegemony in Germany. Under Wallenstein's command, Austrian forces established a system of military recruitment and taxation designed to support the process of re-catholicizing the lands and peoples of Germany.<sup>18</sup> Even the Catholic rulers of Catholic principalities complained that the burdens of taxation and military rule were too heavy. They forced the Austrian Emperor to dismiss Wallenstein and reduce the size of the coalition army.<sup>19</sup> In effect this act reduced the Austrian coalition to a singleton: The original Austrian army led by General Tilly. As one historian wrote,

"The truth was that Tilly now shouldered the burden of defending the Catholic cause in Germany almost alone, since the Army of Flanders had become too weak to defend any place outside [the Spanish King] Philip IV's patrimonial lands and the troops of [Austrian Emperor] Ferdinand II were too involved elsewhere to help. The principal reason for both developments was simple. In the course of 1629, both branches of the House of Habsburg, [the Spanish and the Austrian], thanks to the tireless diplomacy of the count-Duke of Olivares, had become fatally involved in a major war with France in Italy."<sup>20</sup>

Thus Austria was fighting alone in Germany – without allies. *State 2, the state suffering the spillover effects, will find its military action unsustainable without allies (implication 2b.).*

---

<sup>16</sup>Parker and Adams 1997, p. 69.

<sup>17</sup>Parker and Adams 1997, pp. 129–31.

<sup>18</sup>Parker and Adams 1997, pp. 70–1, 87–90.

<sup>19</sup>Parker and Adams 1997, pp. 90–1.

<sup>20</sup>Parker and Adams 1997, p. 92.

Austria was brought to the brink of destruction by Sweden's entry into the Protestant coalition in 1630. The Swedish army remained cornered in northern Germany by Austrian forces until the spring of 1631. After Austrian forces captured and ravaged the Protestant city of Magdeburg, however, the Duchy of Brandenburg joined the Protestant coalition. And after Austrian forces pillaged the Saxon countryside, Electoral Saxony also joined the Protestant coalition. These expansions of the Protestant coalition enabled a resounding victory over the Austrian forces in the battle of Breitenfeld. After the remnants of the Austrian army fled south, Swedish forces then overran much of Germany.<sup>21</sup>

#### 4. The type and size of victory coalitions

*State 2 will secure its victory with a hegemonic coalition (**implication 3.2s**).  
State 2's victory coalition will be selective (**implication 7.2s**).*

Austria established a hegemonic victory coalition in Germany twice during the war. The first time was in 1627 after it defeated Denmark. Austrian asserted military control over captured Protestant principalities and collected taxes to fund its occupation.<sup>22</sup> While it restored the Catholic religion in the occupied regions,<sup>23</sup> its hegemony was driven by a deeper motive. It wanted to secure its earlier gains in restoring Catholicism in Austria and Bohemia from 1620 to 1624 from a resurgence of Protestant power in Germany. This motive for establishing hegemony over Germany outlasted the goal of restoring Catholicism in Germany – and in this sense it was a more fundamental goal.<sup>24</sup> After the Austrian forces won victories in Germany in 1634 and 1635 and established the Peace of Prague, Austria retained the goal of imposing hegemony over Germany yet abandoned the goal of restoring Catholicism there.

---

<sup>21</sup>Parker and Adams 1997, pp. 110–6.

<sup>22</sup>Parker and Adams 1997, pp. 71, 90–1.

<sup>23</sup>Parker and Adams 1997, pp. 81–4, 87–9.

<sup>24</sup>Parker and Adams 1997, 129. By 1635 "...the formation of Imperial policy was in the hands of pragmatists... who realized that sacrifices would be required if the gains of the 1620s were to be recovered".

Again it established an extractive hegemony in Germany to fund its occupying forces.<sup>2526</sup> Yet it conceded that the Protestant principalities could retain their religious rights.<sup>27</sup>

In both cases, in the late 1620s and again in the mid-1630s, Austria's victory coalition was selective. It only included the states that it was occupying to secure its victory and fund its forces.

*State 1 will secure its victory with a voluntary coalition (implication 3.1s).  
State 1's victory coalition will be all-inclusive (implication 7.1s).*

Between 1638 and 1642, Sweden and France, the main members of the Protestant coalition, won enough battles to convince the leader of the Austrian coalition to settle for peace.<sup>28</sup> Although it took another five years to reach a final settlement, the Treaty of Westphalia, Austria conceded defeat on the issues over which it had been fighting in Germany. It conceded religious freedom to the German Protestants and it committed not to impose its will on the German principalities by force. In effect Austria gave up everything it had gained since 1624. It was left with only its original gains in the religious struggle in Austria and Bohemia from 1618 to 1624.<sup>29</sup>

The victorious powers – Sweden, France and their allies in Germany – guaranteed the final settlement by securing for the German territorial rulers the right to form alliances with non-German powers (such as Sweden and France) to defend their treaty rights.<sup>30</sup>

Together they secured a victory coalition that was voluntary and all-inclusive. It was all-inclusive in that it included all of the principalities in Germany – both Protestant and Catholic – and all of the states beyond Germany that had fought in the war – Sweden, the Dutch Republic, France, Spain and Austria.<sup>31</sup> The aim in including all states was to maximize

---

<sup>25</sup>Parker and Adams 1997, pp. 147–8.

<sup>26</sup>Parker and Adams 1997, more.

<sup>27</sup>Parker and Adams 1997, pp. 128–9.

<sup>28</sup>Parker and Adams 1997, pp. 150–2.

<sup>29</sup>Parker and Adams 1997, pp. 163–4.

<sup>30</sup>Croxton and Parker 2009: 85.

<sup>31</sup>Parker and Adams 1997, pp. 161–2, 164–7.

deterrent strength.<sup>32</sup> The victory coalition was voluntary in that no member of the coalition maintained coercive military control over another member's territory or government. Each member committed to recognize and respect the political independence of all other members and withdraw its troops onto its own territory.<sup>33</sup>

---

<sup>32</sup>Croxton and Parker 2009: 80. "France alone could never have sufficient strength to guarantee that Spain would remain peaceful; only many states acting together could create the necessary balance. Thus, Richelieu conceived the idea of international leagues to enforce the treaties. The idea was not unprecedented...but the Congress of Westphalia applied the concept of an international peace organization to all of Europe."

<sup>33</sup>Parker and Adams 1997, pp. 168–9.



## CHAPTER 26

### Case study: Grand strategy in the French Revolutionary Wars

#### 1. Pre-war alliance formation

*State 2, the state suffering the spillover effects, will search for allies before entering military action (implication 1.2r).*

When the French Revolution turned radical in the summer of 1791, the Austrian government began to search for allies to form a great power coalition and threaten to intervene in France if the revolutionaries did not moderate their domestic political agenda. None of the conservative monarchies was willing to give Austria a military commitment. Nonetheless, Austria and Prussia issued the Declaration of Pillnitz in August of 1791 which threatened that if the revolutionaries did not respect the French king and monarchical form of government, they would mount a military invasion into France – provided a coalition of great powers supported it.

*State 1, the state producing the spillover effects, will search for allies before entering military action (implication 1.1r).*

In the fall of 1791, the French revolutionaries sent inquiries to England to see if it was interested in an alliance against the conservative monarchies.

## 2. Sustainability of military action

*State 2, the state suffering the spillover effects, will find its military action unsustainable without allies (implication 2.2r).*

When the war broke out in April of 1792, the French forces were unable to sustain their actions. In May a French offensive into the Austrian Netherlands was aborted upon meeting "modest" Austrian forces.<sup>1</sup> When a joint coalition of Austrians and Prussians invaded France in August, the French could not maintain much military resistance, losing the garrison of Longwy and the city of Verdun. The German coalition advanced far into French territory.<sup>2</sup>

*State 1, the state producing the spillover effects, will find its military action unsustainable without allies (implication 2.1r).*

When war erupted, it appeared that Austria had a firm ally in Prussia. Each state supplied about 40,000 troops to their joint invasion of France in August, yielding a total force of about 80,000. By the time of the battle of Valmy in September, however, the joint force had fallen in size by more than half due to illness and material shortages – to less than 40,000. While it was nominally a coalition force of two allies, it was in fact smaller than the French force of 64,000. After France won the battle at Valmy, it became clear that Prussia was only weakly committed to Austria. Senior Prussian officers made repeated efforts to reach an understanding with senior French officers. Although no agreements were reached, the Prussian force was able to achieve its goal of retreating into Germany. After their loss at Valmy, the defeated and depleted coalition of Prussia and Austria could not maintain much military resistance to French forces.<sup>3</sup>

---

<sup>1</sup>Blanning 1996, p. 73.

<sup>2</sup>Blanning 1996, pp. 73–5.

<sup>3</sup>Blanning 1996, pp. 74–82.

### 3. Alliance formation during the war

#### 3.1 Impact of state 2's alliance formations

*When state 2 gains an ally, it will force state 1 either to exit the war quickly or to gain an ally to avoid having to exit (**implication 5.2r**). Therefore, when state 2 gains an ally, its victories will come in the form of quick routs (**implication 6.2r**).*

This pattern occurred three times between 1792 and 1802. The first time was in the fall of 1792 when Austria gained Prussia as an ally for its invasion of France. The French army was not large enough to secure both the northern frontier with the Austrian Netherlands (Belgium) and the eastern frontier with Germany. After its failed attempt to invade Belgium in May, the army remained focused on the potential for a successful invasion to neutralize the Austrian forces there. In late August the Prussian army invaded from the east and quickly captured the garrison of Longwy and the city of Verdun. The Prussian king ordered his commander to advance on Paris. Sensing the danger, the French commander on the northern frontier moved his army southward to stop the Prussian advance. For a week or two the opposing armies circled in defensive maneuvering for positional advantage. This delay provided time for the French to mobilize an additional force of 26,000 men and move it to the front in preparation for the upcoming battle with the Prussians.<sup>4</sup>

If it had not been for this delay and the arrival of this large force, the French victory at Valmy on September 20 would have been impossible. The Austrian and Prussian coalition would have taken Paris and France would have been forced to exit the war. France avoided this disaster not by finding an ally, but by mobilizing 26,000 men from its domestic population.<sup>5</sup> Yet in a way, the ability to tap such military manpower on such short notice was akin to the ability to contract an ally that had a standing army ready for battle. And France's capacity to mobilize such manpower was due to the revolutionary fervor of its lower classes rather than to the ordinary conscription methods of 18th century states. In this sense France

---

<sup>4</sup>Blanning 1996, pp. 71–5.

<sup>5</sup>Ross 1973, 39. Refugees from neighboring states were eager to be mobilized into the French military cause. "After April, 1792, the French created a series of foreign legions: Belgian, Swiss, German, Savoyard and Dutch. These units consisted of refugees who, with French help, intended to revolutionize their homelands."

had found an ally: Its revolutionary classes that never would have fought for the old regime so willingly or in such large numbers. This pattern of extraordinary domestic mobilization to avert total military disaster would be repeated again and again. By turning to the domestic population as a substitute for alliance manpower, the new French government would transform domestic politics and society beyond recognition.

The second time was in the winter of 1793. By this time French forces had captured Belgium and a swath of territories in western Germany. Alarmed by these developments, England and Spain joined the Austrian coalition.<sup>6</sup> Coalition forces recaptured the German territories in a few days of fighting in March and April.<sup>7</sup> Austrian forces quickly recaptured Belgium in March and moved into northern France in May. On the southern front Spanish forces moved into France in April.<sup>8</sup>

”If the allies [camped on French soil] could defeat the French field armies and capture the fortified cities along the Franco-Belgian frontier, Paris would lie helpless before them. The fall of Paris would in all probability result in the collapse of the Republic.”<sup>9</sup>

The revolutionary leaders in Paris ordered the French generals to launch counteroffensives to repel the invaders. Yet they failed and coalition forces continued to move cautiously but steadily toward Paris.<sup>10</sup>

To make matters worse, domestic revolts began to erupt in the south and west of France as early as March, threatening to destroy the revolution and restore the monarchy. In the countryside the counterrevolutionary movement was driven by peasants opposed to military conscription and nobles seeking a return to royal rule. In the cities it was driven by a combination of disaffected revolutionaries from Paris and local government officials who preferred a federalist constitution to the centralized regime of the governing revolutionaries. As the

---

<sup>6</sup>Ross 1973, 43-4. Portugal, the Holy Roman Empire, and a number of Italian principalities soon joined as well.

<sup>7</sup>Ross 1973, 49. The French garrison in Mainz held out until July.

<sup>8</sup>Ross 1973, pp. 45-9; Blanning 1996, pp. 98-100.

<sup>9</sup>Ross 1973, p. 48.

<sup>10</sup>Ross 1973, p. 48.

insurgency spread through half of the country's provinces, the revolutionary government had to divert large numbers of troops from the war against the Austrian coalition to fight it. The revolutionary leaders stopped all offensive operations against the coalition and reverted to a purely defensive posture.<sup>11</sup> In effect France had exited the war.

The third time was in 1798 when Austria gained England and Russia as allies.<sup>12</sup> The Austrian coalition then captured three key regions from French forces in fast campaigns. In late March of 1799, the coalition forces won two major battles in southern Germany, inducing the French forces to abandon the region. After they retreated back to the west bank of the Rhine, "...any hope of [French] victory in the German front was doomed".<sup>13</sup>

In mid-April of 1799, a Russian force of 24,000 arrived in northern Italy to join the Austrian force there. In a string of victories interrupted by only a few French counteroffensives, the coalition forces retook almost all of northern Italy from the occupying French forces. In the middle of the campaign France ordered its army in southern Italy to move north to assist the effort to save northern Italy. France thereby lost military control of its satellite ally, the Republic of Naples, which quickly fell to a combination of local and Russian forces aided by the British Navy. The movement of French forces to the north made no difference, however. After repeated losses to the Austrian coalition in June, "...the French army [in northern Italy] was too weak to launch offensives".<sup>14</sup> In July the coalition proceeded to besiege the few fortresses still held by the French. After a French counteroffensive failed completely in August, the French "army was far too disorganized to embark on further offensives."<sup>15</sup> It retreated to the western edge of the north Italian plain, set up a defensive line and hunkered down.<sup>16</sup> "Once French military might had departed...[from] most places in Italy, the [Italian] counterrevolutionaries had no difficulty in seizing power on their own [from the weak French

---

<sup>11</sup>Ross 1973, pp. 51–5.

<sup>12</sup>Ross 1973, pp. 196–7.

<sup>13</sup>Blanning 1996, 232–3. "If the Archduke had pressed home his advantage, a defeat might have become a rout and this critical sector might have been neutralized for the rest of the war"; Ross 1973, 230–2. Quotation from page 231.

<sup>14</sup>Ross 1973, 239–47. The quotation is from Ross, page 249; Blanning 1996, pp. 235–7.

<sup>15</sup>Ross 1973, p. 251.

<sup>16</sup>Ross 1973, pp. 249–51.

puppets]”.<sup>17</sup> Thus the French army had lost Italy and was out of the fighting there.

Domestic revolt erupted again in the south and west of France. The rebels switched from their long-running guerrilla tactics to the organization of large-scale combat forces. At least 32,000 participated. The plan was to launch simultaneous revolts across the land that would enable the foreign forces of the Austrian coalition to invade and restore the French monarchy. It took at least 45,000 troops from the Revolutionary Army to pacify the country and secure the revolutionary government.<sup>18</sup> As the regime focused on saving itself from overthrow, France was again thrown into a purely defensive posture toward its external enemies.

### 3.2 Impact of state 1’s alliance formations

*When state 1 gains an ally, it will not force state 2 to gain an ally in response or drop out of the fighting quickly. State 2 will be able to stay in the fighting even if it does not gain an ally in response (indeed even if it loses allies) (implication 5.2r). When state 1 gains an ally, therefore, its victories will come only after long campaigns of attrition (implication 6.1r).*

This pattern occurred three times between 1793 and 1796. The first time was in 1793. The Austrian coalition had already reconquered Belgium and pushed the French forces well inside French borders (as related above). Fearing the foreign invasion would reach Paris and overthrow the regime, the revolutionary government instituted a mass conscription in August that increased the size of the French army far beyond what the tax revenue of France alone could support.<sup>19</sup> While most of the troops were Frenchmen, the mass conscription was like finding an ally in two senses. The force was as large as a normal-size French army *plus* a normal-size army of at least one other state. And France would need to conquer other states to raise the tax revenue needed to fund the force. In effect, France was raising a coalition-size army *in advance* of conquering the allies needed to fund it. Every time it conquered another state, it gained the additional tax base needed to fund another “ally” worth of troops conscripted from its own domestic population.

---

<sup>17</sup>Blanning 1996, p. 243.

<sup>18</sup>Ross 1973, pp. 253–6; Blanning 1996, pp. 247–8.

<sup>19</sup>Blanning 1996, pp. 100–1, 109.

Despite the growth of the French army, the Austrian coalition forces initially continued their offensives into the French homeland. Over the next three months, from September through December of 1793, the French forces stopped the invasion and pushed the Austrian coalition forces out of France – back into Belgium in the north and across the Rhine River in Germany to the east. It was not a quick rout, though, but a long campaign of attrition, as indicated by a number of metrics. The French forces actually lost a large majority of the battlefield engagements with the coalition forces.<sup>20</sup> The French suffered heavy troop losses in a number of the engagements. The few engagements that they won were defensive engagements – successes in halting offensives by the coalition. When the coalition forces retreated, it was not because they were losing the majority of engagements but because the cost of continuing to win the majority of engagements would be too high. At the end of the campaigning season in late December, Austria was certainly not on the brink of deciding to drop out of the war. Rather it was resolved to bring the fight to the French with more tenacity in the spring when campaigning resumed.<sup>21</sup>

The second time was in the spring of 1794 when French forces went on the offensive in the north. Over a five month period, from mid-April to mid-September, they pushed the coalition forces out of Belgium. Yet again it was a long campaign of attrition. The French forces suffered heavy losses in a number of the engagements and lost a majority of them.<sup>22</sup> The French launched a major offensive in Germany in the summer of 1795 with the goal of forcing Austria out of the war. Two separate French forces pushed across southern Germany in August and September in an effort to reach Austria and bring Vienna to terms. But the campaign failed completely due to strong counteroffensives by the Austrian coalition forces.<sup>23</sup>

The third time was in 1796 in northern Italy. During April and May, French forces captured the provinces of Piedmont and Lombardy without suffering any major losses. Their

---

<sup>20</sup>According to Ross, the Austrian coalition forces won 11 engagements while the French forces won seven engagements. Ross 1973, pp. 69–72.

<sup>21</sup>Blanning 1996, 109–12. "The Austrians were by no means ready to abandon the struggle"; Ross 1973, pp. 69–72.

<sup>22</sup>According to Ross, the Austrian coalition forces won nine engagements while the French forces won seven engagements. Ross 1973, pp. 78–82; Blanning 1996, pp. 112–6.

<sup>23</sup>Blanning 1996, pp. 153–8; Ross 1973, pp. 91–3.

efforts to continue eastward through the rest of northern Italy turned into a long campaign of attrition, however. Between July and January of 1797, the Austrians launched four counteroffensives to try to stem the French tide and regain northern Italy. Although the Austrian forces lost most of the engagements and conceded the entire region in the end, they did inflict heavy troop losses on the French forces in a number of the engagements.<sup>24</sup>

Over the next two months, the French forces in Italy pushed into the Tyrol. They marched to Vienna in April and brought Austria to terms, thereby ending the War of the First Coalition.

### 3.3 State 1's formation of security margins

*State 2 can win simply by increasing its own coalition size. But the more allies state 1 has, the more allies state 2 needs to gain to win in this way. Hence state 1 will increase its coalition size, thereby creating a security margin, to make it more difficult for state 2 to win in this way (implication 4.1r).*

France increased its coalition size to enlarge its security margin three times between 1792 and 1799. The first time was in 1794. After being invaded by the coalition forces the previous year (see above), France bounced back and established security margins on the northern and eastern frontiers. French forces captured Belgium in a difficult campaign of attrition in the spring and summer of 1794.<sup>25</sup> In quick campaigns they expanded the security margin to include the Dutch Republic and Germany up to the Rhine River.<sup>26</sup>

The second time was in the summer of 1796 after French forces captured the northern Italian provinces of Piedmont and Lombardy from Austrian forces. The French quickly expanded their security margin to include central and southern Italy (e.g. Tuscany, the Papal States and Naples).<sup>27</sup> The third time was after the War of the First Coalition ended in 1797. Over the next 18 months France continued to expand its security margin, adding

---

<sup>24</sup>Blanning 1996, pp. 147–53; Ross 1973, pp. 97–104.

<sup>25</sup>Ross 1973, pp. 78–82, 91; Blanning 1996, pp. 112–6.

<sup>26</sup>Ross 1973, pp. 81–2, 88–9.

<sup>27</sup>Ross 1973, pp. 99–100, 104–5; Dwyer 2008, p. 241.



the Dutch Republic, Switzerland, the Papal States and Naples (again).<sup>28</sup>

#### 4. The type and size of victory coalitions

*State 1 will secure its victory with a hegemonic coalition (implication 3.1r).  
State 1's victory coalition will be all-inclusive (implication 7.1r).*

In most of the lands that France conquered, it established an extractive hegemony through military coercion. The victorious French generals extracted large money payments from the conquered governments and used some of the proceeds to pay its troops while sending the rest back to the home government in Paris. French troops also requisitioned food and supplies from the conquered lands. This type of hegemony was established in Belgium, the Dutch Republic, some areas of Germany and most of Italy. The French empire of conquest expanded gradually but steadily over the course of the war to include almost all of Europe. The smaller states were annexed or turned into satellite republics.<sup>29</sup> The larger states – Austria and Prussia – were coerced into becoming nominal allies of France even though they harbored the intention eventually to reenter the war against it.

*State 2 will secure its victory with a voluntary coalition (implication 3.2r).  
State 2's victory coalition will be selective (implication 7.2r).*

After the end of the Napoleonic Wars in 1814, the victorious powers signed the Treaty of Paris and established the Concert of Europe to stabilize the peace. The Concert was a selective coalition in that it only included the great powers – England, Prussia, Austria and Russia – not any of the smaller states that had been involved in the war (such as Spain, Italy or the other German states). This limitation of membership was deliberate on the part of the great powers.<sup>30</sup> Each of the four victorious powers was a voluntary member of the

---

<sup>28</sup>Dwyer 2008, pp. 452–3.

<sup>29</sup>Blanning 1996, pp. 158–63; Ross 1973, pp. 80–1, 88–9, 91, 99–100, 104; Dwyer 2008, pp. 225–7, 265, 275, 281–2, 452–3.

<sup>30</sup>Sinnreich 2009: 142–3. "[The four victorious great powers] agreed to retain control of major territorial

concert: none of them maintained any kind of hegemony over the others.

---

questions among themselves.” France was excluded from the Concert initially, but admitted within a few years.

Part VII

**The path of power at home: Domestic  
politics in hegemonic wars**

## CHAPTER 27

### A Contractual Realist theory of domestic politics in a unitary actor at war

#### 1. Previous literature on the domestic effects of war

Previous studies have concluded that when a state engages in international conflict, its domestic institutions tend to become more centralized.<sup>1</sup> These studies conclude that the centralization effect is uniform: war causes domestic centralization, not decentralization. This conclusion is refuted by the history of hegemonic wars, however. Although some hegemonic wars caused domestic centralization, others caused domestic *decentralization*.

#### 2. A new view of the domestic effects of war

The domestic effect of hegemonic war depended on the economic structure of the international spillover problem that caused the war. In the cases where the international spillover problem *increased* the productivity of military labor relative to civilian labor in the state suffering the spillover effects, the war caused domestic *centralization*. But in cases where the international spillover problem *decreased* the productivity of military labor relative to civilian labor in the state suffering the spillover effects, the war caused domestic *decentralization*.

This chapter presents the logic behind these causal mechanisms and derives their observable implications. The next two chapters present case studies that test the implications against the historical evidence.

---

<sup>1</sup>Huth 1996. Gleditch 2002. Gibler 2010.

### 3. Domestic institutional restructuring

When states engage in international security competition, the primary function of domestic institutions is to allocate manpower between the civilian and military sectors and to tax the civilian workers to compensate the military workers. A shift in the productivity of military labor relative to civilian labor requires a reallocation of labor and a change in the tax rate. The direction of the shift in productivity determines the direction of the labor reallocation and the change in the tax rate. This, in turn, determines the domestic pressures that the government must manage and the type of institution needed to manage these pressures.

If relative military productivity *increases*, then domestic constituents will demand a transfer of their civilian manpower into military use. The government must *organize and control* this transfer. So it will become more *centralized and authoritarian* in nature. Domestic constituents will accept a higher tax rate in return as long as the government exercises its increased authority as they have demanded. By contrast, if relative military productivity *decreases*, then domestic constituents will oppose a transfer of their civilian manpower into military use. And they will oppose a tax hike to fund any such transfer. The government will have to *compensate* them to make them willing to support a transfer. So government institutions will become more *decentralized and contractual* in nature.

#### 3.1 Secession wars

In a secession crisis, there is a *decrease* in the productivity of military manpower relative to civilian manpower (as explained in Part 2 of this study). This decrease is shown in figure 5.1 in the top right panel. The marginal product schedule for military labor shifts down (solid, boldfaced line). The optimal army size decreases from  $A2$  to  $A2s$ . To the right of point  $A2s$ , the military productivity schedule is lower than the civilian productivity schedule. It would not be rational to transfer men from civilian to military use to the right of that point, because their productivity would be relatively lower in military use. Therefore, constituents in other regions of the country will *oppose* transferring their manpower from civilian to military use to fight the secession.

Military productivity does increase once the state gains some military options to combat

the secessionists (dashed, boldfaced line). But the *actual proceeds* of this increase only begin to flow *after* the central government has recaptured some territory from the secessionists. For only then can it collect tax revenues from that territory again. Before any territory has been retaken, however, constituents in other regions of the country do not know whether military action will succeed. Hence they prefer to keep their manpower in its current civilian use – where its productivity is already clear. If the government wants to attract their support for military action, therefore, it must offer to give them the *future* proceeds of military action *now*. It must bid for their support with the only assets it has to give away: the rights to retain tax revenue and the right to control policy.

**Implication 1s.** In a secession war, the central government will bid for the support of domestic constituents by offering to give them shares of its governance rights (e.g. the rights to receive tax revenue and control policy).

**Implication 2s.** Governance rights will get shared out among the key providers of military manpower in the fight against the secessionists. Domestic institutions will become decentralized and contractual in structure.

### 3.2 Depression wars

In a depression crisis, by contrast, there is an *increase* in the productivity of military manpower relative to civilian manpower. This increase is shown in figure 5.1, middle right panel. The marginal product schedule for civilian labor shifts down (solid, boldfaced line). The optimal army size increases from  $A2$  to  $A2d$ . In the range between  $A2$  and  $A2d$ , the military productivity schedule is higher than the civilian productivity schedule. It would not be rational to keep this segment of the population in the civilian economy, because their productivity would be relatively higher in military use. Therefore, constituents will *favor* transferring their manpower into military use to combat the depression.

Even before the state takes any military action, domestic constituents prefer to have their civilian manpower transferred into military use. So the government does not have to do anything to attract their support for military action. Quite the opposite: Domestic constituents must attract a leadership that favors military action by offering to give their

*current* manpower in return for a leader's promise of *future* military action. They must bid for military leadership with the only assets they have to offer: their labor, their loyalty and their political and civil rights.

**Implication 1d.** In a depression war, constituents will bid for strong leadership that promises to transfer manpower from civilian to military use. They will give the leadership loyalty and the right to exercise authoritarian control over their lives, even at the expense of political and civil rights they previously safeguarded.

**Implication 2d.** Governance rights will become concentrated in a small number of executive agencies. These agencies will have the authority to transfer men into military use and to control the remaining men in their civilian uses. Domestic institutions will become more centralized and authoritarian in structure.

#### 4. Cleavages in domestic politics and society

The relationships among domestic constituents are determined by the direction of the demand for military labor. Is it the government that must bid for the support of domestic constituents, or the constituents who must bid for strong government leadership? In a secession war, the central government must bid for the constituents' support. So each constituent complains that the other constituent received a higher bid. And each constituent demands to receive a higher bid for himself, or else he will withhold his support from the government. In a depression war, by contrast, the domestic constituents must bid for strong government leadership. So each constituent complains that the other constituent did not bid high enough. And each constituent demands that the other one must raise his bid, or else be penalized by the government.

**Implication 3s.** In a **secession** war, domestic constituents will be divided over how many *new* rights each one has *received* from the government. Each constituent will claim that he has not *received* enough.

**Implication 3d.** In a **depression** war, domestic constituents will be divided over how many *old* rights each one has *conceded* to the government. Each constituent will claim that the other has not *given* enough.

## 5. Impact of military successes and failures on domestic institutions

Changes in military fortune cause shifts in the domestic distribution of bargaining power. The nature of the shift depends on the primary axis of contention in domestic politics. In a decentralized and contractual system, the axis of contention is between the central government and the provincial governments and military contractors. Military success induces the central government to try to regain some of the governance rights it traded away earlier to gain provincial support. Military failure induces the provincial governments and contractors to demand more concessions of governance rights from the central government.

By contrast, in a centralized and authoritarian system, the axis of contention is between different agencies of the central government. Military success induces the central government to increase the authority of the agencies that already had authority at the expense of those that did not. Military failure induces the central government to transfer authority from the agencies that already had it to others that lacked it – or to new agencies created specifically to reverse the military setback.

**Implication 4s.** In a **secession** war:

- a. If the war goes well, the central government will regain the revenue rights and policy control rights that it traded away earlier.
- b. If the war goes badly, provincial contractors will gain more revenue rights and policy control rights at the expense of the central government.

**Implication 4d.** In a **depression** war:

- a. If the war goes well, the agencies that wielded relatively more authority will increase in authority at the expense of others that had less.
- b. If the war goes badly, the agencies that wielded less authority will gain authority at the expense of those that had more.



## 6. Impact of military successes and failures on domestic cleavages

Changes in military fortune also influence the cleavages between domestic constituents. Military success mitigates the cleavages, while military failure intensifies them. The impact on domestic politics, in turn, depends on the direction of demand for military labor. In a decentralizing system, military failure intensifies the conflict between provincial governments over their shares of governance rights. This causes further decentralization. In a centralizing system, by contrast, military failure intensifies the pressures among domestic constituents to prove their loyalty and their obedience to the central government. This causes further centralization and a deepening of authoritarian control.

**Implication 5s.** In a **secession** war:

- a. If the war goes well, friction between the provincial governments over their shares of the country's governance rights will decrease. They will become more tolerant of differences in their respective shares. They will be less insistent on decentralizing power and more accepting of the central government's current power.
- b. If the war goes badly, friction between the provincial governments over governance rights will increase. They will become less tolerant of differences in their respective shares. They will become less accepting of the central government's power and more insistent on decentralizing power.

**Implication 5d.** In a **depression** war:

- a. If the war goes well, frictions between domestic constituents over burden sharing will decrease. They will exert less pressure on each other to prove their loyalty to the government and make sacrifices for the national interest. Society will become more tolerant of differences and less insistent on authoritarian obedience.
- b. If the war goes badly, frictions between domestic constituents over burden sharing will increase. They will exert more pressure on each other to prove their loyalty to the government and make sacrifices for the national interest. Society will become less tolerant of differences and more insistent on authoritarian obedience.

## 7. Impact of military successes and failures on domestic policy conflicts

### 7.1 Secession wars

The central government must pay the provincial governments and military contractors in advance to attract them into combat against the secessionists (as explained above). Once they are being paid enough, their incentive is to continue fighting and expanding militarily as long as the central government continues to pay them. Moreover, military success *increases* the productivity of military labor relative to civilian labor, thereby *increasing* the degree of military expansion that is optimal. This magnifies the provincial contractors' incentive to continue expanding militarily as long as the central government continues to pay them. In fact, their incentive is to oversell the domestic government on how much military expansion is optimal, simply to continue being paid to expand. Thus the central government is caught on the horns of a dilemma. On one hand, it must pay the contractors enough to attract them into the fighting. On the other hand, it must restrain their narrow interests in over-expansion in order to secure the national interest in optimal expansion.

**Implication 6s.** If the war goes well, there will be a policy conflict within the central government between those who want to indulge the widening ambitions of the provincial military contractors in order to retain their military support, and those who want to restrain the military contractors at the risk of losing their military support.

**Implication 7s.** If the war goes badly, there will be a policy conflict within the central government between those who want to grant the military contractors' demands for more governance rights to retain their military support, and those who want to deny their demands at the risk of losing their military support.

### 7.2 Depression wars

Domestic constituents must offer their labor and loyalty to the government to attract a leadership that favors military action (as explained above). Once the military leadership

is in place, its incentive is to pursue military expansion as far as the domestic constituents originally demanded – and for as long as the constituents continue supporting it. Military success *decreases* the productivity of military labor relative to civilian labor, thereby *decreasing* the degree of military expansion that is optimal. This reduces the government’s incentive to expand militarily. Indeed, military successes turn the original expansion goal demanded by the domestic constituents into an “over-buy” (in retrospect). And the more military successes are achieved, the more of an over-buy it becomes. Thus the government is caught on the horns of a dilemma. It must pursue military expansion as far as the domestic constituents originally demanded simply to retain their political support. Yet expanding that far is really an over-expansion that is contrary to the national interest. The government cannot say that outright, however, because it might sound like it is breaking its original promise to the domestic constituents. And this might destroy the popular political support needed to sustain the war effort at all. Thus the government and the constituents are caught in their mutual over-buy and oversell. The government must continue to over-supply the military expansion that the constituents originally over-demanded.

**Implication 6d.** If the war goes well, there will be a policy conflict between:

- a. those in the majority (of both the government and the domestic constituents) who want to retain the original ambitions for military expansion to maintain popular political support for the government, and
- b. those in the minority (of both the government and the constituents) who want to restrain the government’s ambitions at the risk of being unpopular domestically.

**Implication 7d.** If the war goes badly, there will be a policy conflict between:

- a. those in the majority (of both the government and the domestic constituents) who want to intensify the war effort to salvage the original vision of expansion – despite increasing evidence that it is impossible to achieve, and
- b. those in the minority (of both the government and the constituents) who can see the writing on the wall and seek to expose the original vision of expansion as delusional –despite the weight of majority opinion/delusion against them.

## CHAPTER 28

### Case study: Domestic politics in Austria in the Thirty Years War

The war was caused by a secession crisis in the Austrian Habsburg monarchy.<sup>1</sup> The central government was Catholic as were most of its noble supporters. The nobility also included large numbers of Protestants, however. They formed a majority in some provinces such as Bohemia and upper Austria. The Protestant nobility feared the central government would try to re-catholicize them. It had already done so in lower Austria by restricting titles of nobility and administrative positions to Catholics. In 1618, the nobility of Bohemia declared independence from Vienna and chose as their new king a Protestant prince from the neighboring state of Lower Palatinate. The nobility of upper Austria quickly joined the secession movement. The province of lower Austria – around Vienna – had a more even mix of Protestants and Catholics. The Catholics chose to remain loyal to the Habsburg ruler, Ferdinand II, while the Protestants chose to join the rebellion. Ferdinand knew that he had to stand firm against the secessionists or else his kingdom would fall to the Protestants.

The Catholic noble estates in lower and inner Austria remained loyal to Ferdinand. But they did not volunteer to send military support to fight the secessionists (**implication 4s**). Ferdinand had an army of his own but it was too small relative to the task at hand.<sup>2</sup> To whom could he turn for support to form a larger army? What would they demand in return for their support? He turned to those whose support he needed the most and to those who had the most support to offer. In both cases, those to whom he turned demanded material

---

<sup>1</sup>Sources that provide evidence that supports the model's implications are cited specifically. Otherwise the narrative follows the general accounts contained in: Asch (1997), Barker (1982), Bireley (1981, 1994, 2003), Chesler (1979), Fichtner (2003), Gindely (1884), Heilingsetzer (1991), Limm (1984), MacHardy (2003), Melton (2007), Pages (1970), Parker (1997), Polisensky (1971), Schramm (1991), Steinberg (1966), Wedgwood (1999).

<sup>2</sup>Pages 1970: 55-57, 60; Parker 1997: 46; Steinberg 1966: 37.

concessions in advance before agreeing to give their support (**implication 1s**). He met their demands variously by trading away his government's rights to levy taxes, set policy and intervene in provincial government affairs (**implications 2s, 3s**). He turned first to the estates of lower Austria because this province was geographically central to his kingdom. They demanded the right to practice Protestantism freely and the right to control provincial government in the way they had done under Ferdinand's predecessor. He granted these demands.<sup>3</sup> He then turned to his fellow Catholic rulers, the Duke of Bavaria and the King of Spain, because they too had to stand firm against the secessionists or their kingdoms might fall to the Protestants. The Duke offered military support but demanded the right to rule upper Austria and retain its tax revenues until Ferdinand could return payment for the support. The Spanish king also offered military support but demanded that his ambassadors in Vienna have the right to be consulted on all Austrian policy decisions. In practice this amounted to a veto right that gave Spain effective control over Ferdinand's policies.<sup>4</sup>

Military victories gave Ferdinand opportunities to restore and fortify his rule. On one hand, victories allowed him to take back some of the most cherished rights he had traded away earlier to gain military support (**implication 6.1s**). On the other hand, victories allowed him to award his military supporters other rights that were less valuable to him in comparison (**implication 3s**). He and his allies defeated the secessionists of upper Austria and Bohemia in 1620. Soon thereafter he took back the right of the lower Austrian nobility to practice Protestantism.<sup>5</sup> The victory in Bohemia also enabled him to award Catholic nobles there the estates confiscated from Protestant nobles who had joined the secession movement.<sup>6</sup> He awarded many of those estates to Albert of Wallenstein and, in 1623, named him the Duke of Friesland within Bohemia. By this point, Ferdinand had been elected Emperor of the Holy Roman Empire (which comprised all of Germany as well as the Austrian Habsburg lands).

Ferdinand and his allies then defeated the Protestant state that had supported the Bo-

---

<sup>3</sup>Bireley 1981: 9; Bireley 1994: 41; Fichtner 2003: 35-36; MacHardy 2003: 73-74.

<sup>4</sup>Bireley 1981: 5-6; Fichtner 2003: 35; Pages 1970: 58-59, 61-62, 66; Steinberg 1966: 38-39, 42.

<sup>5</sup>Barker 1982: 9; Bireley 1981: 30-45; Bireley 1994:45-47; MacHardy 2003: 75; Melton 2007: 178.

<sup>6</sup>Barker 1982: 8; Fichtner 2003: 35, 37-39; Limm 1984: 17, 21-22; Melton 2007: 174-175; Pages 1970: 74; Polisensky 1971: 141-144, 182; Steinberg 1966: 41.

hemian secession (the Lower Palatinate) and moved into central Germany. He awarded his main ally, the Duke of Bavaria, *de facto* control of the Lower Palatinate in return for his military support. This award posed a dilemma for Ferdinand and his advisers (**implication 5s**). While they knew the award was necessary to retain the Duke's support, they suspected that it might upset the balance of power among the Catholic principalities of the Empire.<sup>7</sup>

Ferdinand's main ally in Germany, the Catholic league, now expected a counter-offensive by the Protestant states on Germany's northern borders. So the league members advised him to form a larger army. A few years earlier – at the time Wallenstein had received the Duchy of Friesland – Wallenstein had proposed to form and lead an imperial army. Ferdinand had refused out of concern that Wallenstein's territorial gains would upset the internal balance of power in the Habsburg lands and the Empire.<sup>8</sup> But now Ferdinand gave his approval and contracted loans to fund the new imperial army. Denmark now entered the war in defense of the Protestants (as the Catholic league had predicted). The Catholic imperial army under Wallenstein defeated Denmark and advanced into northern Germany. These victories, in turn, allowed Ferdinand to make further territorial awards and reclaim more of the governance rights that he had traded away earlier for military support. He awarded Wallenstein the territory of Mecklenburg on the Baltic Sea and named him the Duke of Mecklenburg. Ferdinand then took back his own province of upper Austria from his own ally, the Duke of Bavaria, and awarded him formal title to the Lower Palatinate in compensation.<sup>9</sup>

The Duke suspected that his title to the Lower Palatinate would not survive a peaceful settlement of the religious conflict in Germany. So he pressured Ferdinand to pursue maximal aims in Germany: the restitution of all Catholic Church properties that had been secularized or alienated to the Protestants since 1552 and the right of Catholic rulers to dictate their subjects' religion. In response to this pressure and that of his religious counselors, Ferdinand issued the Edict of Restitution proclaiming exactly these policies.<sup>10</sup> The Edict provoked

---

<sup>7</sup>Pages 1970: 77-78, 82-83; Steinberg 1966: 42.

<sup>8</sup>Pages 1970: 101.

<sup>9</sup>Bireley 1981: 23, 64-65; Pages 1970: 104-107; Polisensky 1971: 172-174; Steinberg 1966: 45-48, 51-52.

<sup>10</sup>Asch 1997: 94-97; Bireley 1981: 85-94; Pages 1970: 111; Steinberg 1966: 50-51.

strong opposition from the Protestant states within and beyond Germany, of course. To counter their opposition, Ferdinand decentralized the governance rights of the Empire even further (**implication 6.2s**). He gave Wallenstein greater authority to raise troops, levy taxes and set policy in the areas of Germany under his control. Although Wallenstein did command a large army, he did not run any regular institutions for the administration of justice, taxes or local government (beyond his own Duchies of Friesland and Mecklenburg). His authority was thus dictatorial but not governmental in the sense of organized or centralized administration.<sup>11</sup>

This problem caused intense debate and strife within the Catholic camp for two reasons. On one hand, there was a sense that Wallenstein had received more than his fair share of the governance rights won in the war (**implication 8s**). On the other hand, there was a fear that Wallenstein's heightened authority would upset the balance of power among the Catholic principalities within Germany (**implication 5s**). The pressure group opposing Wallenstein was led by Ferdinand's chief ally himself, the Duke of Bavaria. By emphasizing the imbalances and risks created by Wallenstein's unruly state within a state and his ambition to enlarge it, this group convinced Ferdinand to dismiss the overmighty general and disperse his army.<sup>12</sup> This move showed that military successes could enable the central government to take back some of the governance rights it had traded away earlier to obtain military support (**implication 6.1s**). Yet, at the same time, this move created a power vacuum within Germany that invited Sweden to invade to defend the Protestants from the threat posed by the Edict of Restitution. At this point, the war expanded into a truly international struggle among all of the great powers of Europe. The conflicting claims already staked would prolong the conflict for another 18 years.

---

<sup>11</sup>Bireley 1981: 65; Pages 1970: 112; Polisensky 1971: 177; Steinberg 1966: 52.

<sup>12</sup>Asch 1997: 98-100; Bireley 1981: 47, 117-118; Limm 1984: 24; Melton 2007: 192-193; Pages 1970: 105-106; Polisensky 1971: 177-178, 184; Steinberg 1966: 52-53.

## CHAPTER 29

### Case study: Domestic politics in Japan in World War II

The war was caused by Japan's pursuit of economic self-sufficiency through military expansion.<sup>1</sup> Japan was reliant on China for imports of raw materials and export markets for finished goods. As early as 1926, disagreements over the terms of that trade and the security of Japanese investments produced tensions between the countries. The tensions increased in 1927 when the Japanese economy suffered a severe recession. An even deeper recession in 1930 eliminated the market for Japanese silk in the United States – making the Chinese market even more important. The Japanese military had long felt the need for an international military strategy – given the nature of the modern export economy – and a comprehensive mobilization strategy – given the nature of modern warfare. The military's plans for institutional and budgetary growth to support these strategies had been blocked by civilian politicians prior to 1927. With the recessions of 1927 and 1930, however, there arose a huge domestic constituency for the military's goal of economic self-sufficiency through military expansion. This tipped the balance of domestic political forces in favor of the military's policy platform. The domestic politics of the late 1920s and 1930s is a long story of the consummation of this marriage between domestic constituents impoverished by recession and political-military leaders offering a centralized solution. The story raises fundamental questions about the relation between domestic politics and the national interest.

Would this coalition have formed without the domestic constituent demands created by the recessions? Was the formation of this coalition a victory of special interests over the

---

<sup>1</sup>Sources that provide evidence that supports the model's implications are cited specifically. Otherwise the narrative follows the general accounts contained in: Barnhart 1987, Benson and Matsumura 2001, Berger 1977, Crowley 1966, Hanneman 2001, Iriye 1997, McClain 2002, Nish 1997, Nish 2002, Peattie 1975, Thomas 1996, and Waswo 1996.



national interest or a necessary shift in the domestic coalition structure to advance the national interest? Once the coalition formed and government institutions were restructured to mobilize more military manpower, how did leaders know the optimal amount of manpower to extract from domestic society? Did the creation of the new institutions in itself answer this question? Or might the institutions be used to over-extract men and material in a way that hindered the national interest rather than advancing it? By what metrics and methods did leaders try to distinguish between these two possibilities? Were differences of opinion between leaders due to parochial interests or simply to different understandings of the national interest?

The model presented earlier implies that the coalition never would have formed without the domestic political demands created by the twin recessions. The coalition was not a victory of special interests but a necessary shift in coalition structure to advance the national interest. While the restructuring of domestic institutions was necessary to extract the optimal amount of military manpower to promote the national interest, the new institutions in themselves would not automatically choose the optimal amount. There was a risk of over-extraction that at least some leaders might perceive and try to avoid. Yet differences of opinion between leaders were not necessarily signs of parochial interest. Different understandings of the national interest would be sufficient to explain such disagreements. The model predicts the two main understandings of the national interest over which leaders would disagree. To test whether these are reasonable answers to the above questions, let us compare the model's implications to the historical record.<sup>2</sup>

The military leaders of Japan had learned about the need for a national mobilization strategy years earlier – during World War I. They lobbied civilian leaders to pass laws to create institutions that would have the capacity to mobilize the country for modern warfare. However, civilian governments blocked or reversed such initiatives in 1918, 1922 and 1924. The most the military could achieve was to establish planning offices that could issue guidelines but lacked the power to enforce them and mobilization offices that had

---

<sup>2</sup>Crowley (1966: 278-279) argues that Japan's foreign policy in the 1930s was not the result of factionalism at odds with the national interest, but rather a result of the sincere pursuit of the national interest by groups who simply differed in their conceptions of it. My model explains how and why they disagreed.

powers that could only be exercised during full-scale wars.<sup>3</sup>

This balance of political forces began to change with the onset of financial crisis and recession in 1927. Domestic constituents demanded that the government increase its authority to protect and restore the economy (**implication 1d**). Constituents were dissatisfied with liberal democracy and its centrifugal forces which seemed to be causing and prolonging the economic downturn.<sup>4</sup> When the economic crisis brought down the civilian government, military leaders bid for domestic political support by offering stronger, more centralized leadership (**implication 2d**). A retired general assumed political power by accepting the offices of prime minister and foreign minister. Once in power, his cabinet increased the government's authority to solve the financial crisis, to control the population and to engage in military expansion in China.<sup>5</sup> The cabinet also began to concentrate government authority in a smaller number of agencies that had the right to transfer resources into military use by authoritarian methods (**implication 3d**). The army minister established a Cabinet Resources Bureau that revolutionized mobilization planning by extending its scope from resources that were of direct military use to all of the nation's resources. This included all goods and services produced in the civilian economy.<sup>6</sup> Although the agency did not immediately receive the authority to enforce its plans, its creation was the first in a series of steps by the leadership to build this authority.

The foreign territory of most interest to military and economic planners was Manchuria. Japan had enjoyed various mining and economic rights there since 1915. The local warlord had been cooperative in safeguarding these rights for many years. With the recession in 1927, however, the Japanese felt a need to secure these rights and make more infrastructure investments to enable higher production and greater security for the Japanese economy. At the same time, these goals became threatened by the Chinese Nationalist movement which was gaining strength and moving its forces toward Manchuria. Japanese military leaders fell into two camps in addressing this problem. One camp thought economic relations with

---

<sup>3</sup>Barnhart 1987: 23-25; Peattie 1975: 12-15; Waswo 1996: 82-84.

<sup>4</sup>Benson and Matsumura 2001: 38, 42; McClain 2002: 414; Waswo 1996: 88-89, 97-98.

<sup>5</sup>Benson and Matsumura 2001: 38; Hanneman 2001: 28, 38; McClain 2002: 395; Nish 2002: 58; Iriye 1997: 60.

<sup>6</sup>Barnhart 1987: 25.

Manchuria could be secured peacefully through diplomatic agreements. The other camp thought Manchuria should be taken under Japanese Imperial control to resolve the problem definitively. At first the diplomatic camp appeared to win the debate when it reached an economic agreement with the Manchurian warlord. The Chinese Nationalists refused to recognize the agreement, however, so the warlord backed away from it. As the recession deepened in 1928, the Japanese prime minister ordered military action to neutralize the warlord. In the event, he was assassinated first by a Japanese military officer who was associated with the movement for economic self-sufficiency through military expansion. The dead warlord's successor was not interested in cooperating with Japan, of course. So the camp in Tokyo that preferred to assert imperial control over Manchuria stepped to the fore.<sup>7</sup>

The episode showed that the younger army officers were strongly in favor of military expansion to restore the nation's economy and were prepared to force matters at the front line – ahead of their rear commanders' schedules and even wishes. Most of the young officers came from the rural agricultural regions of the country. In these areas, individual loyalty to the community and its needs was instilled from birth and, in turn, military service was given the highest social approbation. The recession was causing widespread unemployment and impoverishment in these communities. The young officers felt that the policies of the civilian politicians and business leaders were prolonging this distress and preventing the military spending needed to alleviate it through territorial expansion. Others in the military felt this way too. The young officers were simply in the vanguard.<sup>8</sup>

Once these young officers were on the front lines, it seems, their main way to balk at the slow pace of military expansion was to push forward at a faster pace – even if this meant insubordination. Their impetuous actions may have suggested to the rear commanders that their authority had limits if it were not used to satisfy the young officers' demands for military expansion to restore the economy. Of course, the commanding officers in the rear needed the younger officers at the front to heed military authority or else the commanding officers would have to go forward and do the front line jobs themselves! For this reason the rear officers probably felt a need to acquiesce to the young officers' *fait accompli* at least to

---

<sup>7</sup>Barnhart 1987: 29-31; Nish 2002: 60; Iriye 1997: 61.

<sup>8</sup>Benson and Matsumura 2001: 41-42; Hanneman 2001: 38-39; McClain 2002: 413-415.

a point. The assassination of the Manchurian warlord marked the beginning of a repeated pattern of insubordination by young officers on the front lines and acquiescence by senior officers in the rear. The senior officers did not always take steps to punish the insubordinate officers afterward. In any case, the preference of the young officers for military expansion to restore the economy became national policy.<sup>9</sup>

The onset of economic depression only intensified the young officers' preferences and increased the pressures on the top leadership. From 1929 to 1930, the bottom fell out of the economy. Silk exports fell by 45%. Silk had constituted almost 20% of total farm production in the mid-1920s. Now farm incomes were dropping precipitously – by a total of 66% from the two downturns combined.<sup>10</sup> Rural communities were devastated by famine and, by some reports, cannibalism. In the face of this crisis, the military leaders of the economic self-sufficiency movement saw Manchuria not only as a source of industrial raw materials but also as a site for cheap food production for the Japanese home market. Planning to take Manchuria under Japanese control was intensified. The planners in Tokyo were again divided into two camps. The first camp wanted to try diplomacy first to prod the Chinese Nationalists into cooperating and cow the great powers into standing by. The second camp wanted to use military force from the start. The first camp won a decision to delay military action for a year to buy time for more diplomacy with the Chinese and more domestic political efforts to build a consensus for military action. A member of the second camp immediately told some young officers in Manchuria about this decision. They quickly engineered a threat to Japanese-controlled railway lines in Manchuria that appeared to justify immediate military action. By early 1932, Japan had conquered Manchuria and asserted control over its economy and government.

The first strategic effect of the Manchurian campaign was to provoke the Chinese Nationalists and Soviet Russia in a way that caused fears in Tokyo that the war would widen. This prospect immediately caused two policy conflicts within the central government. First, the leaders most responsive to the demands of impoverished domestic constituents wanted a larger army than those focusing on the long-run strategy of restructuring the national

---

<sup>9</sup>The argument in this paragraph about the contract enforcement function of military insubordination is my own.

<sup>10</sup>Hanneman 2001: 38; Kindleberger 1986: 140.

economy for self-sufficiency (**implication 4d**). The first group, led by the army minister, proposed a large increase in force levels and material support programs to prepare for a probable conflict with the Soviets. The second group, led by the foreign and finance ministers, objected that such a large military buildup would stall the nation's economic recovery.<sup>11</sup> This conflict, in turn, created a second policy conflict over two different kinds of mobilization. The group that wanted a relatively larger army wanted a *fast and steep* mobilization geared to the current war effort – supported by a *shortsighted* institutional restructuring program designed to serve military needs during a war. The other group wanted a *slow and gradual* mobilization geared to the future national economy that would be constructed from foreign conquests – supported by a *farsighted* institutional restructuring designed to serve a restored economy after the war (**implication 5d**).<sup>12</sup>

The shortsighted faction led by the army minister won the debate at first. It pushed through a big rearmament program to gear the country for war. Within a few years, however, the new status quo in Manchuria had become stable and successful enough to enable the farsighted faction to switch the focus from shortsighted military goals to farsighted economic goals (**implication 6.1d**). In early 1934, this faction forced the army minister to resign, took key posts and strove to set the nation on a path of farsighted economic and political restructuring. It lobbied for three goals: (1) scaling back the rearmament program, (2) reorienting the military mobilization and the institutions supporting it to support a long-run strategy, and (3) shifting to more diplomatic methods in relations with the Chinese Nationalists to pursue three aims: (a) to gain a period of peace in which to implement the long-run strategy, (b) to secure Manchuria's southern flank in case of a conflict with the Soviets down the line, and (c) to secure access to the natural resources and markets of North China (just south of Manchuria).<sup>13</sup>

The rearmament program and its institutions could only be scaled back modestly in the short term due to opposition from the shortsighted faction in the army and navy (as well as to the irreversibility of the programs already begun). But the medium term programs

---

<sup>11</sup>Barnhart 1987: 33-35; Crowley 1966: 152-153, 202.

<sup>12</sup>Barnhart 1987: 33-34; Berger 1977: 93-99; Crowley 1966: 254; Iriye 1997: 65-66; Peattie 1975: 186-188.

<sup>13</sup>Barnhart 1987: 35-40; Berger 1977: 100-102; Crowley 1966: 200-210, 255-256, 276-279; Peattie 1975: 188-190.

for mobilization and institutional restructuring were redesigned to support the farsighted strategy. And a diplomatic effort was launched to establish a Japanese sphere of influence in North China and, at the same time, stabilize relations with the Chinese Nationalists to avoid conflict over that sphere. As long as this effort could maintain peace with the Chinese, the farsighted strategy could succeed.

This strategy bore significant fruit over the next few years in two areas. First, Japan created autonomous zones in North China that served its needs for a military buffer zone and secure access to resources and markets without hindrance from the rest of China. This process was not without difficulties and crises that portended future conflict – due in part to the intimidation tactics of the Japanese. Yet the farsighted faction that led these efforts from Tokyo prioritized peaceful resolution of disputes with the Chinese wherever possible and maintaining a low military profile in North China.<sup>14</sup> Second, this faction designed and implemented a long-run plan to strengthen the economy and centralize the government's control over both the economy and the military mobilization process (**implication 2d**). The plan was intended to restructure the Japanese economy from the ground up for self-sufficiency, maximum national income and long-run dominance in East Asia. Since the depression had taken hold in 1930, rural agricultural regions had remained devastated and restive. Plans were made to create new industries that would employ some of the idle agricultural workers and generate domestic purchasing power for the remaining workers' produce.<sup>15</sup> These industries – to produce aluminum, iron, magnesium, natural gas, and petroleum substitutes – would form the strategic core of both a civilian economy oriented toward export competitiveness and a military economy oriented toward self-sufficiency through foreign conquest. The plan would require a revolutionary increase in central government control over the economy (**implication 2d**). Yet because the economy desperately needed resuscitation, civilian politicians and business leaders approved it (**implications 1d, 3d**).<sup>16</sup>

This period marked the high point of the farsighted faction's influence in the government and impact on the nation's economy and military strategy. In early 1936, a group of young

---

<sup>14</sup>Crowley 1966: 210-219; Nish 1977: 210-212; Peattie 1975: 269-270.

<sup>15</sup>Barnhart 1987: 67-70, 77; Crowley 1966: 256-257; Peattie 1975: 210-213.

<sup>16</sup>Barnhart 1987: 73-74; Peattie 1975: 214-216.

army officers belonging to the shortsighted camp staged a coup d'état that almost succeeded. They were reacting against what they perceived to be government foot-dragging in the campaign for economic self-sufficiency through military expansion. Simultaneously, the Chinese front began to destabilize. By 1936 a series of disputes and incidents had brought strength, unity and ambition to the Chinese Nationalists who sought to free North China, and perhaps even Manchuria, from Japanese domination. The farsighted faction in Tokyo believed that it could reassure the Nationalists and insure stable economic relations with all of China by *reducing* Japan's political control over North China. At first they prevailed over the shortsighted faction – which advocated a policy of deeper entrenchment and confrontation. Yet they failed to solve the underlying problem of securing the natural resources of North China that were essential to Japan's expanding economy. As the Chinese Nationalists became bolder in confronting Japanese forces and civilians on the ground, it became increasingly difficult to believe that diplomacy, reassurance and accommodation could secure Japan's vital interests.

The deterioration in the military outlook reawakened the policy conflicts between the shortsighted faction and the farsighted faction (**implications 6.2d and 7**). The shortsighted faction pressed hard for mobilization of a larger army that would be capable of solving the China problem by force immediately. The farsighted faction, which was still in charge of the key bureau whose approval was required for the mobilization, was caught on the horns of a dilemma. Approval of the shortsighted mobilization plan would divert so many men and resources from the farsighted mobilization plan as to paralyze it. As efforts to resolve each confrontation with the Chinese peacefully appeared on the brink of success or failure, the farsighted faction either canceled or approved (respectively) the myopic mobilization of troops needed to force a solution by arms.<sup>17</sup>

Eventually it became clear that a military campaign would be necessary. The shortsighted faction predicted that the campaign would be short and low in cost. The farsighted faction predicted that the campaign would be long and so costly as to jeopardize the long-run strategy of restructuring for national self-sufficiency. This faction proved right. Although the Chinese Nationalists were defeated in each battle, they regrouped and fought back after

---

<sup>17</sup>Barnhart 1987: 86-90; Crowley 1966: 291-299; Peattie 1975: 273-275, 284-286.

each defeat rather than surrendering and making peace as the shortsighted faction had hoped. Japanese military commanders demanded and received more troops and material to expand further into China and knock the Nationalists out for good. With this backing, the Japanese forces kept winning battles and territory, but at the cost of eviscerating the Japanese economy at the structural level. Whole industrial sectors were reoriented from civilian production for both export and domestic markets to military production for the war in China. The fall in exports was exacerbated by domestic price inflation that was induced by higher government demand for all kinds of goods. At the same time, the fall in production for domestic markets meant that more goods had to be imported from abroad – just as foreign exchange was becoming scarcer due to the fall in exports! Even some key military goods needed for the war in China had to be imported from abroad due to the reorientation of Japanese industry toward other key military goods. In short, the whole productive, trade and financial structure of the economy was moving *away* from self-sufficiency rather than *toward* it – to pay for a war that was being fought ostensibly to secure self-sufficiency.<sup>18</sup>

This dilemma brought the policy conflicts in the central government to a new head with even higher strategic stakes. The shortsighted faction wanted to salvage the original goals of military expansion by intensifying the war effort, increasing the demands on domestic society, and centralizing the government further. The farsighted faction wanted to abandon those goals out of fear that further centralization and intensification of the war effort would make the over-extraction problem even worse and ruin the long-run strategy for good (**implication 7d**). The shortsighted faction wanted to press for total victory against the Chinese, while the farsighted faction preferred peace even if the cost were high.<sup>19</sup> The Prime Minister brought in a new foreign minister from the farsighted faction to negotiate a peace settlement with the Chinese. At this point, the Soviets entered the fighting and the shortsighted faction insisted on fighting back. This response simply escalated the war on a new front and encouraged the Chinese to escalate on the old front. These developments increased the demand for Japanese troops yet again and deepened the over-extraction problem even further. No one in the Japanese leadership could see a way out of the dilemma.

---

<sup>18</sup>Barnhart 1987: 109-110.

<sup>19</sup>Barnhart 1987: 111-112; Crowley 1966: 372-377; Nish 1977: 224-226; Thomas 1996: 217-218.



Fighting on the Soviet front ended quickly due to shortages of key materials on the Japanese side. But the shortsighted faction in Tokyo won the debate over whether to seek total victory over the Chinese. This decision made the future course of the war almost inevitable. For each side's decisions to escalate left the other side with no other option but to escalate back. In the quest for total victory over the Chinese, the Japanese launched a new offensive to capture the city of Hankow. The steep requirements in men and material led planners in Tokyo to conclude that drastic new sacrifices would be needed from Japanese society and the economic drain would render the long-run strategy dead in the water. Yet the city of Hankow was taken by Japanese forces nonetheless. The Chinese responded by opening a southern supply route from Burma. This would enable them to continue fighting indefinitely if nothing were done to close the route down. The only way for Japan to cut off the route and salvage its chances for a victory in the China campaign was to invade Indochina. With this move Japanese ambitions appeared to be unlimited. The United States felt that its only way to contain Japan – short of war – was to place an embargo on oil and fuel sales to the nation. Japan then had no choice but to invade the Dutch East Indies for its oil resources and, knowing that war with United States was inevitable, attack Pearl Harbor in the hope of knocking out the United States at the start.

Part VIII

From the path of power back to the  
path of law: Settlements of hegemonic  
wars

## CHAPTER 30

### Stability and instability in hegemonic war settlements: Theory and cases

Grand strategy is the art and science of waging war in a way that will produce a sustainable peace. In the study of grand strategy, there has been much focus on the hegemonic wars waged by the great powers in Europe from 1500 to 1950. Many studies have been written about these wars. Not as many studies have been written about the peace settlements that followed these wars. Only a few studies have attempted a comparative analysis of these settlements as a group. None of these studies offer a metric of stability or an explanation of the comparative stability of the settlements. In this study I offer both.<sup>1</sup>

I define *instability* as the prevalence of unilateral asset seizures that do not provoke an immediate response of a reciprocal or retributive kind.<sup>2</sup> Stability is the relative absence of unilateral asset seizures. Using this metric, three of the seven settlements of hegemonic wars were stable. The other four settlements were unstable. I argue that the difference is explained by anchor based governance. The stable settlements were based on *anchor contracts* – agreements whose terms were invariant to shifts in relative power among the signatories. The unstable settlements were based on *variable term contracts* – agreements whose terms varied with shifts in the relative power of the signatories. The aim of the case

---

<sup>1</sup>The political science literature on the settlements of hegemonic wars is small. Most of the studies that focus on these settlements take an historical approach. Some studies try to answer deeper causal questions by focusing on a single settlement (Gulick 1957; Kissinger 1957). In not making comparisons across settlements, however, these studies fail to discern causal mechanisms that are common to the settlements of long and costly wars. Some studies attempt comparison using the historical method (Oslander 1994; Lesaffer 2004). Because they do not provide a rigorous theory based on rational microfoundations, however, they fail to distinguish between root causes, subsidiary causes and epiphenomenal factors. The most important effort to use theory to answer causal questions about the settlements of hegemonic wars is Ikenberry's study of the settlements of 1815, 1918 and 1945 (Ikenberry 2001). I review and critique that study in the last section of this chapter.

<sup>2</sup>Conflict is the prevalence of unilateral asset seizures that *do* provoke an immediate response of a reciprocal or retributive kind.

studies in this chapter is to present the evidence that supports this conclusion.

This conclusion suggests the need for a new definition of grand strategy.

A grand strategy is a combination of a compellance strategy, a deterrence strategy and a contracting strategy that is designed to bring the enemy to settlement terms in a sustainable way.

This definition highlights the fact that a good contracting strategy can make the difference between stability and instability. It also highlights the fact that even the best compellance strategy may fail to produce stability in the long run if it is not accompanied by a good contracting strategy. Consider an historical example. Following the Glorious Revolution in 1688, England fought off France for over 20 years to defend the Protestant Succession to the English throne. The Peace of Utrecht in 1713 was not based on an anchor contract, however. Over the next three decades England had to fend off three French invasion gambits and wage a constant campaign against domestic subversives supported by France. England may have had an optimal compellance strategy in the wars against France from 1688 to 1713. But the success of that strategy was nearly undone by the lack of a good contracting strategy. This example suggests that studies of grand strategy should devote more focused attention to the contracting element in the grand strategy equation. The case studies in this chapter are intended to make a start on that work.

In sections 1 through 5, I provide five brief case studies of hegemonic war settlements.

## **1. The Peace of Westphalia (1648)**

The primary cause of The Thirty Years War was Habsburg Austria's drive to defend Catholic lands from the spread of the Protestant religions – and the Protestants' drive to defend their lands from Catholic armies.<sup>3</sup> While the peace treaties of Munster and Osnabruck (1648) ended the fighting, the tensions between the opposing camps remained high. Each camp feared that the religion of the opposing camp would continue to spread into its lands and populations by nonviolent means. Since a large percentage of the German territories had

---

<sup>3</sup>In writing this case study I relied on a number of secondary sources. In addition to the sources cited in the text, I also relied on: Polišenský (1971b), Asch (1997), Croxton (1999), and Wedgwood (1999).

mixed populations, contention and disputes between Catholics and Protestants continued to arise. Local co-existence depended on some form of cooperation to reduce tensions and resolve disputes.<sup>4</sup>

The treaty negotiators reasoned that the main way to prevent a resurgence of Habsburg Austrian power into Germany was to craft a settlement that would stabilize relations between the Catholic and Protestant confessions in Germany. The settlement provided for stability in two ways. It amended the constitutional, political and legal structures of the Holy Roman Empire to enable religious disputes to be settled by recourse to law and courts rather than arms and war.<sup>5</sup> And it gave every territorial ruler in Germany – of which there were hundreds – the rights to declare war, raise an army and form alliances within and beyond Germany.<sup>6</sup>

One of the primary sources of tension between Catholics and Protestants was the competition for control of church buildings. The peace settlement aimed to avoid conflict over these assets primarily through the use of anchor terms: Clauses that fixed the distribution of an asset so it was *invariant* to shifts in relative power between the confessions. Control of church buildings was distributed simply by choosing an historical date and awarding control to the confession that had possessed control on that date. The date might have to be chosen carefully. Yet once chosen, it was invariant to shifts in relative power.

”One of the ways treaties brought security was by freezing the status quo.... [L]aws and treaties used the notion of the status quo to fix where groups could worship.... The Peace of Westphalia established 1624 as the normative date (*Normaljahr*) for the empire: wherever Calvinists or Lutherans had worshiped in officially Catholic territories as of that date, they could continue or resume doing so, and vice versa for Catholics in Protestant territories. If the worship had been public, so it could be in future: if only private, then it had to remain so.... In fact, determining any ‘status quo’ was an act of creative interpretation, especially when

---

<sup>4</sup>Whaley 2012b: 322-5. “[C]oexistence... does not of itself guarantee mutual respect and harmony. It did, however, tie Catholic and Protestant territories into an institutional framework within which a degree of cooperation was essential.” (324)

<sup>5</sup>Thompson 2009: 52-3. “[I]t would be shortsighted to argue that the confessional issue had simply disappeared. The settlement brought confessional conflict within a legal framework. Confessional disputes were no longer settled by artillery and infantry but by advocates and supposedly incisive legal arguments.”

<sup>6</sup>Croxton and Parker 2009, pp. 81–5.

the normative date was sometime in the past. If the fortunes of the confessions had undergone any dramatic swings, choice of date was crucial. To restore the supremacy of the Reformed faith in the Palatinate, negotiators at Westphalia agreed to set 1618, before the territory was occupied by a Spanish army, rather than 1624 as the normative year there. For the Reformed in Lutheran territories and vice versa, they set 1648.”<sup>7</sup>

While this approach did not eliminate each camp’s fear of encroachment by the other’s religion, it did prevent each camp from expanding its share of *this material asset* – church buildings – at the other’s expense.

Tension between Catholics and Protestants arose not only over churches, but also over government offices, public budgets and policymaking authority. Anchor terms were also used to reduce these tensions:

”At the local level, rules such as the *Normaljahr* principal [and the parity principal]... were designed to provide *criteria* for resolving disputes.... [T]he confessional balance was translated into constitutional and political parity. Catholics and Protestants were given equal weight on city councils; all public offices were simply occupied by a Catholic and Protestant simultaneously....”<sup>8</sup>

Other methods of achieving confessional parity included the *simultaneum* rule – sharing a church building – and the rule of alternating succession – first a Catholic bishop and then a Protestant bishop, and so on.<sup>9</sup> Again, these anchor terms did not eliminate each camp’s zeal to expand its followership and defend its faithful from the other. Yet such terms did prevent each camp from expanding its share of *key material assets* – government offices, public budgets and policymaking authority – at the other camp’s expense.

By preventing each side from expanding its share, moreover, such terms eliminated each side’s *incentive* to increase its relative power in order to expand its share. The value of anchor terms was described in this way by an historian of religious peace:

---

<sup>7</sup>Kaplan 2007, p. 220.

<sup>8</sup>Whaley 2012b, 324-5. Italics added.

<sup>9</sup>Kaplan 2007: 203, 216-7. Whaley 2012b: 325.

”[P]arity could also take the form of separate institutions – separate, equal in number, and receiving equal funds from government.... Parity in general, and full-blown, numeric parity in particular, had several great virtues. First, it...imposed rigid, narrow parameters on conflict. Second, it eliminated uncertainty. *Like treaty clauses defining a status quo, parity froze into place a specific state of affairs [e.g. a specific distribution of a material asset] that could be defined precisely and policed fairly easily. Numeric parity created a balance of power between the confessions [e.g. a specific distribution of a material asset between them] that did not depend on their relative numbers in the population, their wealth, princely patrons, or any other factor.... By ruling out any change in the balance of power, parity ensured the confessions they would never lose their current assets and resources, or be able to seize their opponents’.* Finally, and most obviously, numeric parity ensured that neither confession could get its way simply by outvoting the other. No law could be passed or decision taken without the consent of members of both [confessions].... Proportional division of offices did not have these virtues. Even though it guaranteed a minority representation in government, it did nothing to prevent the majority from tyrannizing over the minority. Such power on one side and vulnerability on the other were not conducive to peace.”<sup>10</sup>

Why, exactly, is it a virtue to stabilize the distribution of a material asset so that it is invariant to shifts in relative power? While this historian is not specific, there is an answer from another source. Nicholas Rowe (1989, 1990) says that the virtue of a stable distribution is that it eliminates each party’s incentive to invest in raising its relative power, and thereby eliminates competitions in raising relative power.

While tensions continued to flare after the settlement of 1648, there were no more religious wars in Germany.

”Disputes over confessional matters continued at almost every level until the end of the Reich. Many involved low-level violence and the abuse of power.... [T]he perceived threat on both sides was intensified by the print war that each

---

<sup>10</sup>Kaplan 2007: 224-5. Italics added.

conducted against the other. Each side accused the other of serious breaches of the Peace of Westphalia.... [Yet] the constitutional structures of the Reich... contributed to the stabilization of the situation.... Again and again, these issues ended up at the Reichstag, the regional Kreis assemblies or in the Imperial courts and were ultimately resolved either by judgment in favor of one party or by compromise, or were diffused through such long-winded discussions and procedures that the conflicting parties simply lost interest or the issue ceased to be relevant.... [O]utright confrontation never materialized.”<sup>11</sup>

In sum, the settlement of the Thirty Years War was based solidly on anchor contracts. And the settlement was stable in the long run.

## **2. The Treaties of Utrecht and Rastatt (1713/14)**

In the 1680s King James II of England embarked on a campaign to restore the preeminence of the Catholic religion in England.<sup>12</sup> The Protestant majority in parliament objected and eventually forced James to abdicate the throne and flee to France. Since European monarchs ruled by the law of divine right, the effective deposition of a sitting king by his own parliament appeared to be a violation of this law and, as such, to pose a threat of similar insubordination by parliaments in neighboring monarchies. King Louis XIV of France felt especially threatened, because his own aristocracy included a "reformist" faction that was maneuvering at the heights of government power to force Louis to delegate and divest his royal power. The deposed and agrieved James was harboring in France under Louis's protection. Together these Catholic monarchs posed the threat of a restoration of Catholic power to the English throne by military invasion.

France had just instigated a minor war on its eastern frontier with Germany. But the situation was transformed by the abdication of James in London and the vote of the English Parliament to invite the Dutch head of state to replace him as the new king of England.

---

<sup>11</sup>Whaley 2012b: 322-3.

<sup>12</sup>In writing this case study I relied on a number of secondary sources. In addition to the sources cited in the text, I also relied on: Clark (1954), Clark (1970), Carter (1968), J. Jones (1968), Gibbs (1969), Clark (1970), Langford (1976), Nordmann (1976), J. Jones (1980), and McKay and Scott (1983).



One of Parliament's goals in issuing the invitation was to secure Dutch military assistance in the fight to protect the English throne – and the Protestant succession to the throne – from French Catholic power. In accepting the English Parliament's invitation, one of the Dutch sovereign's goals was to bring England into the war against France in order to secure Germany and the Dutch Republic from a French invasion.

With England and the Dutch Republic in the war against France, the stakes were high on both sides. The English Parliament perceived its capacity to protect the country from royal despotism to be at stake. The French monarch perceived his prerogative to rule by divine right to be at stake. The war lasted from 1688 to 1697 and later became known as The Nine Years War.

After five years of inconclusive fighting, negotiations opened between France, the Dutch Republic and England and continued regularly for the next four years. King Louis proved willing to make concession after concession, both to peel away subsidiary allies from the coalition and to induce its main members – the Dutch and English – to settle. There was one concession he would not make, however. He would not renounce James' claim to the English throne, acknowledge the legitimacy of the transfer of royal power in England, or recognize the Protestant succession to the English throne. This evidence of Louis' priorities offers proof that he was fighting for the divine right of kings more than any other issue. Additional proof comes from Louis' support of military invasions of England to restore James to the throne.

After the Nine Years War ended in 1697, neither Louis nor James was willing to renounce James' claim to the English throne. The English perceived their intransigence as an ominous and potentially endless threat to the security of the Protestant succession in England.

Four years later the War of the Spanish Succession erupted, pitting France and Spain against Austria and a coalition of German princes. King William of England immediately perceived a threat to the nation, but the English Parliament expected only a war between Austria and France over Italian territories – one that would pose no real threat. Even after France invaded the southern Netherlands and formed a dynastic union with Spain that multiplied its power greatly, King William could not obtain Parliament's support to enter the war. This would require two more direct threats to English interests.

First, Louis enacted trade restrictions against Dutch and English commerce in Italy and South America. Second, when James II, the deposed English King, died in exile in France, King Louis of France responded (against the advice of counselors) by naming the deceased monarch's son, James III, the rightful heir and King of England. Again the political foundation of Louis' divine right to rule in France was at stake.

"This was a logical, indeed a necessary move [for Louis to make], since James's right to the English throne was juristically and ideologically identical with that of Louis's son and grandson to the throne of Spain. Having (successfully) asserted the latter, Louis was bound to uphold the former. But in doing this Louis certainly did not intend to give James direct assistance to recover England, nor does he appear to have had any idea of how this recognition of [James] the Pretender (an infringement of the spirit, if not the letter, of the Ryswick Treaty) would help William in his management of English opinion – but that was a technique that Louis had never had to acquire or use. The effects were immediate."<sup>13</sup>

Together, Louis' two hostile actions induced the English Parliament to support England's entrance into the war against France. Prominent in the parliamentarians' minds was the need to secure the Protestant succession from French power.

For England, the War of the Spanish Succession ended with the Peace of Utrecht in 1713. Yet the Protestant succession in England remained vulnerable. James's son persisted in maintaining his claim to the English throne. More threateningly, James' Catholic supporters in England, the Jacobites, persisted in pursuing their goal of a restoration of Catholic royal power in England. The "Jacobite threat" continued to be a destabilizing factor in Anglo-French relations for decades following the peace settlement of 1713.<sup>14</sup> In 1715 the Jacobites staged a domestic rebellion in support of a military invasion by James to take the throne.

"[T]he French attitude was ambiguous: Louis had abjured the pretender at Utrecht, but this did not prevent him providing covert encouragement and even support to the cause. Loose tongues at the French court exulted in the expected

---

<sup>13</sup>J. R. Jones 1978, p. 288.

<sup>14</sup>Lord 2004.

triumph of Jacobitism. Thus in mid-1715 the French envoy to Sweden suggested to a then sceptical Charles XII of Sweden that he should support an invasion by the Pretender [James III].”<sup>15</sup>

After the Jacobite rebellion of 1715 was put down, Jacobites ”swarmed across Europe armed with letters of introduction from James III.”<sup>16</sup>

The English nobles who had supported the Jacobite rebellion were impeached. The leaders of the impeachment proceedings made clear that the problem, in their view, was the failure of the British diplomats who had negotiated the Treaty of Utrecht to structure the settlement properly. The treaty had left the Catholic Pretender James III, his Jacobite ”conspirators”, and France with both the motive and the opportunity to shift the balance of power in their favor at England’s expense.

”The Jacobite rebellion of 1715...showed...how shallow the roots of the new dynasty were, and the extent to which foreign powers could take advantage of British domestic divisions. As the articles of impeachment against the seven rebel peers who had supported the rebellion made clear, it also reflected the failure of British diplomacy at Utrecht. ’The dissolution of the late glorious confederacy against France,’ the argument ran, ’and the loss of the balance of power in Europe, were further steps necessary to complete the designs of the said conspirators.’ ’The same being effected by the late ignominious peace with France,’ it went on, ’the French king was rendered formidable, and the Protestant succession was thereby brought into the most imminent danger.’ Britain’s first line of defense, in other words was where it had always been: in Europe. Jacobitism and ’universal monarchy’ were two sides of the same coin. They could only be dealt with through a series of interlocking diplomatic and domestic measures.”<sup>17</sup>

---

<sup>15</sup>Simms 2007: 101. Lord writes: ”James also needed men, arms and money from France. The French were lukewarm towards him – they did not want war, and would not venture anything that would threaten the peace with Britain.... Although there were officers from France and Northumberland, they did not seem to take any part in directing affairs.” (Lord 2004: 66, 74.)

<sup>16</sup>Simms 2007: 102, 142.

<sup>17</sup>Simms 2007: 102.

In other words, the "designs" of the Jacobite conspirators were to attack England with French military support and restore the Pretender James to the English throne. They would have gotten nowhere if the Utrecht Treaty had settled the dispute between England and France over the English succession properly. For in that event France would have had no motive to give military support to the Jacobites and they, in turn, would have had no opportunity to threaten England. The Utrecht treaty had failed to settle the dispute properly, however. So it left the English succession to become simply a function of the balance of power between England and France. This is why the threat of a Jacobite rebellion at home and the threat of French power abroad were "two sides of the same coin". The unsettled nature of the succession issue meant that a shift in the balance of power in France's favor could flip the issue her way. Hence she had an incentive to jockey for a power advantage against England – and she did.

George I of Hanover had just become the King of England. Now he faced a threat to his German territories from Sweden. In search of allies to counter the threat, he brought England into an alliance agreement with France. It required France to expel James III to Italy and acknowledge the Protestant succession in England.<sup>18</sup> Yet the alliance was in the end simply a power aggregation tool that served each state's interest in not falling behind in the continental game of balance of power politics. France still hoped and looked for opportunities to restore the Catholic Stuarts to the throne in England. Strategists in London therefore remained wary of French power.

It was not long before the next Jacobite threat emerged. By 1718 Spanish expansionism in the Mediterranean was threatening British commercial interests. When Britain launched a naval attack to unseat Spain from Sicily, it led to war. To threaten Britain at home, Spain coordinated an invasion of Scotland with a domestic Jacobite rebellion aimed at deposing King George and returning the English throne to James. Although the uprising and the invasion were defeated by government forces,<sup>19</sup> it reminded Protestants and Whigs throughout England that the Jacobite threat was still real – and it was a foreign threat mediated by the international balance of power as much as, if not more than, a domestic threat.

---

<sup>18</sup>Simms 2007, pp. 108–15.

<sup>19</sup>Lord 2004: 124–5.

Although France fought on Britain's side in the war, French power was ambivalent. And King George's chief strategist knew it. After peace returned, he defended the war to Parliament by claiming that Spanish power had to be checked or else it might reunite the Spanish and French thrones. This scenario was not farfetched because the Spanish king at this time was a [great]nephew of the French king.

"Stanhope...made clear that [the British attack at] Cape Passaro had been an act not merely of tactical but also strategic preemption. It was aimed, first of all, at stopping Spain from perhaps even reuniting the French and Spanish thrones."<sup>20</sup>

"[W]ith the Austrians only a short sail away across the Channel [in the Austrian Netherlands], the development of Ostend as a naval hub seemed to portend another avenue of Jacobite infiltration.... The threat from the Pretender [James III and his Jacobite supporters] was strategic rather than domestic in origin; it waxed and waned in relation to Britain's position within the state system."<sup>21</sup>

In 1730 the English Jacobites made another plan with James III, now in Italy, to restore James to the throne of England. "...France...was going to back the plans with an invasion...."<sup>22</sup> When the English government uncovered the plot, France withdrew its support. Yet again, it was clear that the Jacobite and French threat was real.

In 1738 the English Jacobites began planning another domestic rebellion and coordinating with the Scottish and French to launch a military invasion in support.

"From early 1740 onwards...Jacobitism had to be reckoned a real threat if supported by Franco-Spanish naval power. Worst fears were confirmed in September 1740 with news of a French squadron setting sail for the West Indies in order to deter British attacks on Spanish forces there."<sup>23</sup>

---

<sup>20</sup>Simms 2007: 187-8.

<sup>21</sup>Simms 2007: 187-8.

<sup>22</sup>Lord 2004: 138.

<sup>23</sup>Simms 2007: 279.

By 1743 the French were amassing an invasion force on the channel. The plan was aborted after the English government discovered it and the French invasion fleet was scattered by a storm. Again in 1745 the French assembled an invasion force and the English Jacobites launched a domestic rebellion that spread widely before it was eventually defeated by government forces.<sup>24</sup>

### **3. The Vienna Settlement (1815)**

After the defeat of Napoleon, the governments of Europe convened in Vienna to craft a formal settlement of the war. The victorious powers – Russia, Prussia, Austria and Britain – faced three main problems. First, they wanted to ensure that French military power would not expand into Europe again and trigger another round of war. To solve this problem, they agreed to reduce French territory within narrower borders, insert an occupying military force into France for a few years, and establish strong buffer states along France's eastern borders from the English Channel to the Mediterranean.<sup>25</sup> Second, each of the four victorious powers wanted to ensure that the other three were satisfied with their territorial gains from the settlement, so none of them would have an interest in unilateral expansion that might trigger another round of war. To solve this problem, the four powers adopted a negotiation procedure whereby they alone would determine the distribution of territories in the final settlement and simply inform the other states at the Congress of their decisions. The resulting distribution enabled Prussia, Austria and Russia to regain the size and strength they had lost under the Napoleonic hegemony. Prussia was restored to its former size in northern Germany and given new territories in the lower Rhine region to strengthen its military posture toward France. Austria was restored to its previous position of predominance in Italy. And Russia was given a position of dominance in eastern Poland.<sup>26</sup>

The third problem addressed at the Congress of Vienna was the most difficult to solve. The victorious powers were concerned that the French ideology of republicanism might

---

<sup>24</sup>Lord 2004: 187-92.

<sup>25</sup>Dakin 1979: 25, 29-31.

<sup>26</sup>Dakin 1979: 14-6, 27-9.

provoke another revolution in France, inspire revolutions against monarchical governments throughout Europe, and trigger another round of domestic instability and international war. The French government shared this concern because Napoleon's final defeat had resulted in the restoration of the Bourbon monarchy in France.<sup>27</sup> While the Vienna settlement of 1815 did not solve this problem, it established a set of political principles and negotiating procedures for solving it in the future. The victorious power agreed to convene periodically and, in the event of a crisis, to act in "concert" to insure peace and stability.<sup>28</sup> Thus the Concert of Europe was born. A few years later France was admitted to the Concert by the Treaty of Aix-la-Chapelle.

Among the five concert powers, a disagreement emerged as to the best way to prevent revolutions against monarchical governments in Europe.<sup>29</sup> The three eastern powers – Russia, Prussia and Austria – were absolute monarchies. Lacking constitutions, they perceived any reform movement in any state in Europe as a threat to their domestic regime stability – even a movement with moderate goals that aimed simply to reform an absolute monarchy into a constitutional one. Hence they thought the best way to prevent a wave of revolutions was to allow no concessions to any domestic opposition movement in *any* state. By implication, their preferred response to any kind of liberalizing movement abroad – be it a revolution, a revolt or just an accommodation of liberals – was to crush the movement and restore the

---

<sup>27</sup>Webster 1925, p. 54; Schroeder 1992, pp. 695–6.

<sup>28</sup>Dakin 1979: 30. Webster 54.

<sup>29</sup>Heydemann 2002.

original regime.<sup>30</sup>

Their view was not shared by the Western powers in the concert. Britain, in particular, thought the best way to avoid a *radical* revolution was to make concessions to *moderates*. Hence Britain's preferred response to a liberalizing movement in any state was to encourage the moderates to accept limited concessions from the reigning regime, such as settling for a transition from absolute monarchy to constitutional monarchy. Britain was wary of giving the eastern powers the right to intervene in any state at will to crush a domestic reform movement.<sup>31</sup> A society without any right to express grievances was more likely to experience a violent revolution and cause an international war as far as the British were concerned.<sup>32</sup>

This disagreement between the Eastern and Western powers was highlighted by the numerous revolts and revolutions in Europe between 1820 and the early 1830s. None of these crises triggered a great power war, however. Nor did the revolutions that swept Europe in 1848. What was the key to stability and peace?

According to the Contractual Realist theory, the settlement of 1815 produced peace and stability for decades because it was an anchor contract. Recall that an anchor contract is one whose terms are tied to anchors in the real world that do not move even when relative power shifts among the signatories of the contract. The main anchor in the Vienna settlement was

---

<sup>30</sup>Sked 1979.5: 99, 113-4. The conservative powers were averse to the formation of constitutional monarchies because it meant granting representative institutions. They felt the monarch should not be restrained by such institutions, but only by his advisors, the aristocracy, and his respect for the social order, the law, and the civil service and judiciary that enforced his will. Metternich, the Austrian Chancellor and the main exponent of the conservative powers' view, advocated

"pure monarchy rather than the constitutional brand.... [At the same time, he] was no supporter of arbitrary or absolute monarchy. A sovereign had to be advised and had to follow orderly government practices. He had to uphold the law and social order and, of course, the laws of God. If he were not to lose respect he would have to demonstrate his capacity to exercise justice as well as authority. In practice this meant that there should be a powerful sovereign whose wishes, formulated in consultation with his advisers, should be enforced by a centrally organized judicial and civil service." (Sked 1979.5: 99-100.)

To capture the conservative powers' view, I employ the standard distinction between constitutional monarchy and absolute monarchy, rather than Sked's distinction between constitutional monarchy and pure monarchy, simply for ease of expression.

<sup>31</sup>Sinnreich 2009, pp. 150-2.

<sup>32</sup>Austria's Metternich also recognized that the worst excesses of absolute monarchy must be reformed and advised the Italian rulers to reform. But they ignored the advice because they knew that Austria was committed to saving them from domestic opposition movements that posed a real threat. (Sked 1979.5: 115-6.)



the term "legitimate monarchical government". This term referred to an anchor in the real world in each state in Europe: the sitting monarch and his actual authority to command obedience from his subjects. This anchor did not move even when relative power shifted among the concert's members. Hence no member had an incentive to increase its relative power at the expense of another, because doing so would not shift the terms of the settlement in its favor. In other words, doing so would not gain it a greater share of the political assets in the European system.

After the Vienna settlement was reached in 1815, a number of crises arose from the eruption of a domestic revolution or reform movement in a smaller state. Each crisis forced the great powers to clarify the terms of the Vienna settlement and apply them in practice. With each new crisis the mutually agreed meaning of the terms became clearer and enabled the powers to resolve the crisis short of war. The terms specified each power's rights and obligations with respect to intervention in another state's domestic politics:

1. Fixed obligation clause. Each power has the *obligation not to foment domestic political change* in another state – in opposition to its legitimate monarchical government – *where such change was not occurring already*.
2. Variable right clause. Each power has the *right to support domestic political change* in another state – in opposition to its legitimate monarchical government – *where such change was occurring anyway* – but the *obligation not to take advantage* of it in search of excess gains.
3. Variable right clause. Each power has the *right to prevent domestic political change abroad* if it is necessary to avert the threat of contagion and domestic political instability at home.

The purpose of the fixed obligation clause was to prohibit each power from committing intentional, opportunistic interference in another state's domestic politics to reap unilateral gains. This clause protected every state in Europe from the threat of a domestic revolution or counter-revolution fomented by a great power. The anchor in this clause – legitimate monarchical government – was the key to the protection, because this term referred to a domestic distribution of political authority that was *stationary*. The anchor would not move

even when relative power shifted among the concert's members. Hence no member had an incentive to boost its relative power in order to foment domestic political change in another state. In other words, wherever a regime had enough domestic legitimacy to retain its hold on power against its domestic opponents, the fixed obligation clause protected it from great power opportunism to foment regime change.

In return for committing to this fixed obligation, each great power gained variable rights. It had the right to support regime change in another state *if* such change was occurring anyway due to the regime's domestic failings, loss of domestic legitimacy, and incapacity to integrate domestic opponents. In other words, whenever the anchor of legitimate royal authority moved on its own due to random forces – unintentionally on the part of any great power – then a great power was permitted to support regime change in that state.

The purpose of this variable right clause was to recognize that sometimes the anchors moved due to random forces beyond any great power's control. Sometimes a state experienced a domestic political change for purely domestic reasons (e.g. the regime's domestic failures, loss of domestic legitimacy, and inability to integrate domestic opponents). This happened both in absolutist regimes that had become too conservative and reformist regimes that had become too liberal. In either case, domestic political change might be permitted by the great powers, because permitting it would not give any great power an incentive to foment such change *de novo* through intentional, opportunistic interference in another state's domestic politics.

Finally, the purpose of the other variable right clause was to enable each great power to avert net losses in its own domestic political system due to contagion from political change abroad.

The main function of the contract was to place *limits* on the great powers' right to intervene in the domestic politics of smaller states. The limits applied to *both* the conservative and the liberal powers. Austria's Chancellor Metternich recognized this point:

"Intervention, however, was to be used only, as [Metternich] saw it, as a weapon in the service of legitimacy; it was not to be used either to abet the revolution

or to upset the balance of power.”<sup>33</sup>

By adhering to this contract, each power could solve the problem of domestic political change abroad in the way that it preferred without encroaching on the core interests of the other powers. For the liberal Western powers – Britain and France – this meant:

1. not initiating movements to create constitutional monarchies in the states of Central and Eastern Europe;
- 2a. supporting the emergence of constitutional monarchies in Western Europe in cases where it was led by domestic reform movements;
- 2b. not taking advantage of their emergence for excess gains; and
3. intervening to prevent conservative reactions in Western Europe from turning constitutional monarchies back into absolute monarchies.

For the conservative Eastern powers – Russia, Prussia and Austria – this meant:

1. not initiating movements to restore absolute monarchies in Western Europe;
- 2a. supporting the restoration of absolute monarchies in Central and Eastern Europe in cases where it was led by domestic absolutists;
- 2b. not taking advantage of their restoration for excess gains; and
3. intervening to prevent liberal revolutions in Central and Eastern Europe from turning absolute monarchies into constitutional monarchies.

With these lane markers painted on the highway of great power relations, the great powers were able to solve the problem of domestic political change in a way that suited all of them well enough. Not perfectly, but well enough. In Western Europe, where the liberal powers had the strongest interest in their preferred solution, constitutional monarchies emerged in Portugal, Spain and Belgium.<sup>34</sup> In Eastern Europe, where the conservative powers had the

---

<sup>33</sup>Sked 1979.5: 119.

<sup>34</sup>Bullen 1979. ADD.

strongest interest in their preferred solution, absolute monarchies were preserved or restored in Naples, Lombardy-Venetia, Hungary, Poland and the German Confederation.<sup>35</sup>

To sum up, the combination of these three clauses – the fixed obligation clause and the variable rights clauses – created a self-enforcing contract among the great powers. The contract was self-enforcing as long as each great power's gains from receiving its variable rights – *plus* its savings from not having to go to war to punish violators of the contract – exceeded its costs of compliance with its fixed obligation.

### 3.1 Contract enforcement

How did the great powers enforce the contract in practice? Did they use power-based enforcement – threatening to go to war against an impending violator? Or did they use anchor-based enforcement – threatening to deny the violator the benefits of membership in the concert of Europe. The historical record shows that each method was used at various times. There were instances where one or more powers had to threaten another power with war to deter it from violating the contract. There were other instances where benefit-based enforcement *alone* – without threats of war – was sufficient to induce compliance with the contract.<sup>36</sup>

### 3.2 Comparison to other theories of the Concert's effectiveness

A number of theories have been offered to explain why the settlement of 1815 enable stability and peace across Europe for almost forty years. These theories reflect the full range of conventional paradigms of explanation in the field: realism,<sup>37</sup> neoliberal institutionalism,<sup>38</sup> constructivism<sup>39</sup> and social-theoretic analysis<sup>40</sup>. However, none of these theories identifies the causal mechanism that is proposed in the present study under the heading of *Contractual Realist theory*. Moreover, each of these studies suffers from logical and empirical problems.

---

<sup>35</sup>Sked 1979.5: 112-7, 121. Sked 1979.8: 177.

<sup>36</sup>Schroeder 1992, pp. 699–700.

<sup>37</sup>Gulick 1955; Slantchev 2005.

<sup>38</sup>Jervis 1982; Jervis 1985; Lipson 1994; Richardson 1999; Ikenberry 2001.

<sup>39</sup>Charles A. Kupchan and Clifford A. Kupchan 1991; Cronin 1999.

<sup>40</sup>Mitzen 2013.

The Contractual Realist explanation avoids these problems.

#### 4. The Treaty of Versailles (1919)

Economists have shown that international trade is vulnerable to interstate competitions in raising relative power to affect market prices. A government can make domestically produced goods more attractive to domestic consumers by enacting trade tariffs that raise the price that domestic consumers pay for imported goods. When foreign firms lose sales as a result, they respond by lowering their selling prices to recoup some of the losses. The decline in their selling prices means, in effect, that they receive less favorable terms of trade due to the tariffs on their exports.

”[W]hen a government imposes an import tariff, some of the cost of this policy is shifted to foreign exporters, whose products sell at a lower world price (i.e., at less favorable terms of trade). This temptation to shift costs naturally leads governments to set unilateral tariffs that are higher than would be efficient.”<sup>41</sup>

When other governments impose tariffs in retaliation, all nations are made worse off. Such competitions over relative pricing power are mutually self-defeating.

Governments can avoid such competitions by constructing international trade treaties based on the principles of reciprocity and nondiscrimination. These clauses neutralize the price effects of each state’s tariffs on the other states’ exports. When a trade treaty includes these clauses, therefore, no government can improve the terms of trade for its domestic firms by increasing tariffs. Since no government can improve its terms of trade by raising tariffs, no government has an *incentive* to raise tariffs.<sup>42</sup>

”[T]rade-policy decisions are inefficient if and only if governments are *motivated* by their abilities to change the world price.... [M]utual changes in trade policy that conform to the principle of reciprocity leave world prices [the terms of trade]

---

<sup>41</sup>Bagwell and Robert W. Staiger 1999, 215-6. See their citations to earlier economic literature demonstrating this point.

<sup>42</sup>Bagwell and Robert W. Staiger 1999; Bagwell and Robert W. Staiger 2002.

unchanged.... [Hence] the principle of reciprocity neutralizes the terms-of-trade externality that underlies inefficient behavior... [so the principle] can be efficiency enhancing.”<sup>43</sup>

The principles of reciprocity and nondiscrimination are *anchors* in the sense defined in this study. They anchor the terms of trade between countries by making the terms invariant to changes in relative bargaining power among states. For example, even if domestic politicians in one state build a protectionist coalition at home that increases its bargaining power in international trade negotiations, this cannot improve the state’s terms of trade because the terms are stabilized by the anchors of reciprocity and nondiscrimination. Hence domestic politicians have no incentive to boost relative bargaining power by building protectionist coalitions at home. And competitions between countries in raising relative bargaining power are avoided.

From the 1850s to 1914, international trade became liberalized to a significant extent. During World War I, however, the great powers enacted tariffs to raise revenue. Following the war, the Versailles Settlement contained only a few economic clauses. With respect to trade, Germany alone was required to grant unconditional most-favored-nation status to the Western powers (e.g. the principle of nondiscrimination). This clause lapsed in 1925, however, and Germany adopted the same tariff strategy that the Western powers had followed during and after the war. The powers competed for market shares by reaching preferential trading agreements within their spheres of influence and imposing tariffs on products from beyond their spheres. By the mid-1920s it was already clear that the overall effect was mutually self-defeating: foreign demand for each state’s products was declining and could not be restored through bilateral tariff reductions.<sup>44</sup>

Despite the tensions between Germany and France in the interwar period, they managed to achieve some cooperation on trade issues through international production cartels. Yet their agreements were not based on the principle of nondiscrimination. On the contrary, the goal was to defend and expand their trading interests at the expense of major competitors.

---

<sup>43</sup>Bagwell and Robert W. Staiger 1999, 217. Emphasis added.

<sup>44</sup>Feinstein, Temin, and Toniolo 1997, pp. 74–6, 101–2; Charles P. Kindleberger 1989, pp. 161–70; Simmons 1994, pp. 37–8.

As an historian of the cartels wrote:

”They cooperated in defending their common [domestic] cartels’ interests against the USA, and they cooperated in forcing Great Britain into the steel cartels.”<sup>45</sup>

In sum, after the Versailles Settlement of 1919, international trade was *not* based on the principles of reciprocity and nondiscrimination. When the depression took hold in the early 1930s, therefore, it provoked a competition in raising relative bargaining and pricing power to garner market share. The result was mutual tariff hikes and a competition to establish exclusive trading blocks that eventually led to war.

## **5. The settlement of World War II**

Following the Second World War, international trade was reestablished on the basis of the principles of reciprocity and nondiscrimination. These principles were enshrined in the General Agreement on Tariffs and Trade (GATT) – a key pillar of the postwar settlement. As a result tariff wars were avoided even in times of recession and competitions to establish exclusive trading blocs have not broken out.

---

<sup>45</sup>Schroter 1996, pp. 135–6.

Part IX

**Contractual Realism resolves the  
inter-paradigm debates**



## CHAPTER 31

### **Contractual Realism explains the causal drivers of the conventional paradigms**

Scholars once agreed that the discipline of international relations theory was divided into three paradigms for theory building and theory testing. Scholars no longer agree on this point. Some scholars believe the discipline is still divided and the inter-paradigm debates continue. Other scholars believe the debates are over and the discipline has moved on. Among scholars who believe the debates continue, some scholars believe it is good that they continue – and they should always continue because each paradigm captures a truth about international relations that will always be essential for theory building and theory testing. Other scholars believe it is good that the debates continue because this may lead to their resolution – and this too will be good. Other scholars believe it is bad that the debates continue, however, because they are no longer useful. According to these scholars, each paradigm captured a truth that remains essential for theorists to use, but there is no need to wage or settle an inter-paradigm debate. For according to the doctrine of "theoretical pluralism", each of these truths can be used for its best purpose without undermining the other truth-claims. These scholars suggest that the discipline should simply focus on the task of answering explanatory questions rather than getting lost in philosophical debates that are unnecessary or unresolvable.

The main goal of this chapter is to show that the inter-paradigm debates continue and the main theory presented in this study – Contractual Realism – resolves these debates.

## 1. Overview of the argument

Each of the conventional paradigms emphasizes a main point. And each point is valid. But each paradigm makes its point in the wrong way. It adopts assumptions that appear to be necessary to make its point – and maintains those assumptions to ensure that its point is not lost. Yet each paradigm’s maintained assumptions are not actually necessary to make its point. The same point can be made from alternative assumptions. These alternative assumptions are made explicit by contractual realism.

Each of the conventional paradigm’s has its maintained assumptions. And each one’s maintained assumptions are logically inconsistent with the maintained assumptions of the other two paradigms. These inconsistencies persist to this day (despite claims from some quarters that they are overdrawn or nonexistent). Contractual Realism provides a single set of logically consistent assumptions that succeeds in making all three of the main points of the three conventional paradigms – and thereby resolves the inter-paradigm debates.

The inter-paradigm debaters raised a series of questions over the years . Contractual Realism answers these questions differently than the conventional paradigms did. In the course of this chapter, I will show that each of the questions falls into one of these categories:

1. The question poses a real debate between two of the conventional paradigms. The existing literature claims to have resolved the debate, but its putative resolution is faulty. Either the putative resolution can be refuted, the grounds for debate persist, and Contractual Realism resolves the debate. Or the putative resolution lost sight of the core message of one of the two conventional paradigms, the core message needs to be retrieved to answer outstanding explanatory questions, and Contractual Realism retrieves it.
2. The question poses a real debate between two of the conventional paradigms. The debate is still unresolved by the existing literature. And Contractual Realism resolves it.
3. The question poses a specious debate between two of the conventional paradigms. It is not really a debate at all, because the opposing positions are not actually inconsistent on a logical or empirical level.

4. The question was not raised by the inter-paradigm debaters, but it should have been raised because the conventional paradigms disagree over it. And Contractual Realism resolves the disagreement.

In distinguishing these questions, I provide further demonstration that Contractual Realism resolves the inter-paradigm debates and thereby enables new answers to a wide range of explanatory questions based on a single, unified set of causal assumptions.

Some scholars claim that the inter-paradigm debates are over – or should be over – because they waste time and effort trying to answer “philosophical” questions when scholars should focus on the more important work of trying to answer “explanatory” questions. This claim seems to ignore the main purpose of forming a paradigm in the first place: to provide researchers with a simple set of theory building assumptions that are well-designed to answer a range of explanatory questions. Each of the conventional paradigms claims that its core assumptions are well designed to answer *all* of the main explanatory questions in IR theory (and beyond). The doctrine of “theoretical pluralism” is now used, however, to claim that each paradigm has a comparative advantage in answering *certain* questions about *certain* aspects of international relations – and the other questions should be left to the other paradigms. This claim seems to support the related claim that debates between the paradigms are only about “philosophical” issues. Neither claim is true.

Debates between the paradigms are not *merely* philosophical debates. They are debates over what is the best way to build explanatory theories. Here “best” is measured by a theory’s explanatory power relative to other theories of the same phenomenon. The essence of this competition between theories is refutation and replacement. Theories that are refuted lose the competition. Alternative theories that escape refutation win the competition (so far). To enter the competition, a theory must be couched in such a form that it is potentially refutable. Thus the inter-paradigm debates are debates over how to build refutable theories and how to determine which theories are, indeed, refuted by which other theories. Therefore, to give up on the inter-paradigm debates without resolving them is to give up on the quest to determine which theories are refuted by which other theories and, in this sense, to give up on scientific progress in the discipline. The inter-paradigm debates need to be resolved. Contractual Realism proposes a way to do it.

First, I summarize the main assumptions and implications of contractual realism. Second, I compare its main assumptions and implications to those of the three conventional paradigms. I show that each conventional paradigm makes assumptions that are not necessary to explain the phenomena in question, and Contractual Realism explains the same phenomena using different assumptions. Finally, I clarify the stakes in letting the inter-paradigm debates continue unresolved.

## **2. Summary of contractual realism**

### **2.1 Main motivation for cooperation**

When the terms of coexistence among actors shift with changes in their relative power, each actor has an incentive to increase its relative power to shift the terms in its favor. When other actors respond in the same way, the result is a competition in raising relative power that is mutually self-defeating. Each actor expends resources to increase its power, yet it is a waste of resources because relative power remains unchanged. The main motivation for cooperation among actors, therefore, is to avoid such wasteful competitions in raising relative power.

### **2.2 Main conditions for cooperation**

To avoid wasteful competitions in raising relative power, actors must meet six conditions.<sup>1</sup>

#### **condition 1: anchor term contracts**

Actors must agree on terms of cooperation that do not shift when their relative power changes. To do so they discern objects in the real world that do not move when their relative power changes. Then they form agreements whose terms distribute rights and obligations by reference to these real world objects. Since the objects do not move when the actors' relative power changes, no actor has an incentive to invest resources in raising its relative power to shift the terms in its favor. (I called the objects "anchors" and the agreements "anchor term

---

<sup>1</sup>Conditions 1, 2 and 4 are due to Rowe (1989, 1990). Conditions 3, 5 and 6 are original.

contracts”.)

### **condition 2: variable benefit terms**

Since the world is always changing, the anchors in the world sometimes move too. Hence actors must write anchor term contracts in a way that specifies how the distribution of rights and obligations between parties will change when the anchors move. As long as each actor knows that an anchor moved due to forces beyond another actor’s control, each actor can allow the other actor to receive a greater share of the asset covered by the contract whenever an anchor moves in its favor. (I call this increased share a ”random variable benefit” because each actor receives it only periodically – when an anchor moves in its favor randomly due to forces beyond its control.)

### **condition 3: justice bargains**

Actors must form ”justice bargains” to enforce the anchor term contracts. One actor commits that if it is found guilty of moving an anchor in its favor intentionally, it will accept the verdict and pay the penalty without resistance. In return the other actor commits that it will allow the first actor to receive its variable benefit whenever an anchor moves in its favor unintentionally (due to forces beyond its control). If the first actor reneges on its commitment, then the second actor reneges on its commitment in retaliation (by withholding the first actor’s variable benefit). This method of contract enforcement does not require actors to use or maintain power assets for enforcement purposes. Hence it does not provoke wasteful competitions in raising relative power. Rather, this method harnesses the natural variability of the world for enforcement purposes.

### **condition 4: adjudication institutions**

Actors must create adjudication institutions to distinguish between unintentional movements of the anchors due to random forces beyond any actor’s control and intentional movements of the anchors due to one actor’s actions. As long as there are institutions to perform this function, actors can withhold the random variable benefit *only* from those who move the anchors intentionally for their own gain. But if such institutions are absent or they fail to

perform this function, then actors will not be able to distinguish between those who comply with the anchor term contracts and those who violate them (by moving the anchors intentionally for their own gain). In this case actors will need to withhold the random variable benefit whenever an anchor moves in one actor's favor for *any* reason – whether intentional or unintentional on his part – simply to deter him from moving the anchors in his favor intentionally in the future!

**condition 5: a social system for labeling compliant actors with social approbation and noncompliant actors with social disapprobation**

Actors need a low-cost way to distinguish between those who comply with their anchor term contracts and those who violate them and defy the guilty verdicts. They assign a label of social approbation to the former and a label of social disapprobation to the latter. This enables actors to identify those from whom to withhold the random variable benefits. As long as the adjudication system functions well (condition 4), the compliant actors will be labeled with approval and the opportunistic actors will be labeled with disapproval. And the latter will be labeled with disapproval *only* for the purpose of withholding their random variable benefits. Then mutual trust, cooperation, civility and social integration remain possible. If the adjudication system becomes ineffective, however, then *both* kinds of actors will be labeled disapprovingly whenever the anchors in the world move for *any* reason – *even reasons beyond their control*. And they will be so labeled not merely to withhold their variable benefit but, more threateningly, to identify them as fair targets for retaliation using power-based methods. Then mutual trust, cooperation, civility and social integration become impossible.

**condition 6: methods for handling novel contingencies contractually**

### **2.3 Main obstacles to cooperation**

Actors face six obstacles to cooperation, corresponding to the six conditions for cooperation listed above.

**obstacle 1** Lack of incentive or capacity to discern terms of cooperation that are invariant to shifts in relative power among actors.

**obstacle 2** Lack of incentive or capacity to discern how the real world anchors specified in the contracts move randomly due to forces beyond any actor's control.

**obstacle 3** Lack of incentive or capacity to form justice bargains to enforce the anchor term contracts.

**obstacle 4** Lack of incentive or capacity to create effective adjudication institutions. Lack of incentive or capacity to maintain such institutions (leading to their breakdown).

**obstacle 5** Lack of low-cost methods for assigning labels of social approbation to compliant actors and social disapprobation to noncompliant actors.

**obstacle 6** Lack of methods for handling novel contingencies contractually.

### **3. Conventional realism's main point and main error**

The main point of conventional realism is that relative power matters in international politics. The main implication is that international cooperation will be prevented or constrained by relative power factors. These points are true, but conventional realism makes them in the wrong way. It adopts the assumption that valuable assets are distributed among states according to relative power at all times – in war and in peace. Hence each state faces the risk that if power shifts against it, another state will expropriate its assets. This risk would be mitigated if the first state could be certain that the second state's intentions will never become aggressive in the future. But no state can be certain of that. Hence no state can cooperate with another state if doing so might shift relative power in the other state's favor. Why can a state never be certain of the future intentions of other states? The

conventional realist answer is: anarchy itself. The anarchical nature of the international system prevents such certainty. And since the system will always be anarchic, cooperation will always be prevented or constrained by uncertainty about other states' intentions. Some conventional realists acknowledge that cooperation is possible. Yet they maintain that states are constrained to cooperate only when relative power capabilities enable it: by insuring that enforcement threats are credible.

To see the error in this analysis, consider its maintained assumptions more closely. Valuable assets are distributed among states according to relative power at all times. Hence states must maintain power at all times, and sometimes use it, to secure their assets against aggression and expropriation by other states. The only conceivable source of asset security is countervailing power. If one state seeks a relative power advantage in order to expropriate another state's assets, the targeted state must deploy countervailing power in its defense. This is the only tool that will keep the first state's drive for power in check. Why? Because the international system is anarchic – there is no global government. The same logic applies to any efforts at cooperation among states. If one state violates the terms of cooperation, the other state must deploy countervailing power to retaliate and deter future violations. This is the only tool that will keep the first state's propensity to commit violations in check.

Conventional realists are correct to assume that *something* keeps the opportunistic drive for a relative power advantage in check. But they are wrong to assume that countervailing power is the only tool that can do the job. In fact, there is another tool that can do the job even better. According to contractual realism, states reach agreements whose terms are invariant to shifts in relative power, so no state has an incentive to increase its power to shift the terms in its favor for its own benefit. Such agreements thus check opportunistic drives for a relative power advantage. I call such agreements "anchor term contracts", because the terms refer to anchors in the real world that do not move even when relative power shifts. These contracts are enforced by justice bargains rather than countervailing power. So cooperation among states is not always constrained by relative power capabilities.

Contractual Realism implies that the primary source of asset security is anchor term contracts. When these contracts are in place and the justice bargains that enforce them are functioning well, the opportunistic drive for power will be checked and valuable assets will



be secure from expropriation. When these contracts are absent or the enforcement devices malfunction, there will be competitions in raising relative power and assets will be insecure.

Thus the conventional realists are right to say that relative power matters in international politics. But they are wrong about how and why. And while they are right to say that cooperation is hindered by relative power factors, they are wrong to say that it is *always* hindered. The hindrance can be overcome by anchor term contracts and justice bargains. As long as these solutions work, relative power factors will *not* hinder cooperation. But when these solutions fail, relative power matters again – in more or less the way that the conventional realists always claimed it matters. States will be driven primarily by relative power concerns. They will avoid policies that might lead to adverse power shifts. And they will forgo lucrative opportunities for mutual gains from cooperation in order to avoid adverse power shifts.

Contractual Realism thus salvages the main point of conventional realism without adopting its key assumptions. In particular, Contractual Realism does not assume that valuable assets are distributed among states according to relative power at all times. It assumes that assets are sometimes distributed among states according to anchor term contracts – ones whose terms are invariant to shifts in relative power among actors.

Conventional realism was first developed to explain great power politics and, in particular, great power wars. Since the 1950s its adherents in the United States have used its paradigmatic assumptions to produce ever more refined versions of its main hypotheses:

1. Valuable assets are divided among states according to relative power at all times, not only in war but in peace as well.
2. Wars are caused by shifts in relative power, uncertainty about relative power, or commitment problems related to relative power.
3. International cooperation is prevented or constrained by relative power at all times.

The historical evidence does not support these hypotheses, however, as we saw in Parts V through VIII of this study. The oscillations between great power peace and great power war since 1550 confirm the Contractual Realist theory instead. They confirm it in four ways.

### 3.1 Patterns of war and peace explained by Contractual Realism

First, each of the seven hegemonic wars since 1550 was caused by a novel contingency that was not covered by the existing anchor term contracts between the great powers (see Part V of this study). Second, each war turned into a competition in raising relative power that destroyed far more economic resources than were at stake in the dispute that caused the war (Part VI). This fact alone supports the Contractual Realist claim that the main purpose of peace is to divide valuable assets among states on terms that are invariant to changes in relative power *in order to avoid competitions in raising relative power*.

Third, once each hegemonic war began, assets were divided among states based on relative power. This provoked a competition in raising relative power that forced each state to reorganize its domestic institutions and policies completely (Part VII). This fact suggests that during the preceding peace, the domestic institutions and policies could be organized so differently because states were *not* dividing assets according to relative power in peace time. They were dividing assets according to contracts whose terms were invariant to shifts in relative power. Fourth, the peace settlements following the hegemonic wars fell into two categories: the settlements that were stable because they were based on anchor term contracts, and the settlements that were unstable because they were not based on anchor term contracts (Part VIII). This pattern also supports the Contractual Realist claim that the function of peace settlements is to divide valuable assets among states on terms that are invariant to changes in relative power in order to avoid competitions in raising relative power.

## 4. Neoliberal institutionalism's main point and main error

In response to conventional realism, neoliberal institutionalism proposed that anarchy need not prevent or constrain cooperation as long as states meet certain conditions. The main condition is that states must establish and maintain international institutions designed to lower the transaction costs of negotiating and modifying international agreements. From the neoliberal institutionalist perspective, the main obstacle to cooperation is not uncertainty about the *future* intentions of other states or the direction of *future* power shifts, but

the possibility that some states will cheat on international agreements in the *present*. The function of international institutions is to enable honest states to screen out cheaters, negotiate agreements that are cheater proof, modify the agreements in response to new forms of cheating, and enforce the agreements by punishing cheaters. As long as institutions are well designed to perform these functions, states can cooperate *despite* their uncertainty about the direction of future power shifts and the intentions of other states in the future. According to neoliberal institutionalism, the main tool for punishing cheaters is boycotting. Honest states simply boycott the cheater states that violate their international agreements. In bilateral agreements the enforcer state simply withdraws from the agreement and declines to cooperate further with the violator state on that issue. In multilateral agreements the violator state is excluded from membership in the group cooperating on that issue and denied the benefits of membership.

Neoliberal institutionalism is correct in saying that cooperation is possible under anarchy. But it makes the point the wrong way. It assumes that the success of international cooperation depends on the relative power of states: the power of the enforcer states to detect and punish cheating *relative to* the power of violator states to evade detection and resist or withstand punishment. Boycotting is costly not only to the violator state that is boycotted but also to the enforcer state(s) that do the boycotting. They forgo the gains from cooperation with the violator state. Hence the feasibility of boycotting depends on their capacity to do without their share of those gains relative to the capacity of the boycotted state to do without its share. In equilibrium, the distribution of the mutual gains from cooperation is determined by the relative power of all states to detect and punish cheating, evade detection and resist or withstand punishment. Hence each state has an incentive to invest resources in maximizing its power on these margins. But the investments are mutually canceling and relative power does not change, so they are a pure waste of resources (or deadweight cost). Such competitions in raising relative power may destroy most or all of the mutual gains from cooperation. (See the technical appendix for a formal model that proves this point.)

In short, if states were really cooperating in the way that neoliberal institutionalism assumes, they would be left in a world where competitions in raising relative power simply continue in a new guise. And the *net* mutual gains from international cooperation (net of the

costs of competitive investments in raising relative power) would be negligible or negative.

If this is not the way that states actually cooperate with each other, then how do they cooperate? After all, there is much evidence that states engage in international cooperation. There must be some explanation for it. Contractual Realism explains it by distinguishing between two very different ways that enforcer states deny the benefits of cooperation to violator states in retaliation for their violations. The first way is boycotting – the main method of enforcement assumed by neoliberal institutionalist theories. The second way is *random-variable benefit withholding* (RBW). This enforcement tool is based on the main feature of anchor term contracts: the terms refer to anchors in the real world that do not move when relative power shifts among states. Because the world is always changing, the anchors sometimes move due to random forces beyond any state's control. The contract terms specify how the distribution of the gains from cooperation will be allowed to shift between partners whenever the anchors move due to random forces beyond any partner's control. In particular, each state receives a *random variable benefit* whenever an anchor moves in its favor due to forces beyond its control. Other states allow it to receive its random variable benefit as long as they know it did not move the anchor in its own favor intentionally. But if it is found guilty of moving an anchor in its favor intentionally, then the other states withhold its random variable benefit in retaliation.

The key feature of this enforcement tool is that it *redistributes* the mutual gains from cooperation without *reducing* their magnitude. This feature contrasts sharply with the boycotting tool, which reduces the magnitude of the mutual gains from cooperation and is thus costly for both the enforcer and violator states. To appreciate the significance of the contrast, consider RBW more closely. Normally, in the absence of any violations, each state allows the distribution of the mutual gains from cooperation to shift in favor of another state whenever an anchor moves in its favor due to random forces beyond its control. But when one state moves an anchor in its own favor intentionally, the enforcer state does *not* allow the distribution of the mutual gains to shift in favor of the violator state. So the enforcer state retains that share of the gains – *which it would not normally receive*. (And the violator state is denied that share of the gains which it normally would receive.) In this way enforcement activity is actually beneficial for the enforcer state – not costly as in boycotting.

This difference changes the economics of enforcement completely, thereby enabling cooperation to occur on issues and among states where it could never occur if boycotting were the only available enforcement tool. Moreover, anchor term contracts change the economics of enforcement in another way that also makes cooperation feasible where it would not otherwise be. Such contracts distribute the mutual gains from cooperation in a way that is independent of the relative power of states to detect and punish cheating, evade detection and withstand punishment. So states have no incentive to invest resources in raising their relative power on these margins. And competitions in raising relative power on these margins are avoided, thereby making the net benefit from cooperation positive on issues where it would be negative if boycotting were the only enforcement tool available.

This is how Contractual Realism explains most forms of international cooperation that actually generate significant benefits for the participating states. Thus neoliberal institutionalism has a point to say that international cooperation is possible under anarchy. But it is wrong to assume that the threat of boycotts is the main way cooperation is enforced and therefore the main explanation for the observed levels of cooperation.

The two competing explanations of international cooperation – neoliberal institutionalism and Contractual Realism – generate different implications for what we should expect to observe when cooperation breaks down. The boycotting model only has one punishment equilibrium. When a state violates an international agreement, another state or set of states boycotts the violator. Cooperation ceases. The two sides are returned to where they were before the cooperation began: not cooperating and not reaping mutual gains from cooperation, but otherwise in stable coexistence.

By contrast, the Contractual Realist theory of cooperation has two punishment equilibria, and the second one yields very different observable implications from the first one. First, when a state is found guilty of moving a benchmark in its favor intentionally but defies the guilty verdict, another state withholds its random variable benefit in retaliation. This punishment redistributes the mutual gains from cooperation but does not reduce them (as explained above). Second, when a novel contingency arises that is not covered by the anchor term contracts (or the adjudication institutions for distinguishing between intentional and unintentional movements of the anchors break down), it causes a dispute among states

over whether one of them moved a benchmark intentionally for its own gain – or the other one is ignoring evidence that the benchmark moved unintentionally due to random forces beyond the first one's control. Then one state must punish the other for having moved the benchmark intentionally for its own gain, and the other one must punish the first for ignoring evidence that the benchmark moved unintentionally on its part. In this equilibrium, RBW enforcement is no longer feasible, because there is no way to distinguish intentional from unintentional movements of the anchors. So states resort to the other main enforcement tool at their disposal: imposing costs on each other by power-based methods. The result is a competition to raise relative power to enforce costs on each other that requires each state to invest resources in raising its power. But the investments are mutually canceling and relative power does not change, so the investments are a pure waste of resources (or deadweight cost). In this equilibrium, not only are the potential gains from cooperation lost, but if the stakes in the dispute are high enough the competition in raising relative power is potentially bottomless (in deadweight cost).

This equilibrium explains the main episodes of cooperation breakdown in European history – both internationally and domestically. In each episode there was a clear, observable correlation between the causal variable and the effect variable. The causal variable was either a breakdown of adjudication institutions or a novel contingency that was not covered by existing anchor term contracts. The effect variable was the breakdown of cooperation, the outbreak of conflict, and a competition in raising relative power that destabilized politics and destroyed economic resources far beyond those at stake in the issue in dispute.

These episodes confirm the causal mechanism proposed by contractual realism. When cooperation occurs, it is because anchor term contracts enable states to escape the world of competitions in raising relative power – even the relative power to detect and punish cheating, evade detection and resist or withstand punishment. When cooperation breaks down, that escape mechanism fails and competitions in raising relative power break out to the detriment of all.

## 4.1 Types of cooperation success and failure explained by Contractual Realism

### 4.1.1 Human rights accords

Both the Peace of Augsburg (1555) and the Peace of Westphalia (1648) were settlements of religious wars. Each settlement contained terms that were intended to protect religious minorities. In this sense each settlement was a human rights accord. The two settlements were not equally effective, however. The Peace of Augsburg was not stable in the long run, because it was not based on an anchor term contract. By contrast, the Peace of Westphalia was stable in the long run, because it was based on an anchor term contract. Consider each settlement in turn.

The Peace of Augsburg established the principle of *cuius regio, eius religio* (to he who rules, his religion can be established). This principal gave each territorial ruler the right to require his subjects either to adopt his confession or to emigrate to another territory. It also gave each ruler the right to switch his own confession if the majority of his subjects belonged to the other confession (or simply if it suited his interests in balance of power politics).

These clauses were *not* anchor terms. On the contrary, these clauses referred to anchors in the real world that *did* move with changes in the relative power of actors. One such anchor was "the ruler's confession". Another was "the percentage of a territory's population that belonged to each confession". Each of these anchors moved with shifts in relative power among territories. Hence each ruler had an incentive to invest resources in increasing his relative power in order to move these anchors in his favor.

For example, each ruler had an incentive to increase his power by building a regional coalition on a confessional basis, because this would induce other rulers to switch confessions so they could join his coalition. Hence the anchor of "ruler's confession" moved with changes in the relative power of regional coalitions. Similarly, each ruler had an incentive to increase his power by promoting conversions among the subjects of a neighboring ruler, because this would increase the percentage of the neighbor's population that belonged to one's own confession, and thereby induce the neighbor to switch confessions and bring the rest of his population with him. Thus the anchor of "percentage of a territory's population that belonged to each confession" moved with changes in the relative power of rulers to promote

conversions in neighboring territories.

The long-run effect of these terms of settlement was a prolonged competition in raising relative power. Between 1555 and 1570, the Protestant camp increased in size and power. By 1580 the Catholic camp started to mount a counter-balancing campaign to reverse the trend. By the 1590s the outcome was stalemate, deadlock and the paralysis of adjudicative institutions for resolving disputes between the confessions. Between 1603 and 1615 a series of minor military conflicts erupted due largely to inter-confessional rivalries and tensions. The crisis of 1618 erupted at the nadir of a prolonged competition in raising relative power.

By contrast, the Peace of Westphalia (1648) was based on anchor term contracts, so it was stable in the long run. The heart of the settlement lay in the local agreements in areas of mixed populations where some people belonged to one confession and others belonged to another confession. These accords distributed valuable assets of all kinds between the confessions according to anchor terms. For example, seats on city councils were divided according to a "parity rule", so the distribution of seats was immune to changes in relative population size across the confessions. To take another example, the usage of church facilities was apportioned between the confessions according to fixed rules, so that the distribution of usage rights did not change with changes in relative population size among the confessions. The long-run effect of the settlement was to neutralize religion as a source of military conflict in Germany.

#### **4.1.2 Limits on foreign interference in domestic politics**

Britain and France fought against each other in three hegemonic wars between 1688 and 1815. All of these wars were fought in part over the right of foreign interference in domestic politics. Britain and France signed peace treaties to end the wars in 1697, 1713 and 1815. Each settlement was intended in part to enact limits on foreign interference in domestic politics. The three settlements were not equally effective, however. The settlements of 1697 and 1713 were not stable in the long run, because their commitments to refrain from foreign interference were not rooted in anchor term contracts. By contrast, the settlement of 1815 was stable in the long run, because its commitments to refrain from foreign interference were rooted in an anchor term contract. (See Parts V and VIII of this study for the logic and



evidence for these conclusions.)

### 4.1.3 International trade agreements

The settlements of World Wars I and II were intended in part to restore international trade among the European great powers under the assumption that multilateral free trade would help to stabilize great power politics. The two settlements were not equally effective, however. The settlement of World War I was not stable in the long run, because its trade component was not based on an anchor term contract. By contrast, the settlement of World War II was stable in the long run, because its trade component was based on an anchor term contract. (See Part VIII of this study for the logic and evidence for these conclusions.)

## 4.2 Neoliberal institutionalism's second error

Neoliberal institutionalism also commits a second error that prevents it from understanding how cooperation succeeds and fails in the real world. Neoliberal institutionalism assumes that boycotting and shaming are two different tools for enforcing cooperation that operate by two distinct mechanisms – the one tool being used in some instances and the other tool being used in other instances. Boycotting is assumed to operate through material incentives while shaming is assumed to operate through emotional mechanisms (such as embarrassment, humiliation, loss of face, honor, prestige, etc.).

By contrast, Contractual Realism assumes that shaming is an integral part of the enforcement method of *random-variable benefit withholding* (RBW). States need a way to distinguish between partners that comply with their anchor term contracts and partners that violate them (by moving anchors intentionally for their own gain). States assign a label of social approbation to the partners that comply and a label of social disapprobation to the partners that violate and then defy the guilty verdicts. This enables states to identify the ones from whom to withhold the random variable benefits. Scholars might think that the label of social disapprobation is a shaming mechanism of enforcement, where the enforcer states shame the violator states to impose an emotional cost on them (through embarrassment, etc.). But in fact, the label of social disapprobation is part of a material enforcement mechanism: random-variable benefit withholding. From this perspective, the distinction between

emotional enforcement through shaming and material enforcement through withholding of material benefits is wrongheaded. In fact social disapprobation and material coercion are integral parts of the same enforcement mechanism: the mechanism would not work if either part were absent.

Contractual Realism shows there is a tight connection between social disapprobation and material coercion in the enforcement of cooperation. By specifying this connection precisely, Contractual Realism provides a single theory by which to study two subjects: the material incentives that shape the forms and fate of cooperation (e.g. the traditional concerns of neoliberal institutionalism) and the ideational factors that shape the forms and fate of shared identities of interest (e.g. the traditional concerns of constructivism). This single theory explains why, throughout European history, the breakdown of cooperation was correlated with the disintegration of shared identities and the rise of social animosities. The explanation is straightforward. As long as the existing anchor term contracts and judicial institutions were functioning well, the labels of social disapprobation were only applied to actors who moved anchors intentionally for their own gain and defied the guilty verdicts. In this environment, social cohesion remained possible. But when new contingencies arose that were not covered by the existing contracts (or the judicial institutions broke down), actors had no choice but to apply the labels of social disapprobation to anyone who gained from any movements of the anchors, even movements that were unintentional on their part. In this environment, social cohesion was impossible.

This mechanism explains why each of the episodes of cooperation breakdown listed in the previous subsection was also accompanied by social breakdown. And again, there was a clear, observable correlation between the causal variable and the effect variable in each episode. The causal variable was a novel contingency not covered by existing contracts or a breakdown of adjudication institutions. The effect variable was the breakdown of cooperation, the disintegration of social identities and the destruction of social fabrics. These episodes provide further confirmation of the causal mechanism proposed by contractual realism.

## 5. Constructivism's main point and main error

Constructivism's main point is that ideas and shared knowledge matter in international politics. Ideas shape people's conceptions of their material interests, their reasons for choosing one action over another, and their behavior. Shared ideas define the boundary of a social identity and distinguish it from other social identities whose members share *other* ideas. In this way a person's social identity shapes his conception of his material interests, his reasoning about options for acting this way or that way, and his behavior. People's behavior is thus driven in large part by the ideas and knowledge they share in communities of social consensus. This is what constructivism means when it says that the world is "socially constructed". Constructivism acknowledges that people's behavior is driven by their material interests as well, but only insofar as these interests are constructed in one way or another by the ideas that people hold and share. There is no straight line from people's material interests to their actions.

Constructivism is right to say that ideas and shared knowledge matter in this way. Constructivism's main error is that it fails to identify the fundamental reason *why* ideas and shared knowledge matter in this way. Contractual Realism identifies the fundamental reason why.

### 5.1 The Contractual Realist explanation for why ideas and shared knowledge matter

The main motivation for cooperation of all kinds is to prevent competitions in raising relative power. The main method for preventing such competitions is anchor term contracts. These contracts distribute assets among actors according to terms that do not shift when relative power shifts among actors. Hence no actor has an incentive to increase his power to shift the terms in his favor. Thus the value of adhering to the contract terms is to avoid competitions in raising relative power. If actors are to reap this value, however, they cannot change the contract terms at will. They must adhere to the contract terms *as originally written*. When they do so, the original contract terms gain a transcendental value: they enable actors to avoid competitions in raising relative power. Contract terms are thus a special *category* of

ideas. They are the ideas that enable actors to avoid competitions in raising relative power. *This* is why ideas matter so much in politics, economy and society generally. The long history of conflicts and settlements in Europe – economic, social and political – is filled with ideas of this kind.

This theory explains *which* ideas matter the most to actors and why they matter the most. The ideas that matter the most are those that serve as the terms of anchor term contracts. These ideas matter the most because they refer to the objects in the real world that matter the most: the anchors. The reason that the anchors matter so much is that they do not move when relative power shifts among actors. So the anchors enable actors to divide assets in a way that is independent of relative power, and thereby avoid competitions in raising relative power. In short, some ideas matter more than others because some objects in the real world matter more than others.

## **5.2 The Contractual Realist explanation for why ideas matter in the seven ways that they do**

Constructivism claims that ideas matter in seven main ways. Constructivism is right. Yet constructivism does not explain *why* ideas matter in these seven ways. Contractual Realism explains why. For each one of the seven ways, I first quote the constructivist view that ideas matter in this way. Then I provide historical evidence that ideas do indeed matter in this way. And then I provide the Contractual Realist explanation for *why* ideas matter in this way.

### **5.2.1 Ideas *fix the meanings* of material objects and physical bodies in the real world**

According to a leading theorist in the school:

”Constructivism considers intersubjective knowledge and ideas to have constitutive effects on social reality and its evolution. When drawn upon by individuals, the rules, norms and cause-and-effect understandings *that make material objects meaningful* become the source of people’s reasons, interests and intentional

acts....”<sup>2</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that ideas fix meanings.

The main method that actors used to fix the meanings of objects and bodies, according to contractual realism, is anchor term contracts. The contract terms refer to anchors in the real world that do not move when relative power shifts among actors. Because the world is always changing, the anchors sometimes move due to random forces beyond any actor’s control. So each contract specifies how the distribution of rights and obligations will shift between the contracting parties when the anchor moves randomly. Consider a simple but non-trivial example from the common law of medieval England.

When a landholder died, his land was inherited by his son or his brother:

case	anchor	description of anchor	contract term
default case	0	deceased has a son	son is "the heir"
exception case	1	deceased has a brother	brother is "the heir"

The anchor in the real world was actually a *composite* of two parts. As long as the two parts always moved together in perfect correlation in the real world, however, there was never any cause for dispute. That is, as long as each deceased landholder had *either* a son and no brother *or* a brother and no son, then disputes never arose. Denoting the first part of the anchor with a single dot and the second part of the anchor with a double dot, the situation where the two parts always moved together in perfect correlation was:

case	anchor	description of anchor	contract term
default case	ôö	ô deceased has a son; ö = deceased has no brother	son is "the heir"
exception case	ïï	ï deceased has no son; ï = deceased has a brother	brother is "the heir"

But what happened if the two parts started to move independently of each other? Then

---

<sup>2</sup>Adler 2002: 102. Emphasis added.

there might be cases where the deceased landowner had a son *and* a brother (exception case B):

case	anchor	description of anchor	contract term
default case	ö	ö = deceased has a son; ö = deceased has no brother	son is "the heir"
exception case A	ī	ī = deceased has no son; ī = deceased has a brother	brother is "the heir"
exception case B	ō	ō = deceased has a son; ī = deceased has a brother	son or brother?
exception case C	ī	ī = deceased has no son; ī = deceased has no brother	?

This is not a trivial example. According to the leading legal historian of medieval England, many private wars were fought between sons and their uncles over this issue.<sup>3</sup> To avoid such conflicts, the common law courts had to decide whether the sons or the brothers would inherit in such cases. Either rule would be sufficient to avoid conflict, and there was no particular reason to expect that one rule would be more efficient than the other in economic terms. But one rule or the other had to be the *authoritative* rule.

This example illustrates the main point of the previous subsection: Some ideas matter more than others because some material objects and physical bodies in the real world matter more than others.

When a landholder died, the physical body that mattered the most was "the heir". Now the idea of "an heir" was just that – an idea. But it was an important idea because it referred to a physical body in the world whose identity would not change even when relative power among actors did change. So actors could avoid competitions in raising relative power by basing inheritance contracts – wills and bequests – on "heirs". Avoiding these conflicts and competitions was a transcendent value. It *induced* actors to elevate the idea of an "heir" to a privileged status in the domain of inheritance contracting. In other words, the idea of an "heir" mattered more than others, most fundamentally, because the idea enabled actors to avoid such conflicts and competitions. The process of identifying an heir in a particular case may have been costly. Yet it was a cost worth paying to avoid the even greater cost of those

---

<sup>3</sup>Baker 2002.

conflicts and competitions. Thus, as long as the costs of forming and reforming anchor term contracts were kept affordable, then actors could avoid conflicts and competitions in raising relative power. But if the costs of anchor term contract management rose too high, then actors would not be able to avoid such conflicts and competitions.

The example of inheritance illustrates a number of general points that have wide applicability (as we will see).

### **A Contractual Realist model of the evolution of contracts over time**

1. When a contract is first formed, the contracting parties often assume unconsciously that the anchor in the real world is a single object, simply because its component parts have always moved together in perfect correlation in their experience.
2. When the parts of an anchor start to move independently of each other for the first time, it creates a novel contingency that is not covered by the original contract.
3. There are always at least two different ways to amend the contract to cover such contingencies. One way awards the right in question to one of the contracting parties. The other way awards the right in question to the other party.
4. Since this situation pits one party's material interest against the other party's material interest, such contingencies can generate disputes.
5. These disputes can be resolved short of conflict as long as the two parties reach agreement on which of the two ways the contingency will be handled from then on.
6. Often there is no efficiency reason to handle it one way or the other, but since the choice has distributional consequences between the contracting parties, it is still contentious.
7. If the choice favors one party again and again with each novel contingency that arises, there will eventually be pressure from the other party to adopt a more evenhanded approach.
8. With an evenhanded approach, each party is given the next contingent right created by the next novel contingency at least some of the time. So over time, a rough balance of contractual rights is maintained.
9. Within each novel contingency, it is essentially arbitrary which of the two parties is given the contingent right in question. And each party is content not to receive it as

long as a rough balance of contractual rights is maintained across contingencies over time.

10. The resulting distribution of contractual rights between the contracting parties can be highly arbitrary at the micro level, and it can vary significantly from one environment to another (e.g. from one culture to another, one civilization to another, etc.). For example, inheritance laws varied significantly from one region to another in medieval Europe.
11. As parts of an anchor that previously moved together start to move independently of each other for the first time, the definition of the anchor becomes more differentiated over time. Each time this happens, one part of the anchor is given legal recognition and the other part is not. For example, sons are given legal recognition as "heirs" while brothers are not. As a result of these decisions, some objective facts in the real world become visible to the law while others become invisible to the law. Which facts are visible is essentially arbitrary, and it can vary significantly from one environment to another (for the reasons explained above).

### **How and why ideas fix meanings**

This model explains how and why it is the case that ideas fix the meanings of material objects and physical bodies in the world. Actors are motivated to divide valuable assets according to anchor term contracts, because they want to avoid competitions in raising relative power. So actors write contracts in terms of anchors – objects and bodies in the real world – that do not move when relative power shifts among actors. In this way, actors distribute assets according to *verbal* contracts whose terms are independent of material *power*. Actors are still motivated by two material interests in doing this, however. First, they want the assets for their material *value*. Second, they want to preserve the assets' values from wasteful competitions in power-raising. Thus the meanings of real-world objects and bodies – as anchors – are determined *simultaneously* by the ideas embodied in the contract terms and the material interests embodied in the contractors.

The contract terms are amended periodically to keep up with changes in the real world. When a contract is amended (#11 in the model above), it gives one part of the anchor legal recognition and leaves the other part of the anchor without legal recognition. Since an anchor



is an object or body in the real world, this gives one fact in the real world visibility and renders another fact in the real world invisible. The facts that remain visible are the ones that will determine how valuable assets are distributed among contracting parties in those contingencies. So the meaning of these facts for these parties is that they determine whether each party will gain or lose a share of a *materially* valuable asset. Consequently, actors have a material interest in the way that contracts are amended in response to novel contingencies. This gives them a material interest in which of the *ideas* embodied in contract terms will receive priority in that amendment process and which other ideas will lose priority. In short, some ideas matter more than others because some material objects and physical bodies in the real world matter more than others.

These points lead naturally to the topic of why and how ideas play the role that they do in the evolution of contracts, norms and laws in response to novel contingencies.

### **5.2.2 Ideas *enable the changes* in shared understandings that are needed to avoid conflict**

According to the same leading theorist:

”Constructivism’s added value...is to take *change* less as the alteration in the positions of material things than as the emergence of new constitutive rules, the evolution and transformation of new social structures, and the agent-related origins of social processes.”<sup>4</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that ideas enable change.

First consider the example of inheritance laws in medieval England. Between 1100 and 1300, the common law courts evolved a system of inheritance laws that covered every possible contingency in which a legal heir needed to be identified by the court to avoid a conflict. The ”parentelic scheme” specified which living relatives came first, second, third, etc. in line to inherit: .

---

<sup>4</sup>Adler 2002: 102. And citations therein. Emphasis added.

Inheritance law is not the only area in which rules and norms change, adapt and grow to avoid conflict. Much of the history of modern contract law was an elaboration of new contract terms that specified the distribution of rights and obligations across parties in novel contingencies not previously encountered. Contractual Realism hypothesizes that rules and norms evolve in a similar way in all areas of politics, the economy and society. Consider a critical example from the Concert of Europe in the 19th century.

In the Vienna settlement of 1815, the great powers agreed not to depose the incumbent regimes of smaller states by military force. They also agreed that the standard for identifying an incumbent regime was "legitimacy". These agreements constituted an anchor term contract. The anchor was legitimacy. This anchor did not move even when relative power shifted among the great powers. For according to the law of divine right, a sitting monarch was the legitimate ruler as long as he lived. This anchor only moved when the sitting monarch died. Then his nearest relative would inherit the legitimacy and retain it as long as *he* lived. Because the anchor of legitimacy did not move even when relative power shifted among the great powers, none of them had an incentive to increase its relative power for that purpose.

case	anchor	description of anchor	contract term
default case	0	monarch rules by law of divine right	"legitimate" regime
exception case	1	nearest relative rules by law of succession	"legitimate" regime

This contract reflected the main assumption about legitimacy during the *ancien regime* (e.g. before 1789). A sitting monarch was the ultimate source of justice in his kingdom. So his rule was just, *ipso facto*, simply because he was the sitting monarch. According to the law of divine right, the sitting monarch could not be replaced while he lived simply because some of his subjects thought his rule was unjust. Dissent was illegitimate *per se* – if it meant calling for the monarch to be replaced while he was still alive. In contractual terms, this meant that the anchor of "legitimacy" was a composite of two parts: the sitting monarch and his history of ruling justly. But by definition the second part never varied in the eyes of his subjects. His rule was just in their eyes *ipso facto*:

case	anchor	description of anchor	contract term
default case	öö	ô = monarch rules by law of divine right; ö = monarch rules justly	"legitimate" regime
exception case	ïö	î = nearest relative rules by law of succession; ö = successor rules justly	"legitimate" regime

The French revolution changed this assumption about legitimacy. After 1815, subjects voiced open dissent when a monarch ruled unjustly. And subjects called for his replacement by a close relative according to the law of succession – even while he was still alive. In contractual terms, the second part of the anchor – just rule – was now a variable in the eyes of the subjects. While some subjects thought he ruled justly, others thought he ruled unjustly and should be replaced by a relative who would rule more justly. Now the two parts of the anchor were moving independently of each other. This situation generated disputes over what constituted the legitimate government. Did the sitting monarch's government retain legitimacy even though it ruled unjustly? Or could a close relative take over the regime and earn popular legitimacy by ruling more justly?

case	anchor	description of anchor	contract term
default case	öö	ô = monarch rules by law of divine right; ö = monarch rules justly	"legitimate" regime
exception case A	ôi	ô = monarch rules by law of divine right; î = monarch rules unjustly	"legitimacy" ?
exception case B	ïö	î = nearest relative rules by law of succession; ö = successor rules justly	"legitimacy" ?

There were two different ways to settle this kind of dispute. The first way was to assign legitimacy to the sitting monarch, however unjustly he ruled. This was the solution adopted by the Eastern powers – Austria, Prussia and Russia – when disputes of this kind erupted in Eastern Europe. The second way was to regard the unjust monarch's replacement as legitimate as long as that replacement ruled more justly. This was the solution adopted by the Western powers – France and Britain – when disputes of this kind erupted in Western Europe.<sup>5</sup>

---

<sup>5</sup>The main examples are the legitimacy and succession struggles in Spain and Portugal between 1815 and 1848. See Dakin in Sked (1979).

Thus, by updating their anchor term contract to cover the novel contingency of domestic reform movements, the great powers were able to avoid conflict. Between 1815 and 1853, many domestic conflicts broke out in smaller states over the issues of domestic reform and legitimacy. Although the great powers had sharply conflicting interests in many of these cases, they did not go to war over any of them. The key factor in maintaining the peace, according to contractual realism, was the great powers' use of an anchor term – legitimacy – and the great powers' capacity to expand the classification scheme for determining a regime's legitimacy when the novel contingency of reform movements forced them to do so.

The great powers' adaptability illustrates how a principled idea – legitimacy – enabled change. First, the idea served as an anchor term that distributed a valuable asset among contracting parties in way that was mutually agreeable, understandable, and invariant to shifts in relative power. Then, when the component parts of the anchor started to move independently of each other for the first time, the contracting parties wrote additional contract terms to cover the new contingency, thereby extending the principled idea's applicability without abandoning its transcendental value of avoiding competitions in raising relative power.

### **5.2.3 Ideas *provide the ingredients* for practical rationality in maintaining shared understandings and identities**

According to the same leading theorist:

"[Constructivists] advance the notion of practical or communicative rationality....[A]gents do not choose between the most efficient alternative [sic], but 'follow rules that associate particular identities to particular situations, approaching individual opportunities for action by assessing similarities between current identities and choice dilemmas and more general *concepts* of self and situations'.... [I]nstrumental action is prompted by expectations and intentions, which are drawn from previously constituted social structures...."<sup>6</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that ideas provide the ingredients of practical rationality.

---

<sup>6</sup>Adler 2002: 102-3, citing March and Olsen 1998: 951. Emphasis added.

Consider the example of court process in medieval England. The main goal of the court system was to maintain shared understandings between plaintiffs and defendants in order to maintain social consensus and peace. To make the court process tractable, judges and litigators developed a simple classification system for the accusations that plaintiffs made against defendants. Complaints were classified into different "forms of action" depending on the type of complaint it was. Each form of action was specified in precise language that clarified exactly where the line was between legal and illegal behavior, and exactly what line the plaintiff claimed the defendant had crossed to the point of warranting a suit. For example, there was a form of action for a tenant's complaint that he had been wrongly evicted from his land by his lord. There was another form of action for a landholder's complaint that a neighbor had trespassed on his land and caused economic damage. There were about 20 forms of action in total.

This system had two notable features from the perspective of Contractual Realist theory. First, if a plaintiff could not couch his complaint in terms of one of the forms of action, then he had no right to make his complaint at all. Second, once the plaintiff chose a form of action, he was obligated to fit the facts of his complaint into the precise language of that form, or else he had no right to make his complaint at all. The courts could reject complaints on the grounds that the plaintiff had failed to fit his facts into the form's language.

These features show that the forms of action constituted a set of anchor term contracts between the members of English society – and between the government and society. The forms were anchor term contracts in that their terms were invariant to shifts in relative power between these actors. The set of forms and the language of each form remained the same from case to case, regardless of the relative power of the litigants and regardless of their power (or weakness) relative to the government. Occasionally the central court officers created a new form of action to acknowledge a new way that members of society were harming each other's interests so that the victims could obtain relief at court. But these were the exceptions that proved the rule. The forms of action were more or less fixed and if a plaintiff could not find a form for his facts then he could not find relief at court. This stability reflected the limits of the government's power at this period. The government could not right every wrong in society. It could only right a limited number of wrongs: those codified in the forms of action.

These forms demonstrate that *certain* ideas have the power to maintain mutual understandings and social consensus – the ideas embodied in anchor term contracts. This point is proven by the way in which trial verdicts were enforced. Sheriffs often lacked the military power to enforce guilty verdicts on defendants who defied them.<sup>7</sup> Nonetheless, medieval England's rate of compliance with guilty verdicts was relatively high compared to the rate of compliance in medieval Germany, which lacked a simple classification system like the forms of action. In England the court system's efficiency gave the government a powerful tool with which to punish lords who evicted tenants from their land illegally and then defied the guilty verdicts. The king simply took the lord's lands "into his hands". This did not involve any physical seizure of the lords' lands (as the king's sheriffs often lacked the military power to do that). Rather, the king simply instructed his court officers to deny the lord access to the court system in the lord's *other* cases where the tables were turned – where the lord was a plaintiff trying to recover land that should have reverted to him on the death of a tenant, but the tenant's relatives were refusing to surrender it.<sup>8</sup> Such reversionary rights were common and valuable. So when the king denied a lord access to the courts to recover his reversionary rights, it cost the lord dearly.

A lord's reversionary rights constituted a random variable benefit, because tenants died at unpredictable and uncontrollable times. Thus, when the king prevented a lord from recovering his reversionary rights by denying him access to the courts, the king was withholding a random variable benefit from the lord. This enforcement tool thus fits the description of *random variable benefit withholding*, as defined by Contractual Realist theory. This was one of the main tools for enforcing the forms of action which protected the landholding rights of tenants from encroaching lords. In sum, the forms of action were anchor term contracts enforced by justice bargains.

This system worked because it was simple and unchanging. There were only about 20 forms of action and each complaint had to fit precisely into one form. So the courts did not need a large number of officers to process complaints. Hence the courts could process complaints at a relatively low cost. (Presumably, if there had been 100 or 200 forms of

---

<sup>7</sup>Morris 1927, Baker 2004.

<sup>8</sup>Turner.

action and the complaints could deviate from the forms in idiosyncratic ways, the courts would have needed a much larger number of officers per complaint to deliver justice. And the cost of justice would have been much higher.) Once the members of society realized that justice could be attained cheaply, their demand for justice increased significantly. But the system remained affordable because it remained simple and unchanging.

This example shows *how* anchor term contracts provide the main ingredients of practical rationality in maintaining mutual understandings and social consensus. Contractual Realist theory also explains *why* anchor term contracts provide the main ingredients of practical rationality. Put simply, the process of maintaining mutual understandings and social consensus is costly itself. As long as these costs are kept affordable, then actors can avoid conflicts and competitions in raising relative power. But if these costs rise too high, then actors cannot avoid such conflicts and competitions. Anchor term contracts keep the process affordable and feasible by keeping it focused on the main factor needed for peace and social consensus: Contracts that distribute valuable assets among actors in a way that is invariant to shifts in relative power.

#### **5.2.4 Through language, ideas *limit the definitions* of problems and solutions – and sometimes enable new definitions**

According to constructivism:

”[L]anguage is the medium for the construction of intersubjective meanings.... [D]iscourse...is power, in the sense that 'it makes us understand certain problems in certain ways, and pose questions accordingly. It thereby limits the range of alternative policy options, and it enables us to take on others'.... [L]anguage expressions represent the potential for new constitutions of reality...[so] discourse is also a source of change.”<sup>9</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that language imposes limits on problem-solving ideas and sometimes enables creativity in problem-solving with ideas.

---

<sup>9</sup>Adler 2002: 103. And citations therein. Quote is from Diez 1999: 607.

The linguistic roles and limits of ideas are also demonstrated by the court process in medieval England, and the forms of action in particular. As explained in the previous subsection, the forms of action provided the ingredients for practical rationality in maintaining mutual understandings and social consensus. The impact of the forms was limited by language itself. If a plaintiff could not fit his facts into one of the forms, then he had no right to make his complaint at all. Once he found a form to fit his facts, he had to fit his facts into that form precisely – without a letter of deviation – or else the court officers could reject his complaint. These limits could disqualify a plaintiff’s claim even when the claim was legitimate in principle. Thus the limits of language itself had a significant impact on the material interests of plaintiffs and defendants.

This impact is also explained by the affordability and feasibility constraint. In a world where actors competed for assets in myriad ways, not every kind of wrong could be righted by the government. If it had tried to right every wrong, the cost would have been greater than the average member of society was able or willing to pay. The government needed to draw the line somewhere. Its limits in righting wrongs were expressed, in the end, in its linguistic rules for what qualified as a legitimate complaint. Thus the constraints of language were, in the end, due to the constraints of economic affordability in a world where the main source of mutual understanding, social consensus and peace was anchor term contracts.

### 5.2.5 Ideas *constrain communication and negotiation* over the definition of problems and the feasibility of alternative solutions

According to constructivism:

”Social communication...enables agents to fix the meanings of material reality. When fixing meanings, agents select from ‘a horizon of possibilities’. In so doing they contribute to the institutionalization of practices and consequently to the unintentional survival of social structures.”<sup>10</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that ideas constrain communication and negotiation.

---

<sup>10</sup>Adler 2002: 102. And citations therein. Quote is from Mingers 1995: 157.



Consider the role and limits of communication in medieval English court process. When a complaint reached a court for litigation, the opposing attorneys were not always in consensus on the issue that was actually to be tried. Yet the judge needed them to be in consensus on it. So they were required to follow strict rules to reach consensus on it. The purpose of the rules was to reduce each case to a single question of fact that could be put to the jury for settlement based on its knowledge of the case. There were both formal rules and informal rules.

The formal rules required the opposing attorneys to resolve questions of law through a stylized pleading process. The plaintiff's attorney had to start the process by couching the complaint in terms of a form of action. The defendant's attorney had only three options in response:

1. He could deny the facts as alleged by the plaintiff. This would reduce the case to a question of fact that could be put to the jury.
2. He could accept the facts, but deny that his client had committed any violation of law. This would produce a question of law.
3. He could accept the facts and admit that his client had violated the law, but allege that it was justified in self-defense because the plaintiff had violated *another* law. This would start the whole process over again in reverse with a new complaint couched in terms of another form of action. In this case the original plaintiff's attorney would have the same three options in reverse in his new role as a defendant's attorney (for the original plaintiff).

Theoretically the process could go on *ad infinitum*. In practice it was limited to four rounds. (Each volley of each round had a name, perhaps to enable the judge to enforce the rules correctly, or perhaps simply to enable him to keep track of where the attorneys were in the process in case they got lost in negotiations.) So the process forced the attorneys to resolve questions of law and reduce the case to a clear factual issue.

The informal rules required the attorneys to resolve questions of law through bilateral negotiation rather than turning to the judge to do it. Judges were extremely reluctant to settle questions of law themselves. They much preferred the opposing attorneys to do it, for two reasons. First, if judges settled most questions of law one way or the other, they

likely would leave one of the opposing litigants in each trial unhappy with the trial verdict (e.g. feeling it was illegitimate). In a warlord society where the disgruntled litigant could easily continue his cause of action "out-of-doors" and start a private war against the other litigant, it was important not to leave litigants disgruntled with trial outcomes. By contrast, if the opposing attorneys settled most questions of law through bilateral negotiation, then at worst one litigant might be unhappy with how the *jury* had settled the *factual* questions. But neither litigant could fault the *judge* for how the questions of law had been settled, since their own attorneys had done that. Thus the preference of the judges *not* to settle questions of law themselves actually reinforced the legitimacy of the legal and trial system – a significant achievement at that early stage in its formation.

Second, judges knew that if they settled the questions of law themselves, over time they would establish many judicial precedents, some precedents might be inconsistent with others, and eventually the judicial system would be paralyzed by the inconsistencies. By contrast, if the opposing attorneys settled the questions of law, then each case would be shoehorned into one of the *existing* forms of action at trial and very few *new* precedents or forms of action would need to be created to reach trial verdicts.

When opposing attorneys negotiated to settle questions of law, it forced them to discriminate between different definitions of material interest and decide which definitions mattered most to their clients. By prioritizing interests in this way, they actually defined some material interests *into* the legal universe and others *out* of it (as explained above). Since the stakes were usually high on both sides, each side had to deploy a combination of threats, concessions, and persuasion if the two sides were to reach consensus on the questions of law. They were more likely to reach consensus if each could persuade the other to see the matter a little differently. Yet the scope for persuasion was still limited by hard material interests. The negotiated outcome had to meet both constraints: (1) prioritizing among competing interests through the creative use of existing definitions of interest, and (2) protecting each litigant's legitimate interests from illegitimate encroachment by the other litigant. Over the long run, such bargaining led to the emergence of a characteristic *way* of doing legal business in the central royal courts. It was an oral tradition that took practitioners years to learn – and that evolved in nature over decades and centuries. This was in part because opposing

attorneys made thousands of little agreements about which objective facts in the world to make legally visible and which ones to leave legally invisible. Each one of these decisions was essentially arbitrary (for the reasons explained earlier), so no newcomer could deduce it from first principles alone. The knowledge of this body of agreements that one needed to practice law could only be learned from deep immersion over years.

In sum, the communication between opposing attorneys at trial was constrained to follow fixed rules, so that the attorneys would settle all questions of law rather than asking the judge or jury to do it, and each case could be settled by the jury's verdict on factual questions alone. This system illuminates the three main roles of ideas in structuring communication for the purpose of forming mutual understandings and maintaining social consensus.

1. **Ideas as hypotheses.** The formal pleading rules embodied ideas – hypotheses – about the most efficient way to reduce a dispute to a tractable form through verbal communication.
2. **Ideas as definitions.** The need to settle questions of law through bilateral negotiation forced the opposing attorneys to rely on existing ideas – definitions – of material interests as much as possible.
3. **Ideas as decisions.** Ideas – decisions – brought some objective facts into social visibility and left others out.

In all of these ways, ideas constrained communication and negotiation over the definition of problems and the feasibility of alternative solutions. The ultimate value of reaching consensus at trial was that it would enable the opposing litigants to avoid a private war over the dispute. Thus the constraining role of ideas was driven at root by this transcendent material value.

The constructivist quote at the beginning of this subsection stated that communication is *either* about bargaining over instrumental material interests *or* it is about persuasion and blank. From the Contractual Realist perspective outlined here, however, it is not either/or. It is both/and.

### 5.2.6 The *institutional capacity to form contracts and fill ideational gaps in contracts* determines whether disputes will be settled by words or power

According to constructivism:

”The imposition of meanings on the material world is one of the ultimate forms of power.... When...someone uses guns or tanks, or makes a threat, it usually means that the ability to impose meanings, status or functions on physical objects by collective agreement has already failed.”<sup>11</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that peace and stability depend on the institutional capacity to form contracts and fill ideational gaps in contracts.

There are two main ways that breakdowns in this capacity cause conflict and competitions in raising relative power. First, when a novel contingency exposes a legal incompatibility between two terms of an anchor term contract (or norm or law), it gives each side uncertainty about whether the other side created the problem intentionally for its own gain. Then each side faces the risk that if it lets the other side gain anything from the dispute, the other side will have an incentive to create similar disputes (again?) in the future. So each side must deter the other side from committing such opportunism by imposing more costs on the other side than it might gain from the current dispute. Then it will not reap a net gain from the current dispute.

This mechanism of conflict initiation explains the outbreak and duration of the seven hegemonic wars in Europe between 1550 and 1950 (see Part V of this study).

Second, if conflicts and competitions in raising relative power are to be avoided, institutions must have three capacities:

1. the capacity to form new anchor term contracts and justice bargains;
2. the capacity to reform these understandings in response to novel contingencies; and
3. the capacity to resolve the legal incompatibilities between them that arise over time in a simple and uniform way.

---

<sup>11</sup>Adler 2002: 103.

There were wide variations in these institutional capacities across the regions of Western and Central Europe from 1000 to 1900. Consequently there were wide variations across these regions in the levels of conflict and competitions in raising relative power (Part IV).

### 5.2.7 Ideas *structure the practices* that constitute states' interactions

According to constructivism:

"Culture it is not only in people's minds, discourse, and interactions; it is also in the very performance of practices. From that perspective, practices not only organize the world – they are also the raw materials that make it up. When states face each other for myriad reasons, their interaction is affected, indeed constituted, not only by the cost-benefit analyses leaders make, the ideas and knowledge people carry in their heads, and the discourse they use to communicate. Rather, what states do in relation to other states...is constituted by the practices they share."<sup>12</sup>

Contractual Realism agrees with constructivism on this point. But Contractual Realism also explains *how and why* it is the case that ideas structure the practices, and practices constitute the interactions among states.

A "practice" is simply a distribution of rights and obligations to take certain actions and refrain from taking other actions. This distribution of rights and obligations is determined by an anchor term contract. Thus, when we say that *practices* constitute the interactions among actors, it is simply another way of saying that *anchor term contracts* constitute the interactions among actors. Therefore, if one wants to explain why practices remain stable at times but change at other times, the place to look for such explanations is in the anchor term contracts that determine the practices. And a good place to start is with the explanations of stability and change offered in the previous six subsections.

In the constructivist quote at the beginning of this subsection, the author wrote that

"interaction [among states] is affected, indeed constituted, not only by the cost-

---

<sup>12</sup>Adler 2013: 126. And citations therein.

benefit analyses leaders make, the ideas and knowledge people carry in their heads, and the discourse they use to communicate....[but also] by the practices they share.”<sup>13</sup>

Contractual Realism of excusably sees the relation between these four elements differently. The actors’ cost-benefit analyses are not separate from their practices. Rather, the practices *embody* the cost-benefit analyses, because the practices are determined by anchor term contracts. And these contracts are rationally motivated and structured by (1) the value of the material assets that the contracts divide among actors and (2) the value of dividing them on terms that are invariant to changes in relative power among actors (which is that it avoids wasteful competitions in raising relative power).

---

<sup>13</sup>Adler 2013: 126, and citations therein.

Part X

# Appendices

## CHAPTER 32

### A formal model of the world of power, conflict and high cost deterrence

When the world's goods are divided according to relative power, costly conflicts and competitions in raising relative power are unavoidable. The main focus of recent scholarship has been on costly conflicts. According to the standard argument, when actors are uncertain about relative power, conflicts are necessary to deter opportunism. One actor may bluff that his power is greater than it really is to gain a greater share of the world's goods from another. The only way for the potential victim to deter this kind of bluffing is to initiate conflict against the bluffer periodically to impose more costs on him than he would gain by such bluffing (Fearon 1995). This kind of conflict is costly for both sides. Although this is a serious problem, however, it is not the most fundamental problem that arises in a world where the goods are divided according to relative power. For this kind of conflict can be avoided by taking measures to eliminate the uncertainty about relative power. Even after this problem is solved, however, costly competitions in raising relative power still occur for a deeper reason.

A conflict is an interaction in which each actor imposes costs on the other to induce him to make concessions. Imposing costs on the opponent requires the deployment of specialized assets – ones that are only useful for prosecuting a conflict. In the case of violent conflicts, this means guns and ammunition. Even in non-violent conflicts, however, imposing costs on an opponent may require the use of specialized assets. For example, a labor union mounting a strike may need to have organizers who are specially trained in strategy and tactics for strikes. Whenever actors require specialized assets to engage in conflict, each actor faces a choice whether to invest in such assets, and if so, how much to invest. His investments constitute a *transformation cost*. This is the cost of transforming resources from their standard use



in civilian economic production into the specialized form they must take to be useful in a conflict. If either actor fails to make such investments in peacetime, then the peace will be unstable. That actor will be unable to deter the other actor from seizing some of his goods unilaterally.

To see this point, let us define “peace” as an agreement between two actors to respect a given distribution of a good between them.<sup>1</sup> If one actor appropriates some of the other’s goods, it is rational for the victim to retaliate only if his cost of investing in the specialized assets needed to retaliate is less than the value of the goods taken by the predator. If this investment cost is greater, then it is not worth retaliating. By implication, a predator can seize a victim’s goods without triggering a retaliation as long as the predator seizes less in value than it would cost the victim to invest in armaments and retaliate. The optimal strategy of the potential victim, therefore, is to maintain armaments in peacetime, that is, *before* any predations occur. Such investments reduce the amount of resources he must invest *following* a predation to amass his optimal retaliation capacity, thereby reducing his cost of retaliating. This reduces the value of goods the predator can seize without triggering retaliation, because the predator can only seize goods up to the value of the victim’s investment cost following a predation. Thus, by maintaining armaments in peacetime, the victim can reduce his losses to predators and obtain a higher income. The defender’s optimal strategy, in fact, is to make peacetime investments up to the point that the marginal cost of the investments is just equal to the marginal benefit in terms of the value of goods secured from predation. When both sides make such investments, however, the result is a costly competition in raising relative power. [ADD HERE]

When engagement in conflict requires specialized assets, therefore, costly competitions in raising relative power are unavoidable. They must occur simply for each actor to maximize his income in the face of predators who pose threats to his income. This problem arises even if actors are certain about each other’s power. The root cause of the problem is deeper. The root cause is simply the fact that actors are dividing the world’s goods according to relative power. As long as they continue to divide the goods by this method, such costly

---

<sup>1</sup>This definition covers many types of agreements, such as peace agreements between nations, constitutional contracts between branches of a government, employment contracts between firms and a workers, etc.

competitions will occur even if *no* other causes of costly conflicts or competitions in raising relative power are present. That is, even in a world where there are no shifts in relative power, no uncertainty about relative power, and no first-mover advantages in starting a conflict, there will still be costly competitions in raising relative power due simply to the problem of resource transformation costs. This problem alone makes a world in which the goods are divided according to relative power costly.

This conclusion is fundamental. For actors have many ways to solve the problems of shifting power, uncertainty about relative power, and first-mover advantages – and insure that they do not cause costly conflicts. Yet most of these solutions require specific investments – ones that are only useful for these purposes. Hence the better an actor becomes at solving these problems – as measured by the amount of resources he devotes to solving them – the higher his resource transformation costs will become. The higher these costs are, in turn, the more costly it is to live in a world where the goods are divided according to relative power – and the greater is the incentive to find a different way to divide the world's goods.

This chapter models a world in which the goods are divided according to relative power. The model isolates the conditions that cause costly conflicts and competitions in raising relative power and the magnitude of the costs involved. The next chapter models a world in which the goods are divided according to contracts based on benchmarks that are independent of relative power. That approach enables actors to avoid costly conflicts and competitions in raising relative power. The two models are then integrated into a single model to derive the boundary conditions under which actors are better off switching from one world to the other. These boundary conditions enable the derivation of a range of comparative statics implications and observable implications for empirical testing.

## **1. The basic model**

Two actors divide a good through a process of conflict based on relative power. Each actor imposes costs on the other to compel him to release some of the good in his possession. Each actor continues to impose costs on the other up to the point that his marginal benefit from the last unit of cost imposed (in terms of the amount of the good he receives from the

other) is equal to his marginal cost of imposing it. The outcome of this process is that the good becomes divided between the actors according to their relative allocations of resources to imposing costs on each other. For example, if both actors allocate the same amount of resources, then they divide the good equally. If one actor allocates more resources than the other, then the first receives a larger share of the good than the second.

To formalize these ideas, assume that there are two actors in the system, 1 and 2. They divide a good  $g$  through conflict. Each actor allocates an amount of resources  $f_i$  to the conflict, where  $i \in (1, 2)$ . Each actor's allocation is drawn from a pool of his own resources that is separate from the good  $g$ . Each actor captures a share of the good  $p_i$  that is directly proportional to his allocation  $f_i$  (given the other actor's allocation  $f_j$ ):

$$p_i = \frac{f_i}{f_i + f_j} \quad (32.1)$$

$$p_j = \frac{f_j}{f_i + f_j} = 1 - p_i. \quad (32.2)$$

The amount of the good that actor  $i$  wins in the conflict is equal to his share  $p_i$  times the total amount of the good available to be divided:

$$g_i = p_i g. \quad (32.3)$$

Actor  $i$ 's final income is the sum of his income from other sources,  $q_i$ , and the amount of the good that he wins,  $g_i$ , minus his total cost of allocating resources to winning it,  $c_i f_i$ :

$$y_i = q_i + g_i - c_i f_i. \quad (32.4)$$

The parameter  $c_i$  represents the slope of actor  $i$ 's cost schedule for allocating resources to the conflict. Each actor's power is measured by this marginal cost parameter  $c_i$ . The lower this parameter is, the more powerful the actor is.

Each actor allocates resources to the conflict up to the point that his marginal benefit from the last unit allocated is equal to his marginal cost of allocating it. The marginal benefit schedule is  $\frac{f_j}{(f_i + f_j)^2} g$ . The marginal cost schedule is  $c_i$ . Setting these expressions equal to

each other and solving for  $f_i$ , the result is:

$$f_i(f_j) = \frac{\sqrt{f_j g}}{\sqrt{c_i}} - f_j \quad \text{for } i \in (1, 2). \quad (32.5)$$

The same result can be obtained by taking the first derivative of actor  $i$ 's income,  $y_i$ , setting it equal to zero, and solving for  $f_i$ . This is actor  $i$ 's reaction function: his optimal allocation of resources to the conflict as a function of the other actor's allocation. By substituting actor  $j$ 's reaction function into actor  $i$ 's reaction function and solving for  $f_i$ , we obtain actor  $i$ 's optimal allocation of resources to the conflict as a function of the basic parameters:

$$f_i^* = \frac{c_j}{(c_i + c_j)^2} g \quad \text{for } i \in (1, 2). \quad (32.6)$$

When both actors adopt their optimal allocations,  $f_i^*$  and  $f_j^*$ , each actor's share of the good captured in the conflict is:

$$p_i^{**} = p_i(f_i^*, f_j^*) \quad (32.7)$$

$$= \frac{c_j}{c_i + c_j} \quad (32.8)$$

$$= 1 - p_j^{**}. \quad (32.9)$$

Each actor's income is:

$$y_{iw}^{**} = q_i + p_i^{**} g - c_i f_i^* \quad (32.10)$$

$$= q_i + \frac{c_j^2}{(c_i + c_j)^2} g. \quad (32.11)$$

This result is intuitive. The more powerful actor  $i$  is – that is, the lower his marginal cost  $c_i$  is – the higher his income is. The more powerful his adversary  $j$  is, the lower his income is. This result captures the most important feature of a world in which the good  $g$  is divided according to the actors' relative power: Each actor's income is directly proportional to his relative power.

So far the model assumes that the conflict is continuous. There is never any peace and never any peace dividend. Each actor bears his conflict cost continuously ( $c_i f_i^*$ ). If either

actor's marginal cost parameter  $c_i$  changes, then his optimal allocation  $f_i^*$  changes and the division of the good between the actors changes. Thus the distribution of the good between the two actors shifts continuously with every shift in relative power between them. The conflict can be terminated, however, and the actors can save their conflict costs, if they can agree to stop fighting and to respect the division of the good produced by the conflict to date. Under what conditions can this kind of peace agreement be reached and be stable?

## 2. Peace equilibrium with peace dividend

Suppose that the actors receive an amount  $g$  in each period  $t_0, t_1, \dots, t_\infty$ . Each actor has a choice in each period whether to remain at peace with a conflict allocation of  $f_i = 0$  or to arm to an allocation of  $f_i^*$  and initiate conflict. The actors choose simultaneously without knowledge of each other's choices. If neither actor arms and initiates conflict, then the amount  $g$  is distributed between the two actors according to their relative power and they save their conflict costs. Each actor's income per period is then:

$$y_{ip}^{**} = q_i + p_i^{**} g. \quad (32.12)$$

Let  $p_i^{**}$  be termed the status quo distribution and  $y_{ip}^{**}$  and  $y_{jp}^{**}$  be termed the status quo incomes.

Suppose that in period  $t_0$  actor  $i$  arms and initiates conflict. If actor  $j$ , ignorant of actor  $i$ 's choice, chooses to remain at peace in this period, then  $j$  cannot arm and impose costs on  $i$  in this period. But  $j$  can arm in the next period,  $t_1$ , and impose costs on  $i$  from that point forward. In period  $t_0$ , therefore, actor  $i$  reaps excess gains from actor  $j$  at the rate of  $p^{*o} - p^{**}$ , where

$$p^{*o} = p(f_i^*, 0) = 1 \quad (32.13)$$

$$p^{**} = p(f_i^*, f_j^*). \quad (32.14)$$

Actor  $i$ 's one-period gain is therefore:

$$b_i(f_i^*, 0) = (p^{*o} - p^{**})g = (1 - p^{**})g. \quad (32.15)$$

If  $j$  arms and starts to impose costs on  $i$  in the next period,  $t_1$ , then the actors engage in conflict at their equilibrium allocations  $(f_i^*, f_j^*)$  from that point until the actors settle for peace. As long as the conflict lasts, each actor has a choice in each period whether: (1) to offer terms to the other actor and return to peace on these terms if they are accepted, (2) to settle for peace on the terms offered by the other actor, or (3) to continue the conflict for another period.

How long would actor 2 need to prolong the conflict to leave actor 1 worse off having initiated a conflict at  $t_0$  than having remained at peace at the status quo distribution at  $t_0$ ? To answer this question, suppose that actor 2 continues the conflict for  $T_2$  periods and then offers to return to peace at the status quo distribution (with allocations of zero by both actors). Then actor 1's income from initiating a conflict at  $t_0$  and settling for peace at time  $T_2$  is:

$$Y_{1a} = \sum_{t=0}^1 d^t [y_{1p}^{**} + b_1] + \sum_{t=1}^{T_2} d^t [y_{1p}^{**} - c_1 f_1^*] + \sum_{t=T_2}^{\infty} d^t [y_{1p}^{**}] \quad (32.16)$$

$$= \sum_{t=0}^1 d^t [q_1 + p^{**}g + b_1] + \sum_{t=1}^{T_2} d^t [q_1 + p^{**}g - c_1 f_1^*] + \sum_{t=T_2}^{\infty} d^t [q_1 + p^{**}g]. \quad (32.17)$$

His income from remaining at peace at  $t_0$  is:

$$Y_{1p} = \sum_{t=0}^{\infty} d^t [y_{1p}^{**}] \quad (32.18)$$

$$= \sum_{t=0}^{\infty} d^t [q_1 + p^{**}g]. \quad (32.19)$$

If actor 2 prolongs the conflict for long enough to render  $Y_{1a}$  less than  $Y_{1p}$ , then actor 1 will be better off returning to peace at the status quo distribution at time  $T_2$  than continuing the conflict. And he will be deterred from initiating a conflict again. To derive the minimal duration of retaliation by actor 2 that is sufficient to restore peace and deterrence in this

way, set  $Y_{1a} = Y_{1p}$  and solve for  $T_2$ :

$$T_2^* = . \quad (32.20)$$

Notice that the lower  $c_1$  is, the *shorter* is the duration of actor 2's retaliation that will be sufficient to restore and maintain deterrence. That is, the more powerful actor 1 is, the *less* long actor 2 needs to retaliate. This result is counter-intuitive: it contradicts common intuitions about relative power. The logic behind the result is straightforward though. The more powerful actor 1 becomes, the higher his optimal allocation  $f_1$  is and the greater his income is. However, as his power increases (as  $c_1$  decreases), his one-period gain from attacking before actor 2 can arm and retaliate rises *less* quickly than his income from remaining at peace rises. So a shorter duration of retaliation is required to offset this gain. At the same time, his conflict costs during the retaliation period,  $c_1 f_1^2$ , rise *more* quickly than his income from remaining at peace rises (because his optimal allocation  $f_1$  increases, and so his conflict cost multiplier  $f_1^2$  increases by more than his marginal cost  $c_1$  decreases). For this reason, too, the more powerful actor 1 becomes, the shorter is the duration of retaliation by actor 2 that will be sufficient to restore deterrence.<sup>2</sup>

Provided that actor 2 is willing to retaliate for  $T_2 > T_2^*$ , he can disarm during peacetime ( $f_2^* = 0$ ), because he knows that he can always rearm and retaliate if actor 1 ever violates the status quo distribution  $p_i^{**}$ . The peace equilibrium is only stable, however, if actor 2 has

---

<sup>2</sup>Actor 1's income from an attack ( $Q_{1a}$ ) is greater than his income from remaining at peace at the status quo ( $Q_{1p}$ ) – at least for one period until actor 2 joins the fray. Both incomes increase as actor 1's power increases (e.g. as  $c_1$  decreases). Yet his income from an attack increases at a *lower rate* than his income from remaining at peace at the status quo. Hence his one-period *net* gain from an attack (before actor 2 joins the fray) shrinks in size as his power grows. Hence it does not take as long for actor 2 to impose sufficient costs to offset these gains and restore deterrence (other things being equal). Intuitively, the more powerful actor 1 is, the greater his share of the good and his income become even when both actors remain at peace at the status quo distribution. So his *excess* gain from attacking before actor 2 can arm and join the fray becomes relatively smaller by comparison. In short, the larger actor 1's share of the pie becomes as his relative power grows, the less *additional* pie he can get from attacking before the target can arm and retaliate.

The same intuition applies to the other side of the deterrence equation. Actor 1's income from remaining at peace at the status quo ( $Q_{1p}$ ) is greater than his income from the conflict period once actor 2 joins the fray ( $Q_{1w}$ ). Both incomes increase as actor 1 becomes more powerful (e.g. as  $c_1$  decreases). Yet the income from conflict increases at a *lower rate* than the income from remaining at peace at the status quo. Hence actor 1's income loss from the conflict period – compared to peace at the status quo – increases in size as actor 1's power grows. So again, it does not take as long for actor 2 to impose sufficient costs to offset actor 1's gains from the attack (other things being equal).

an incentive to retaliate for time  $T_2^*$ . His income from retaliating for  $T_2^*$  is:

$$Y_{2r}^* = \sum_{t=0}^1 d^t [(1 - p^{**})g - b_1] + \sum_{t=1}^{T_2^*} d^t [(1 - p^{**})g - c_2 f_2^*] + \sum_{t=T_2^*}^{\infty} d^t [p^{**}g]. \quad (32.21)$$

His income from remaining unarmed, not retaliating, and simply allowing actor 1 to continue reaping an excess gain of  $b_1$  in each period is:

$$Y_{2u} = \sum_{t=0}^{\infty} d^t [(1 - p^{**})g - b_1]. \quad (32.22)$$

The income from retaliating is greater than the income from allowing actor 1 to continue seizing  $b_1$  as long as actor 2 is not too weak, that is, as long as actor 2's conflict cost parameter is lower than a critical threshold:

$$c_{2max} = . \quad (32.23)$$

When  $c_2 < c_{2max}$ , actor 2 always has an incentive to retaliate for  $T_2 > T_2^*$  and the peace equilibrium is stable. Actor 2 can disarm during peacetime knowing that he can always rearm and retaliate if actor 1 ever violates the status quo. The equilibrium is symmetric, so the same is true for actor 1. As long as  $c_1 < c_{1max}$ , actor 1 always has an incentive to retaliate for  $T_1 > T_1^*$  if actor 2 ever violates the status quo and seizes a greater share of the good for himself. So actor 1 can disarm during peacetime too.

### 3. Costly conflicts due to uncertainty about relative power

Suppose that the actors are at the peace equilibrium with allocations of zero. For simplicity consider the symmetric case where  $c_1 = c_2$ . The status quo distribution is  $p^{**} = .5$  and the status quo incomes are:  $Q_{ip}^{**} = y_i + .5g$  for  $i \in (1, 2)$ . Now suppose that relative power shifts periodically in actor 1's favor. That is to say, his marginal cost parameter  $c_1$  decreases. How large would the shift have to be for actor 1 to be better off initiating a conflict than remaining at peace at the status quo distribution? Let  $\hat{c}_1$  be the level of  $c_1$  at which actor 1's income from initiating a conflict,  $Y_{1w}^{s*}$ , is just equal to his income from remaining at peace



at the status quo,  $Y_{1p}^{**}$ :

$$Y_{1w}^{s*} = Y_{1p}^{**} \quad (32.24)$$

$$p_1^{s*}g - \hat{c}_1\hat{f}_1^* = p_1^{**}g. \quad (32.25)$$

where  $\hat{f}_1^* = f_1^*(\hat{c}_1) =$ ,

$p_1^{s*} = p(\hat{f}_1^*, f_2^*)$ , and

$Q_{1w}^{s*} = Q_{1w}(p_1^{s*})$ .

The solution for  $\hat{c}_1$  is  $\frac{c_2}{2}$ . If  $c_1$  falls below  $\frac{c_2}{2}$ , then actor 1's income from initiating a conflict is greater than his income from remaining at peace at the status quo distribution *even if actor 2 retaliates forever*.

Suppose, then, that actor 1's marginal cost parameter shifts periodically from a level above this threshold ( $\bar{c}_1 > \hat{c}_1$ ) to a level below this threshold ( $\underline{c}_1 < \hat{c}_1$ ). First consider the case where actor 2 cannot observe the value of actor 1's marginal cost parameter  $c_1$ . That is, actor 2 does not know actor 1's power level. Then consider the case where actor 2 can observe  $c_1$ : he does know actor 1's power level.

If actor 2 cannot observe  $c_1$ , then actor 1 can bluff that  $c_1 = \underline{c}_1$  when actually  $c_1 = \bar{c}_1$  as a ploy to obtain more of the good than his power actually enables. To be precise, actor 1 can:

- (1) claim that  $c_1 = \underline{c}_1$ ,
- (2) demand that actor 2 accept a peace agreement that gives him only  $Q_{2w}(p_1^{s*}) + \epsilon$ , and
- (3) back the demand by a threat to fight forever if actor 2 does not submit.

The threat is credible when  $c_1 = \underline{c}_1$ , because at this level actor 1 is better off fighting forever than remaining at peace at the status quo distribution. Hence actor 2 is better off submitting to the demand than standing firm and fighting over it. For by submitting he receives  $\epsilon$  more than he would receive by fighting forever. In this case actor 1 receives  $Q_{1w}(p_1^{s*}) - \epsilon + PD$ , where PD signifies the sum of the two actors' peace dividends. When  $c_1 = \bar{c}_1$ , however, actor 1's threat is not credible. For at this level actor 1 would receive a higher income from remaining at peace at the status quo distribution than from issuing a demand and fighting over it. Hence actor 2 is better off standing firm and fighting. For if he fights for long enough, then actor 1's income from having initiated a conflict will be lower than his income from

having remained at peace at the status quo distribution. Thus actor 2 can restore peace and deterrence at the status quo distribution by fighting for long enough. If actor 2 were to submit to the demand in these cases, then he would be submitting unnecessarily. That is, he would be accepting a loss of income that he could avoid by standing firm and fighting instead.

If actor 2 submits when he does not know the value of  $c_1$ , therefore, he may be submitting unnecessarily for all he knows. For it may be the case that  $c_1 = \bar{c}_1$ . Submitting when he is ignorant of  $c_1$  would allow actor 1 to gain from bluffing that  $c_1 = \underline{c}_1$  when actually  $c_1 = \bar{c}_1$ . Such gains would give actor 1 an incentive to bluff repeatedly. The only way for actor 2 to deter such bluffing, therefore, is to stand firm and threaten to fight over *all* of actor 1's demands. How long would actor 2 have to fight to deter actor 1 from bluffing that  $c_1 = \underline{c}_1$ ? Actor 2 must fight for long enough that if  $c_1 = \hat{c}_1$ , then actor 1's income from bluffing and having to fight over it will be *no greater than* his income from having remained at peace at the status quo. To derive this duration, set  $c_1 = \hat{c}_1$  and suppose that actor 2 continues the conflict for  $T_{2s}$  periods before submitting to actor 1's demand. Then actor 1's income from bluffing and having to fight over it is:

$$Y_{1sw} = \sum_{t=0}^{T_{2s}} d^t [y_{1w}^{**}(\hat{c}_1)] + \sum_{t=T_{2s}}^{\infty} d^t [y_{1p}^{**}(\underline{c}_1)]. \quad (32.26)$$

His income from remaining at peace at the status quo is:

$$Y_{1p} = \sum_{t=0}^{\infty} d^t [y_{1p}^{**}(\bar{c}_1)]. \quad (32.27)$$

Actor 2 must prolong the conflict for long enough to render  $Y_{1sw}$  less than  $Y_{1p}$ . Setting  $Y_{1sw} = Y_{1p}$  and solving for  $T_{2s}$ , we obtain:

$$T_{2s}^* = \frac{\ln\left(\frac{1-d}{d}\right)}{\ln d} \quad (32.28)$$

If  $d < 1$ , then  $T_{2s}^* < \infty$ . For example, if  $d = .9$ , then  $T_{2s}^* = 21$ . Provided that actor 2 is willing to fight for  $T_{2s}^*$  and actor 1 knows it, then whenever  $c_1 > \hat{c}_1$ , actor 1 will be better off remaining at peace at the status quo distribution than bluffing. In this way, actor 2's

willingness to fight for  $T_{2s}^*$  whenever actor 1 issues a demand will deter actor 1 from bluffing that  $c_1 = \underline{c}_1$  when actually  $c_1 = \bar{c}_1$ .

Formally, the following strategies constitute a perfect Bayesian equilibrium in which actor 1 is deterred from bluffing by actor 2's willingness to fight over all of actor 1's demands.

**Actor 2's strategy.** Whenever actor 1 issues a demand, stand firm and threaten to fight over it if actor 1 does not back down and retract the demand. If actor 1 responds by initiating conflict, fight for  $T_2 = T_{2s}^*$  and then submit to the demand.

**Actor 1' strategy.** When  $c_1 = \bar{c}_1$ , do not issue a demand; remain at peace at the status quo distribution. When  $c_1 = \underline{c}_1$ , issue a demand backed by a threat to fight forever if it is not met. If it is not met, then initiate conflict and continue to fight until actor 2 submits to the demand.

In this equilibrium there are two outcomes. When  $c_1 = \underline{c}_1$  (probability  $s$ ), actor 1 initiates conflict, actor 2 retaliates for  $T_2 = T_{2s}^*$ , and then they return to peace. When  $c_1 = \bar{c}_1$  (probability  $1 - s$ ), actor 1 does not initiate conflict. The actors remain at peace at the status quo distribution. The actors' incomes in this equilibrium are:

$$Q_1^{LI} = \frac{d}{1-d} \left[ .5g + s \left( \frac{1}{2\underline{c}_1} - \frac{1}{c_2} \right) \frac{g^2}{k_p^2} \right] \quad (32.29)$$

$$Q_2^{LI} = \frac{d}{1-d} \left[ .5g - s \left( \frac{1}{\underline{c}_1} - \frac{1}{2c_2} \right) \frac{g^2}{k_p^2} \right] \quad (32.30)$$

where LI stands for *limited information*. To summarize, actor 2 can deter actor 1 from bluffing that he is more powerful than he really is by being willing to fight. The result is periodic costly conflicts, however. To assess these costs, consider the alternative situation where actor 2 can observe the value of actor 1's marginal cost parameter.

When actor 2 knows that  $c_1 = \bar{c}_1$ , he can fight in response to any demands by actor 1. So actor 1 will be deterred from making demands (as before), and the actors can remain at peace at the status quo distribution with allocations of zero. In turn, when actor 2 knows that  $c_1 = \underline{c}_1$ , he can submit to the demand peacefully. In these cases actor 1 will always make the demand, and the new distribution will be:  $([Q_{1w}(p^{s*}) - \epsilon + PD], [Q_{2w}(p^{s*}) + \epsilon])$ . Thus, actor 1 only issues demands when  $c_1 = \underline{c}_1$ , as in the limited information equilibrium. But now actor 2 submits to these demands peacefully rather than fighting over them. So

costly conflicts are avoided. The actors' incomes are now:

$$Q_1^{FI} = \frac{d}{1-d} \left[ .5g + s \left( \frac{1}{c_1} - \frac{1}{c_2} \right) \frac{g^2}{k_p^2} \right] \quad (32.31)$$

$$Q_2^{FI} = \frac{d}{1-d} \left[ .5g - s \left( \frac{1}{c_1} - \frac{1}{c_2} \right) \frac{g^2}{k_p^2} \right] \quad (32.32)$$

where FI stands for *full information*. By moving from the limited information equilibrium to the full information equilibrium, each actor saves his conflict costs. The actors' savings are:

$$Q_1^{SAV} = Q_1^{FI} - Q_1^{LI} \quad (32.33)$$

$$= s \frac{d}{1-d} \left( \frac{1}{2c_1} \right) \frac{g^2}{k_p^2} \quad (32.34)$$

$$Q_2^{SAV} = Q_2^{FI} - Q_2^{LI} \quad (32.35)$$

$$= s \frac{d}{1-d} \left( \frac{1}{c_2} \right) \frac{g^2}{k_p^2}. \quad (32.36)$$

In sum, certainty about relative power enables the two actors to avoid costly conflicts. But does this certainty enable them to avoid costly competitions in raising relative power?

#### 4. Costly competitions in raising relative power

The basic model assumes that it is costly to *use* resources in a conflict, but costless to *transfer* resources into conflictive use at the start of a conflict and back into productive use at the end of a conflict. When the transfers are costless, the peace equilibrium is stable as long as there is certainty about relative power (as shown above). When the transfers are costly, however, the peace equilibrium is unstable even where there is certainty about relative power. Actor *i* can seize some of the good from actor *j* unilaterally as long as actor *i* takes less than it would cost actor *j* to transfer resources into conflictive use to retaliate. Actor *i* can continue to violate the status quo distribution of the good in this way as long as he doesn't do it for so long that actor *j* would be better off incurring the transformation cost to retaliate. The larger is actor *j*'s optimal allocation of resources to fighting ( $f_2^*$ ), the more arms he needs to transfer into conflictive use to retaliate optimally (e.g. by the least-cost method), and thus

the higher his total transformation cost is. The larger  $f_2^*$  is, therefore, the more of the good actor  $i$  can seize (or the longer he can continue to seize a given amount of the good) before he pushes actor  $j$  to the point where he is better off retaliating than remaining at peace and allowing the predations to continue.

The only way for actor  $j$  to deter the predations is to produce and maintain arms in peacetime. The more arms he maintains in peacetime, the less resources he needs to transfer into conflictive use to retaliate optimally following a predation, and thus the lower his total transformation cost is. If actor  $j$  maintains enough arms in peacetime, then his total transformation cost will be low enough that he is better off arming fully and retaliating optimally after a predation than remaining at peace and allowing the predations to continue. The minimum amount of arms that actor  $j$  needs to maintain in peacetime to retain his incentive to retaliate depends on the trade-off between his cost of maintaining arms in peacetime and his benefit from reducing predation by actor  $i$ .<sup>3</sup>

To formalize these ideas, I present a reduced form of the model analyzed above. The reduced form captures the essential features of the basic model while also yielding closed form solutions. In each period the actors receive a total amount of the good equal to  $2g$ . They divide it either by conflict or by a peace agreement based on the distribution that a conflict would produce. Each actor allocates an amount of resources  $f_i$  to conflictive use from a pool of his own resources separate from the good  $g$ . Each actor pays a marginal cost  $c_i$  for each unit of resources he allocates to conflictive use. This is the opportunity cost in terms of output lost from the productive economy as long as the resources are allocated to conflictive use. The actors are assumed to be equal in power:  $c_i = c_j = c$ . The status quo distribution of the good is thus an even split:  $g_i = g_j = g$ .

Each actor's optimal fighting allocation is normalized to one ( $f_i^* = f_j^* = 1$ ). Each actor chooses a fraction  $a_i$  of his optimal fighting allocation to maintain as arms in peacetime, where  $0 \leq a_i \leq 1$ . Due to the normalization, the *fraction* of his optimal allocation that he maintains as arms in peacetime also represents the *amount* of arms that he maintains

---

<sup>3</sup>This is an example of a Stackelberg leadership strategy. By sinking an investment in armaments in peacetime, actor  $j$  lowers his marginal cost of retaliating once a predation has occurred, thereby increasing his retaliation capacity. The underlying economic principle is that previously sunk costs do not affect current or future decisions. See Tirole (1988: PP) for the standard analysis of inter-firm competition in a duopoly setting, where each firm can invest in a fixed asset that lowers its marginal production cost later.

in peacetime ( $a_i = f_{ia}$ ). And the *fraction* that he needs to mobilize to enter a conflict also represents the *amount* of arms that he needs to mobilize to enter a conflict ( $1 - a_i = f_{ir}$ ). The subscript  $r$  stands for *ramping up* and the subscript  $a$  stands for *arms in peacetime*. Thus if actor  $i$  maintains no arms in peacetime, his income per period in peacetime is  $y_{ip} = g$ . If he maintains a fraction  $a_i$  of his optimal fighting allocation as arms in peacetime, then his income per period in peacetime is  $y_{ip}(a_i) = g - ca_i = g - cf_{ia}$ .

Each actor pays a marginal transformation cost of  $c_m$  per unit of resources that he transforms from productive use into conflictive use. This parameter reflects the technology for transforming resources from productive use into conflictive use and back again. The parameter is assumed to be the same for both actors. Actor  $i$ 's *total* transformation cost is proportional to the amount of resources that he transforms into conflictive use:  $C_{im}^{tot} = c_m f_i$ . His total transformation cost is paid on the date that he transforms  $f_i$  resources into conflictive use. For example, if actor  $i$  maintains no arms in peacetime, then he pays the entire total transformation cost on the date that he transforms  $f_i^* = 1$  resources into armaments to enter a conflict. Alternatively, if he maintains a fraction  $a_i$  of his optimal fighting allocation as arms in peacetime, then he pays a fraction  $a_i$  of the total transformation cost in peacetime to produce and maintain these arms and the remaining fraction  $1 - a_i$  on the date that he ramps up fully to his optimal fighting allocation to enter a conflict.

As before, each actor faces a choice in each period whether to remain at peace at the status quo distribution or to arm to his optimal fighting allocation and initiate a conflict. The actors choose simultaneously in each period, so neither actor knows the other actor's choice in the current period. If actor  $i$  arms fully and initiates conflict on date 0 but actor  $j$  does not, then actor  $i$ 's excess gain in this period depends on the amount of arms that actor  $j$  maintains in peacetime. The more arms he maintains, the smaller is actor  $i$ 's excess gain in period 0 and the smaller is actor  $j$ 's loss in period 0. Formally, if actor  $i$  arms fully and initiates conflict on date 0 but actor  $j$  does not, then the actors incomes in period 0 are:

$$y_{ix} = g - c + (1 - a_j)g = g - c + f_{jr}g \quad (32.37)$$

$$y_{jx} = g - cf_{ja} - (1 - a_j)g = g - cf_{ja} - f_{jr}g. \quad (32.38)$$

Actor  $j$  then chooses on date 1 whether to arm fully and retaliate or to remain at peace at

his peacetime arms level,  $f_{ja}$ , while actor  $i$  continues his predation. If actor  $j$  chooses to arm fully and retaliate on date 1, then he pays a fraction  $1 - a_i$  of his total transformation cost on this date. Then each actor's income is  $y_w = g - c$  for as many periods as actor  $j$  retaliates. He retaliates for long enough to insure that actor  $i$  loses more from the retaliation than he gained from his predation starting on date 0. Then the actors return to peace at the status quo distribution. Alternatively, if actor  $j$  chooses to remain at peace on date 1, then actor  $i$  continues his predation for the same length of time that actor  $j$  would have retaliated had he chosen to retaliate. Then actor  $i$  stops predating and both actors return to peace at the status quo distribution.

The analysis proceeds in three stages. First, I show that when the marginal transformation cost  $c_m$  is zero, the peace equilibrium is stable even if neither actor maintains any arms in peacetime (section 3.4.1). This result demonstrates that the reduced form of the model retains the essential features of the basic model analyzed earlier. Second, I show that when the marginal transformation cost is greater than zero, the peace equilibrium is unstable if either actor fails to maintain enough arms in peacetime (section 3.4.2). Third, I show that the peace equilibrium becomes stable again as long as both actors maintain enough arms in peacetime (section 3.4.3).

#### 4.1 unarmed peace equilibrium is stable if $c_m = 0$

Suppose the transformation cost  $c_m$  is zero and the actors are at peace unarmed ( $a_i = a_j = 0$ ). On date 0 actor  $i$  arms fully and initiates conflict but actor  $j$  does not. On date 1 actor  $j$  arms fully and starts to retaliate. He continues to retaliate for  $T_j$  periods and then both actors return to peace unarmed at the status quo distribution. How long does actor  $j$  need to retaliate to leave actor  $i$  worse off having initiated a conflict at date 0 than having remained at peace at the status quo distribution? Actor  $i$ 's income from initiating conflict at date 0

and returning to peace at time  $T_j$  is:

$$Y_{ix} = \sum_{t=0}^1 d^t [y_{ix}] + \sum_{t=1}^{T_j} d^t [y_w] + \sum_{t=T_j}^{\infty} d^t [y_p] \quad (32.39)$$

$$= \sum_{t=0}^1 d^t [2g] + \sum_{t=1}^{T_j} d^t [g - c] + \sum_{t=T_j}^{\infty} d^t [g]. \quad (32.40)$$

His income from remaining at peace on date 0 is:

$$Y_{ip} = \sum_{t=0}^{\infty} d^t [y_p] \quad (32.41)$$

$$= \sum_{t=0}^{\infty} d^t [g]. \quad (32.42)$$

Setting  $Y_{ix} = Y_{ip}$  and solving for  $T_j$ , we obtain:

$$T_j^* = . \quad (32.43)$$

If actor  $j$  is willing to retaliate for  $T_j$  and actor  $i$  knows it, then  $i$  will be better off remaining at peace at the status quo distribution on date 0 than arming fully and initiating conflict. So he will be deterred from initiating conflict.

Actor  $j$  can disarm completely during peacetime as long as he is willing to retaliate for  $T_j > T_j^*$ , just as before, because he knows that he can always arm fully and retaliate if actor  $i$  ever violates the status quo distribution. The peace equilibrium is only stable if actor  $j$  has an incentive to retaliate for time  $T_j^*$ . His income from retaliating for  $T_j^*$  is:

$$Y_{jr} = \sum_{t=0}^1 d^t [y_{jx}] + \sum_{t=1}^{T_j^*} d^t [y_w] + \sum_{t=T_j^*}^{\infty} d^t [y_p] \quad (32.44)$$

$$= \sum_{t=0}^1 d^t [0] + \sum_{t=1}^{T_j^*} d^t [g - c] + \sum_{t=T_j^*}^{\infty} d^t [g]. \quad (32.45)$$

His income from remaining unarmed, not retaliating, and allowing actor 1 to continue pre-



dating for  $T_j^*$  periods is:

$$Y_{jc} = \sum_{t=0}^{T_j^*} d^t [y_{jx}] + \sum_{t=T_j^*}^{\infty} d^t [y_p] \quad (32.46)$$

$$= \sum_{t=0}^{T_j^*} d^t [0] + \sum_{t=T_j^*}^{\infty} d^t [g]. \quad (32.47)$$

The income from retaliating is greater than the income from capitulating as long as actor  $j$  is not too weak, that is, as long as actor  $j$ 's conflict cost parameter is lower than a critical threshold:

$$c_{j,max} = g. \quad (32.48)$$

As long as  $c_j < c_{j,max}$ , actor  $j$  always has an incentive to retaliate for  $T_j > T_j^*$  and the peace equilibrium is stable. Actor  $j$  can disarm during peacetime knowing that he can always rearm and retaliate if actor  $i$  ever violates the status quo. The equilibrium is symmetric, so the same is true for actor  $i$ . As long as  $c_i < c_{i,max}$ , actor  $i$  always has an incentive to retaliate for  $T_i > T_i^*$  if actor  $j$  ever violates the status quo and seizes a greater share of the good for himself. So actor  $i$  can disarm during peacetime too.

#### 4.2 unarmed peace equilibrium is unstable if $c_m \geq 0$

Suppose the transformation cost  $c_m$  is greater than zero. Assume the actors are at peace and actor  $j$  maintains no arms in peacetime ( $a_j = 0$ ). Again actor  $i$  arms fully and initiates conflict on date 0 but actor  $j$  does not. Then actor  $j$  arms fully on date 1 and retaliates for  $T_j$  periods. Now the minimum duration of retaliation sufficient to maintain deterrence is:

$$T_{j0}^* = . \quad (32.49)$$

This is simply  $T_j^*$  with  $a_j = 0$ . As long as actor  $j$  is willing to retaliate for  $T_{j0}^*$  and actor  $i$  knows it, then actor  $i$  will be deterred from arming and initiating conflict. However, if the marginal transformation cost  $c_m$  is too high, then actor  $j$  is better off remaining at peace and allowing actor  $i$ 's predations to continue for  $T_{j0}^*$  than arming and retaliating for  $T_{j0}^*$ . His

income from remaining at peace is:

$$Y_{j0c} = \quad (32.50)$$

His income from arming and retaliating is:

$$Y_{j0r} = \quad (32.51)$$

Setting  $Y_{j0c} = Y_{j0r}$  and solving for  $c_m$ , we have:

$$c_m^{j,max} = \quad (32.52)$$

If  $c_m > c_m^{j,max}$ , then actor  $j$  is better off remaining at peace and permitting the predations to continue than arming and retaliating.

At the same time, however, if  $c_m$  is too much higher than  $c_m^{j,max}$ , then actor  $i$  is better off remaining at peace at the status quo distribution than arming and predating for  $T_{j0}^*$  – even if he knows that actor  $j$  will permit the predations. Actor  $i$ 's income from remaining at peace at date 0 is  $Y_{ip}$  (equation 32.41). His income from arming at date 0 and predating for  $T_{j0}^*$  (knowing that actor  $j$  will permit the predations) is:

$$Y_{ixd} = \sum_{t=0}^{T_j} d^t [y_{ix}] + \sum_{t=T_j}^{\infty} d^t [y_p] \quad (32.53)$$

$$= \sum_{t=0}^{T_j} d^t [2g] + \sum_{t=T_j}^{\infty} d^t [g]. \quad (32.54)$$

Setting  $Y_{ip} = Y_{ixd}$  and solving for  $c_m$ , we have:

$$c_m^{i,max} = \quad (32.55)$$

As long as  $c_m < c_m^{i,max}$ , actor 1 is better off arming and predating at date 0 than remaining at peace at the status quo distribution (when he knows that  $c_m > c_m^{j,max}$ , and therefore actor  $j$  will permit the predations).

Actor  $i$ 's threshold,  $c_m^{i,max}$ , is always higher than actor  $j$ 's threshold,  $c_m^{j,max}$ . Figure X

shows the difference between the actors' thresholds ( $c_m^{i,max} - c_m^{j,max}$ ) for a range of values of  $g$ ,  $c$ , and  $a_i$ . Notice that actor  $i$ 's threshold is often considerably higher than actor  $j$ 's. There is thus a wide range of marginal transformation costs within which  $c_m$  is high enough that actor  $j$  is better off capitulating than arming and retaliating ( $c_m > c_m^{j,max}$ ), but  $c_m$  is low enough that actor  $i$  is better off arming and predating than remaining at peace at the status quo distribution ( $c_m < c_m^{i,max}$ ).<sup>4</sup> This is the range in which an unarmed peace equilibrium is *unstable*. If actor  $j$  disarms completely ( $a_j = 0$ ), then he will not have an incentive to arm and retaliate in the event of predations by actor  $i$ . So deterrence is impossible: Actor  $i$  will have an incentive to arm and predate. The analysis is symmetric: If actor  $i$  disarms completely, the peace equilibrium is unstable for the same reason.

Suppose that actor  $j$  disarms completely. There is a minimum threshold to which actor  $i$  must arm in peacetime to push actor  $j$  to the point that has no incentive to retaliate. Figure X shows this minimum threshold,  $a_i^{min}$ , for a range of values of  $g$ ,  $c$ , and  $c_m$ . Notice that  $a_i^{min} = 0$  for a wide range of parameter values. Within this range, if actor  $j$  disarms completely, the peace is unstable even if actor  $i$  maintains no arms in peacetime. He has an incentive to go straight from zero armament to his optimal fighting allocation and initiate conflict. For the remaining parameter values, the peace is stable as long as actor  $i$  maintains few or no arms ( $a_i < a_i^{min}$ ). If actor  $j$  remains unarmed, however, then actor  $i$  has an incentive to arm to  $a_i^{min}$  in peacetime, at which the peace becomes unstable. In this sense the peace is unstable even when  $a_i = 0$ .

### 4.3 armed peace equilibrium is stable even if $c_m \geq 0$

Actor  $j$  can render the peace stable by maintaining enough arms in peacetime. By maintaining arms he lowers the amount of resources he needs to transfer into conflictive use to retaliate optimally in the event of predations by actor  $i$ . This in turn lowers actor  $j$ 's total transformation cost of retaliating ( $C_{2m}^{tot}$ ). If actor  $j$  maintains enough arms in peacetime, he can lower his total transformation cost by enough that he will be better off retaliating than remaining at peace and permitting the predations to continue. If both actors maintain enough arms in peacetime, then the peace equilibrium is stable. Neither actor will have an

---

<sup>4</sup>Figure X shows this range of values of  $c_m$  for plausible values of the other parameters of the model.

incentive to arm fully and engage in predation, because each knows that the other has an incentive to retaliate effectively.

Suppose that both actors maintain at least some arms in peacetime:  $0 < a_i \leq 1$  for  $i = 1, 2$ . Suppose again that actor  $i$  arms fully and initiates conflict on date 0 but actor  $j$  does not. Then on date 1 he arms fully and retaliates for  $T_j$  periods. Actor  $i$ 's income from initiating a conflict at date 0 and returning to peace at time  $T_j$  is:

$$Y_{ix} = \sum_{t=0}^1 d^t [y_{ix}] + \sum_{t=1}^{T_j} d^t [y_w] + \sum_{t=T_j}^{\infty} d^t [y_p] \quad (32.56)$$

$$= \sum_{t=0}^1 d^t [g - c + (1 - a_j)g] + \sum_{t=1}^{T_j} d^t [g - c] + \sum_{t=T_j}^{\infty} d^t [g]. \quad (32.57)$$

His income from remaining at peace on date 0 is:

$$Y_{ip} = \sum_{t=0}^{\infty} d^t [y_p] \quad (32.58)$$

$$= \sum_{t=0}^{\infty} d^t [g]. \quad (32.59)$$

Setting  $Y_{ix} = Y_{ip}$  and solving for  $T_j$ , we have:

$$T_{ja} = . \quad (32.60)$$

If actor  $j$  is willing to retaliate for  $T_{ja}$  and actor  $i$  knows it, then  $i$  will be better off remaining at peace at the status quo distribution on date 0 than arming fully and initiating a conflict. So he will be deterred from initiating conflict.

If actor  $j$  is to have an incentive to retaliate for  $T_{ja}$ , however, he must maintain at least some arms in peacetime. His income from retaliating for  $T_{ja}$  is:

$$Y_{jar} = \quad (32.61)$$

His income from remaining at peace unarmed and allowing the predations to continue for

$T_{ja}$  is:

$$Y_{j0c} = \quad (32.62)$$

Setting  $Y_{jar} = Y_{j0c}$  and solving for  $a_j$ , we have:

$$a_j(T_j) = . \quad (32.63)$$

Substituting  $T_{ja}$  into this equation, we obtain actor  $j$ 's reaction function for peacetime armament:

$$a_j(a_i) = . \quad (32.64)$$

This is the minimum fraction of actor  $j$ 's optimal fighting allocation that he must maintain in peacetime to retain an incentive to retaliate, given actor  $i$ 's peacetime armament fraction,  $a_i$ . Actor  $i$ 's reaction function is symmetric:

$$a_i(a_j) = . \quad (32.65)$$

Substituting  $a_i(a_j)$  into equation 32.64 and solving for  $a_j$ , we obtain the minimum peacetime armament fraction that is sufficient for actor  $j$  to maintain an incentive to retaliate, as a function of the basic parameters:

$$a_j^* = . \quad (32.66)$$

As long as  $a_j \geq a_j^*$ , actor  $j$  is at least as well off arming fully and retaliating in response to predations by actor  $i$  as he is remaining at peace unarmed and permitting the predations to continue. As long as actor  $i$  knows that  $a_j \geq a_j^*$ , then he will be deterred from arming fully and engaging in predation. The peace equilibrium will be stable.

When actor  $j$  maintains the fraction  $a_j^*$  in peacetime, his income is:

The difference between this income and his income from remaining unarmed and permitting the predations to continue (eqn. ??) is:

This is the amount of income that actor  $j$  saves by maintaining arms in peacetime. This savings comes at a cost, however. Actor  $j$ 's income when there are transformation costs (eqn. ??) is lower than his income when there are no transformation costs (eqn. ??). The difference between these incomes is:

This is the cost of maintaining peace in a world where the goods are divided according to relative power and it is costly to transfer resources between productive and conflictive use.

These results underscore the most important features of a world in which the goods are divided according to relative power. First, the only way for an actor to maximize his income in this world is to amass and maintain enough power assets to deter others from preying on his goods. Second, each actor's income saving from amassing power assets is directly proportional to his underlying power potential (defined in terms of his marginal cost parameter  $c_i$ ). Third, even when certainty about relative power enables actors to avoid costly conflicts, there are still heavy costs associated with the process of dividing the world's goods according to relative power.

The conclusion of this chapter's analysis is clear. The most fundamental problem that actors face in a world where the goods are divided according to relative power is not the problem of costly conflicts due to uncertainty about relative power. It is the problem of costly competitions in raising relative power which is created by resource transformation costs. The higher these transformation costs become, the more costly it is to live in this kind of world, and the greater the income savings from switching to a world where the goods are divided by some other method. The model in this chapter thus predicts that as transformation costs become higher, and conflict assets become more specialized and more expensive to amass, actors will search more intensively for a different method of dividing the world's goods.

## CHAPTER 33

### A formal model of oscillations between the world of law and the world of power

In the previous chapter I analyzed the world of power. In this chapter I analyze the oscillation between the world of law and the world of power. I present a model in which two actors divide an asset according to the contractual method. The contract is enforced by one of two technologies: anchor based enforcement or power based enforcement. Anchor based enforcement is sufficient as long as the actors have common knowledge of the meaning of the contract terms. If there is a breakdown in common knowledge of the meaning of the terms, however, then anchor based enforcement fails. The actors then revert to power based enforcement. This logic explains why legal incompatibility disputes cause reversion from the world of law to the world of power. Such disputes cause a breakdown in common knowledge of the meaning of contract terms.

The basic set up is shown in figure 33.1. Two actors divide an asset according to the contractual method: they form an anchor contract. The total amount of the asset to be divided is  $2g$ . Assume that the actors divide it equally, so each actor receives  $g$ . Let this be called the status quo distribution. With probability  $v$ , the anchor shifts in actor 1's favor by an amount  $b$  due to random forces beyond his control (e.g. unintentionally on his part). The contract specifies that when this occurs, actor 1 receives  $g+b$  and actor 2 receives  $g-b$ . This is actor 1's "random variable benefit"  $b$ . With probability  $1-v$ , the anchor does not shift in actor 1's favor due to random forces. The contract specifies that at these times the status quo distribution is maintained  $(g, g)$ . At these times actor 1 has an opportunity to move the anchor in his favor intentionally to gain the amount  $b$  at actor 2's expense in violation of the contract terms.

I use this set up to analyze a series of three games. In the first game, actor 2 can tell

whether the anchor moved in actor 1's favor due to random forces beyond his control or his intentional actions. In the second game, actor 2 sometimes *cannot* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions – and actor 1 *knows* whether or not actor 2 can tell. In the third game, sometimes actor 2 cannot tell, but actor 1 does *not* know whether or not actor 2 can tell. The information structure of these games is the same as that in the game tree at the end of the next chapter.<sup>1</sup>

## 1. Game 1: two-sided full information

In this game, actor 2 can tell whether the anchor moved in actor 1's favor due to random forces beyond his control or his intentional actions. When the anchor moves in his favor due to random forces beyond his control, actor 2 grants him the random variable benefit  $b$ . But when actor 1 moves the anchor in his favor intentionally, actor 2 imposes an anchor based punishment of value  $b$  (e.g. keeping the random variable benefit for himself rather than granting it to actor 1). The threat of this punishment is sufficient to deter actor 1 from moving the anchor in his favor intentionally. In equilibrium, he does not move the anchor in his favor intentionally, and he receives the random variable benefit  $b$  whenever the anchor moves in his favor due to random forces beyond his control. In this game the actors' incomes are:

$$Q_{1j}^* = g + vb \quad (33.1)$$

$$Q_{2j}^* = g - vb. \quad (33.2)$$

## 2. Game 2: one-sided limited information

In this game, actor 2 sometimes *cannot* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions – and actor 1 *knows* whether or not actor 2 can tell.

To analyze this game, suppose that with probability  $w$ , actor 2 *can* tell whether the

---

<sup>1</sup>These games are based on my first model of imperfect common knowledge (Koppel 2006a, 2006b).



anchor moved in actor 1's favor due to random forces beyond his control or his intentional actions. With probability  $1 - w$ , actor 2 *cannot* tell. Actor 1 can observe which is the case.

First, suppose actor 2 *can* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions – and actor 1 observes this fact. In this case (probability  $w$ ), the play and outcome of the game are the same as in game 1 above. Thus, actor 2 grants actor 1 the random variable benefit with probability  $v$ .

Second, suppose actor 2 *cannot* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions – and actor 1 observes this fact. In this case (probability  $1 - w$ ), actor 1 has scope to commit opportunism. Actor 1 can move the anchor in his favor intentionally, but bluff that it moved in his favor due to random forces beyond his control. To deter such opportunism, actor 2 must reject the claim and impose the anchor based punishment *whenever* the anchor moves in actor 1's favor for either reason (e.g. randomly or intentionally – actor 2 cannot tell). Again, the threat of this punishment is sufficient to deter actor 1 from moving the anchor in his favor intentionally. But when the anchor moves in his favor due to random forces beyond his control and he claims the random variable benefit  $b$ , actor 2 rejects the claim and imposes an anchor based punishment of value  $b$  (e.g. keeping the random variable benefit for himself rather than granting it to actor 1). Actor 1 can accept this unjust punishment without resistance, because he *knows* that actor 2 cannot tell whether the anchor moved in his favor randomly or intentionally. In this case (probability  $1 - w$ ), actor 1 never receives the random variable benefit – not even when the anchor moves in his favor due to random forces beyond his control.

In this game, the actors' incomes are:

$$Q_{1u}^* = g + wvb \quad (33.3)$$

$$Q_{2u}^* = g - wvb. \quad (33.4)$$

Compared to the outcome of the previous game, actor 1 loses his random variable benefit a fraction of the time  $v(1 - w)$ . This loss is due to the one-sided limited information. There is no need for power based punishment or costly conflict in this game, however, because actor 1 *knows* when the anchor based punishment he receives is justified. That is, he knows when

actor 2 *cannot* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions.

### 3. Game 3: two-sided limited information

In this game, actor 2 *cannot* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions (as in the previous game). But now, actor 1 *does not know* whether or not actor 2 can tell. This uncertainty gives actor 2 scope to commit opportunism. Actor 2 can bluff that he cannot tell whether the anchor moved in actor 1's favor randomly or intentionally – even when he really *can* tell – in order to keep the random variable benefit for himself when it should go to actor 1 according to the contract terms.

To analyze this game, suppose again that with probability  $w$ , actor 2 *can* tell whether the anchor moved in actor 1's favor due to random forces or his intentional actions. With probability  $1 - w$ , actor 2 *cannot* tell. But now, suppose that actor 1 does not know which is the case; he only knows the probability  $w$ . Actor 2 has two options. He can act honestly – never bluffing that he cannot tell when he really can tell. Or he can act opportunistically – bluffing that he cannot tell when he really can tell. Consider each option in turn.

First, actor 2 can act honestly: never bluffing that he cannot tell when he really can. In this case his income is the same as in the previous game:

$$Q_{2uuh} = Q_{2u}^* = g - wvb. \quad (33.5)$$

Second, actor 2 can act opportunistically: bluffing that he cannot tell when he really can. In this case, if actor 1 does nothing to deter such opportunism, then actor 2 always rejects the claim and imposes the anchor based punishment of value  $b$  (e.g. keeping the random variable benefit for himself). Then actor 2's income is:

$$Q_{2uuo} = g. \quad (33.6)$$

Suppose that actor 1 endeavors to deter such opportunism by imposing a power based punishment  $P_2$  whenever actor 2 rejects his claim to the random variable benefit. Then

actor 2's income is:

$$Q_{2uuoP} = g - v(P_2 + c_2). \quad (33.7)$$

To determine the size of the punishment that is sufficient to deter opportunism, set  $Q_{2uuoP} = Q_{2uuh}$  and solve for  $P_2$ . The solution is:

$$P_2^* = wb - c_2. \quad (33.8)$$

As long as the punishment size is larger than this threshold value, actor 2 will be deterred from committing opportunism. This result is intuitive. The higher is the probability  $w$ , the more scope for opportunism actor 2 has, and the larger the punishment needs to be to deter opportunism. The larger is the gain from opportunism ( $b$ ), the larger the punishment needs to be to deter opportunism. The lower is actor 2's conflict cost ( $c_2$ ), the larger the punishment needs to be to deter opportunism.

In this equilibrium, actor 2 is completely deterred from committing opportunism. But he is also completely deterred from rejecting the claim and imposing the anchor based punishment, ever. So the actors' incomes are:

$$Q_{1uu}^* = g + vb \quad (33.9)$$

$$Q_{2uu}^* = g - vb. \quad (33.10)$$

This equilibrium is unstable. For actor 2 never rejects the claim and imposes the anchor based punishment – even when he needs to do so to deter actor 1 from moving the anchor intentionally. Hence actor 1 can move the anchor intentionally with impunity. So the actors' incomes devolve to:

$$Q_{1uu}^{**} = g + b \quad (33.11)$$

$$Q_{2uu}^{**} = g - b. \quad (33.12)$$

To deter this kind of opportunism by actor 1, actor 2 adopts the following strategy.

When he can tell that the anchor moved in actor 1's favor due to random forces beyond his control ( $w$ ), he grants actor 1 the random variable benefit  $b$ . When he cannot tell whether the anchor moved randomly or intentionally ( $1 - w$ ), he grants actor 1 the random variable benefit  $b$  with probability  $1 - x$ . With probability  $x$ , he rejects actor 1's claim to  $b$  and imposes the anchor based punishment. When actor 1 retaliates by imposing a power based punishment, actor 2 retaliates in return by imposing a power based punishment  $P_1$ . Then actor 1's income is:

$$Q_{1uuP} = g + wvb + (1 - w)(1 - x)b - (1 - w)(x)(P_1 + c_1). \quad (33.13)$$

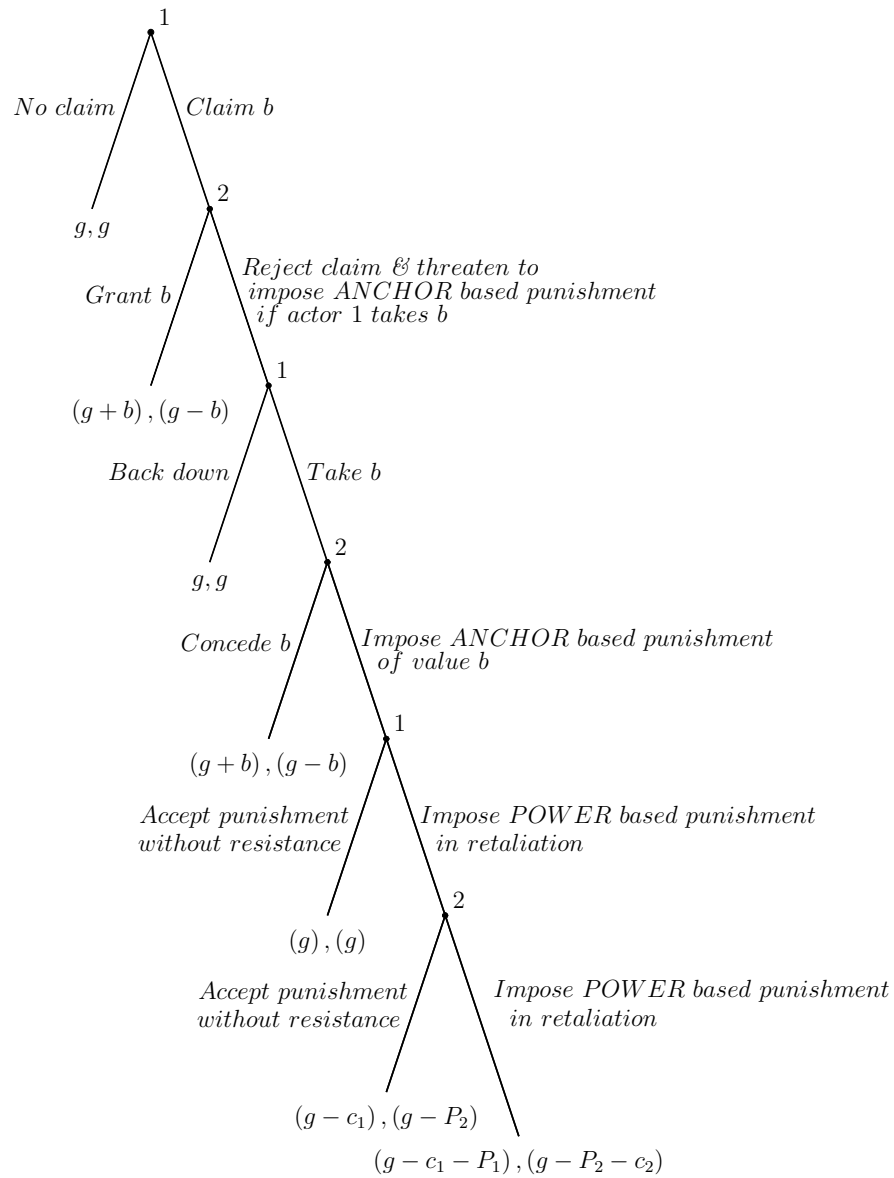
To determine the probability of punishment that is sufficient to deter opportunism, set  $Q_{1uuP} = Q_{1uu}^*$  and solve for  $x$ . The solution is:

$$x^* = \frac{(1 - v)b}{P_1 + c_1 + b}. \quad (33.14)$$

As long as the probability of punishment is higher than this threshold value, then actor 1 will be deterred from committing opportunism. This result is intuitive. The larger is the scope for opportunism ( $1 - v$ ), the higher the probability of punishment needs to be to deter opportunism. The larger is the gain from opportunism ( $b$ ), the higher the probability of punishment needs to be to deter opportunism. The larger the punishment size is ( $P_1$ ), the lower the probability of punishment can be and still deter opportunism. The higher is actor 1's conflict cost ( $c_1$ ), the lower the probability of punishment can be and still deter opportunism.

In this game's equilibrium, actor 1 is deterred from moving the anchor in his favor intentionally. Actor 2 is deterred from bluffing he cannot tell the anchor moved randomly when he really can (in order to keep the random variable benefit for himself when it should go to actor 1 according to the contract terms). However, there is a mutual exchange of power based punishments – a costly conflict – whenever the anchor moves in actor 1's favor due to random forces beyond his control. The cause of the conflict is the breakdown in common knowledge of the meaning of the contract terms – common knowledge of when actor 1 should receive the random variable benefit  $b$ .

#### 4. Figures for this chapter



•Figure 35.1 : The contractual anchor shifting game with two enforcement technologies

Figure 33.1: The contractual anchor shifting game with two enforcement technologies

## CHAPTER 34

### A formal model of the world of law, justice and low cost deterrence

When the world's goods are divided according to relative power, it causes a competition in raising relative power. This outcome can be avoided if actors divide the world's goods by a different method. Suppose that two actors divide a good by means of a contract. They choose a feature of the world that does not shift when relative power shifts and agree that it will constitute the "line in the sand" between one actor's share of the good and the other actor's share. This feature of the world is called an *anchor*. Because the world is always changing, the anchor sometimes moves. The actors need to know how to shift the distribution of the good when the anchor moves. So they write a contract that describes the anchor to make it identifiable, outlines how it moves when the world changes around it, and specifies how the distribution of the good will be allowed to shift cooperatively at these times. The actors can allow the distribution of the good to shift in these ways without conflict as long as the anchor's movements are unintentional on the part of either actor – that is, as long as they are due to natural forces beyond either actor's control. If either actor suspects that the other one moved the anchor intentionally for his own gain, however, then the suspicious actor must fight over the matter for long enough that the suspected actor loses more from the conflict than he expects to gain by moving the anchor intentionally. Then he will be deterred from moving it again.<sup>12</sup>

If there is no adjudication system to distinguish whether a movement of the anchor was due to natural forces or the intentional actions of one actor, then the only way to deter

---

<sup>1</sup>I use the term "fight" in the broadest sense to mean any process in which one actor imposes costs of some kind on the other against his will, and the other resists by imposing costs back. See discussion below.

<sup>2</sup>The ideas in this paragraph are those of Rowe (1989, 1990). The ideas in the next paragraph are my own contribution.

such opportunistic actions is to engage in conflict over *every* shift in the anchor – even the shifts due to natural forces. If there is an adjudication system to make this distinction, however, then the actors can strike a justice bargain that eliminates such opportunism without the need for costly conflicts or competitions in raising relative power. The terms of the justice bargain are straightforward. The first actor agrees that whenever he is found guilty of moving the anchor intentionally for his own gain, he will concede defeat voluntarily without resistance. In return, the second actor agrees that whenever an adverse shift in the anchor (from his point of view) is found to have arisen unintentionally, he will concede defeat voluntarily without resistance. If either actor reneges on his half of the bargain, then the other actor can retaliate simply by not fulfilling his half of the bargain. This kind of retaliation will deter opportunism without the need to engage in costly conflict. In this way, the contractual method of dividing a good between two actors avoids a wasteful competition in raising relative power.

## 1. The contractual anchor shifting game

Suppose that two actors divide a good according to the contractual method (see Figure 34.1). The total amount of the good to be divided is  $2g$ . Assume that the actors divide it equally, so each actor receives  $g$ . Let this be called the status quo distribution.<sup>3</sup> With probability  $v$ , the anchor shifts in actor 1's favor by an amount  $b$  due to natural forces beyond his control (e.g. unintentionally on his part). The contract specifies that when this occurs, actor 1 receives  $g + b$  and actor 2 receives  $g - b$ . With probability  $1 - v$ , the anchor does not shift in actor 1's favor due to natural forces. The contract specifies that at these times the status quo distribution is maintained  $(g, g)$ . At these times actor 1 has an opportunity to move the anchor intentionally to gain the amount  $b$  at actor 2's expense. If there is no adjudication system, then actor 2 is left to his own devices in distinguishing whether a movement of the anchor was unintentional or intentional on actor 1's part. With probability  $w$ , actor 2 cannot tell whether it was unintentional or intentional on actor 1's part. With probability  $1 - w$

---

<sup>3</sup>In a more general version of the model, they divide the good so that actor 1 receives a fraction  $p_o$  and actor 2 receives  $1 - p_o$  (where  $0 < p_o < 1$ ). Since the parameter  $p_o$  would be carried through the entire analysis without affecting the results, I assume that  $p_o = 0.5$  for notational simplicity.



actor 2 can tell.

Suppose the anchor moves in actor 1's favor and he claims  $b$ , but actor 2 does not know whether the movement of the anchor was unintentional or intentional on actor 1's part. If actor 2 assents to the claim, this would betray a willingness to allow actor 1 to gain from movements of the anchor that he might have created intentionally for all actor 2 knows. This, in turn, would give actor 1 the expectation that if he moves the anchor intentionally, then his gains will be permitted. If actor 2 permits such opportunism, then actor 1 will engage in it whenever the anchor does not shift in his favor due to natural forces  $(1 - v)$ . And actor 2 will lose  $b$  at these times.

If actor 2 wants to avoid these losses (e.g. if he wants to maximize his income), then he must insure that actor 1 does not gain anything when the anchor moves in his favor. Specifically, actor 2 must initiate a conflict over actor 1's claim to receive  $b$  and prolong the conflict either until actor 1 abandons his claim and accepts the status quo distribution  $(g, g)$  or until actor 1 has lost more in conflict costs than he expected to gain from receiving  $b$  in the end. If actor 2 fights for long enough to fulfill this condition, then actor 1 will be deterred from moving the anchor intentionally in the future. And if he believes *ex ante* that actor 2 is willing to fight for this long, then he will never move the anchor intentionally in the first place. However, if the anchor moves due to natural forces and actor 1 claims  $b$  for some reason (see below), then actor 2 must initiate and prolong a conflict simply to maintain deterrence of opportunism.

The outcome is different when actor 2 can tell whether a movement of the anchor was unintentional or intentional on actor 1's part. When actor 2 knows the shift was intentional, he can fight for long enough to maintain deterrence (as described in the previous paragraph). When he knows it was unintentional, he can concede actor 1's claim and give him the amount  $b$ . For this would not betray any willingness to allow actor 1 to gain from moving the anchor intentionally. Hence it would not give actor 1 an incentive to move it intentionally again. The outcome is that actor 1 never moves the anchor intentionally (as before). But now there is no conflict following the unintentional shifts, because actor 2 always concedes the claim in these cases.

Another problem arises, however, when actor 2 can tell that a movement of the anchor

was unintentional on actor 1's part. Actor 2 might bluff that he cannot tell – and demand that the distribution remain at the status quo – even though this demand is not necessary to maintain deterrence (as explained in the previous paragraph). The underlying problem is that actor 1 does not know whether actor 2 can tell or can't tell whether the shift was intentional on actor 1's part. So actor 1 cannot allow actor 2 to receive anything more than  $g - b$  when the anchor moves due to natural forces. For allowing that would give actor 2 an incentive to bluff that he cannot tell – and demand that the distribution remain at the status quo – when he really can tell. If actor 1 were to allow this, then he would only receive  $g$  even when the contract awards him  $g + b$ .

If he wants to avoid these losses (e.g. if he wants to maximize his income), then he must insure that actor 2 does not gain anything more than  $g - b$  when the anchor moves due to natural forces. Specifically, actor 1 must demand that the distribution of the good shift in his favor by  $b$  and threaten to fight over it if actor 2 does not concede  $b$ . If actor 2 stands firm in demanding that the distribution remain at the status quo  $(g, g)$ , then actor 1 must engage in conflict and prolong the conflict either until actor 2 concedes  $b$  and accepts  $g - b$  or until actor 2 has lost more in conflict costs than he expected to gain from securing his demand that the distribution remain at the status quo. Then he will be deterred from bluffing that he cannot tell when he really can tell in the future. And if he believes *ex ante* that actor 1 is willing to fight for this long, then he will never bluff that he cannot tell in the first place. Rather, when he can tell (that a movement of the anchor was unintentional on actor 1's part), he will concede  $b$  to actor 1. When he really cannot tell, though, he will fight to deter actor 1 from moving the anchor intentionally. And actor 1 must fight back to deter actor 2 from bluffing that he cannot tell when he really can (because, as mentioned, actor 1 does not know whether actor 2 can tell or can't tell).

The outcome is different still when there is an adjudication system that performs both informational functions. Suppose the system enables actor 2 to know whether a shift in the anchor occurred unintentionally on actor 1's part ( $v$ ) or intentionally on his part ( $1 - v$ ). And the system enables actor 1 to know whether actor 2 can tell this ( $1 - w$ ) or not ( $w$ ). Then no fighting is necessary to deter opportunism by either actor. The two actors can strike a justice bargain that is self-enforcing without the need to fight or threaten fighting. Actor

1 agrees that whenever he is proven guilty of moving the anchor intentionally for his own gain, he will abandon his claim to  $b$  voluntarily without fighting. In return, actor 2 agrees that whenever a movement of the anchor is proven to have occurred unintentionally, he will concede actor 1's claim to  $b$  voluntarily without fighting.

As long as the unintended shifts of the anchor occur frequently enough, actor 1 gains more from actor 2's concessions over them (and from not having to fight over them) than he loses from his own concessions over the shifts he created intentionally. Hence actor 1 is better off adhering to his half of the justice bargain than reneging. If actor 1 ever reneges and refuses to concede defeat voluntarily when found guilty of moving the anchor intentionally, then actor 2 can retaliate by refusing to concede defeat voluntarily after the unintended shifts. Actor 1 would lose more from such retaliations than he would gain from continuing to renege. So he would return to compliance with his half of the bargain. In turn, when the unintended shifts occur frequently enough, actor 2 gains more from not having to fight to deter opportunism by actor 1 than he loses from his own concessions after the unintended shifts. So he too is better off adhering to his half of the bargain than reneging. If he ever reneges and refuses to concede defeat voluntarily after an unintended shift, then actor 1 can retaliate by refusing to concede defeat after a shift he created intentionally. Actor 2 would lose more from such retaliations than he would gain from continuing to renege. So he too would return to compliance with his half of the bargain.

In equilibrium, compliance is self-enforcing without the need to fight or threaten fighting. Hence each actor can disarm safe in the knowledge that he can secure compliance from the other without needing to rearm and fight. When there is an adjudication system that performs both functions, therefore, the actors can avoid costly conflicts and competitions in raising relative power. First consider the situation where there is no adjudication system (4.2). Then consider the situation where there is an adjudication system (4.3).

## 2. Without an adjudication system: Conflict is needed to deter opportunism

Suppose that the anchor shifts in actor 1's favor by  $b$  and actor 2 cannot tell whether the shift was unintentional or intentional on actor 1's part. To deter actor 1 from moving the anchor intentionally, actor 2 demands that actor 1 abandon his claim to  $b$  and threatens to fight for time  $T_2$  if actor 1 does not abandon it. If fighting occurs, actor 2 prolongs the conflict for  $T_2$  and then concedes a fraction  $p_s$  of actor 1's claim (where  $0 < p_s < 1$ ). Then actor 1's income from moving the anchor intentionally whenever he has the opportunity to do so  $(1 - v)$  is:

$$Q_{1m} = \sum_{t=0}^{T_2} d^t Q_{1w}^{**} + \sum_{t=T_2}^{\infty} d^t (g + p_s b). \quad (34.1)$$

By contrast, actor 1's income from never moving the anchor intentionally when he has the opportunity to do so  $(1 - v)$  is:

$$Q_{1nm} = \sum_{t=0}^{\infty} d^t g. \quad (34.2)$$

To derive the minimum duration of conflict necessary to maintain deterrence, set these two incomes equal to each other and solve for  $T_2$ .

$$T_2^* = . \quad (34.3)$$

As long as actor 2 is willing to fight for  $T_2^*$ , then actor 1 will be deterred from moving the anchor intentionally. In equilibrium, he does not move it intentionally. But whenever it moves due to natural forces (e.g. unintentionally on his part), he must claim  $g + b$  for the reason explained above (and formalized below). Then actor 2 fights over the claim and the conflict lasts for  $T_2^*$ .

When actor 2 fights in this way, actor 1 does not know whether it is because actor 2 really could not tell whether the shift was unintentional or intentional on actor 1's part ( $w$ ), or could tell  $(1 - w)$  but bluffed that he could not tell simply to avoid having to concede  $b$ .

To deter actor 2 from bluffing that he could not tell when he really could, actor 1 demands that actor 2 allow the distribution to change to  $(g + b, g - b)$  and threatens to fight for time  $T_1$  if actor 2 does not allow it. If fighting occurs, actor 1 prolongs the conflict for  $T_1$  and then settles for a fraction  $p_s$  of his original claim (where  $0 < p_s < 1$ ). Then actor 2's income from bluffing that he could not tell whenever he has the opportunity to do so  $(1 - w)$  is:

$$Q_{2b} = \sum_{t=0}^{T_1} d^t Q_{2w}^{**} + \sum_{t=T_1}^{\infty} d^t (g - p_s b). \quad (34.4)$$

By contrast, actor 2's income from never bluffing when he has the opportunity to do so  $(1 - w)$  is:

$$Q_{2nb} = \sum_{t=0}^{\infty} d^t (g - b). \quad (34.5)$$

To derive the minimum duration of conflict necessary to maintain deterrence, set these two incomes equal to each other and solve for  $T_1$ .

$$T_1^* = . \quad (34.6)$$

As long as actor 1 is willing to fight for  $T_1^*$ , then actor 2 is deterred from bluffing that he could not tell when he really could. In equilibrium, he does not bluff. But whenever he really cannot tell and demands that the distribution of the good remain at the status quo, actor 1 fights over the issue and the conflict lasts for  $T_1^*$ .

How long does the conflict last:  $T_1^*$  or  $T_2^*$ ? There is a unique value of the settlement fraction  $p_s$  such that  $T_1^* = T_2^* = T^*$ :

$$p_s^* = \frac{g((2c_1 - c_2)g + 2c_1 c_2 k^2(2p_0 - 1))}{(c_1^2 + c_2^2)g - 4bc_1 c_2 k^2} \quad (34.7)$$

$$T^* = \frac{\ln \left( 1 - \frac{4bc_1 c_2 k^2}{(c_1 + c_2)g^2} \right)}{\ln d}. \quad (34.8)$$

This is the combination of parameters  $(p_s, T)$  that minimizes each actor's conflict costs – and

thus joint conflict costs – subject to the constraint that opportunism is deterred.<sup>4</sup> This result will be important for showing that the justice bargain equilibrium to be derived below gives each actor a higher income than the conflict equilibrium that was derived in this section. By showing that a conflict that minimizes each actor's conflict costs leaves each actor with a lower income than he would obtain from a justice bargain, it follows logically that any conflict which is more costly would also leave him with a lower income than he would obtain from a justice bargain.

## 2.1 equilibrium 1: conflict over unintended movements of the anchor

As long as each actor fights for  $T^*$ , neither actor behaves opportunistically. Yet conflict still occurs with probability  $vw$  (e.g.  $100vw$  percent of the time). For the anchor moves due to natural causes beyond actor 1's control with frequency  $v$ . And actor 2 cannot tell that these movements were unintentional on actor 1's part with frequency  $w$ . In these cases, actor 2 must demand that the distribution of the good remain at the status quo  $(g, g)$  and fight over it to deter actor 1 from moving the anchor intentionally. In turn, actor 1 does not know whether actor 2 really could not tell that the anchor's movement was unintentional on his part – or could tell but bluffed that he could not tell. So actor 1 must stand firm in his demand for  $g + b$  and fight over it to deter actor 2 from bluffing.

This mutual deterrence equilibrium is sustainable as long as neither actor is too strong relative to the other. Actor 1's marginal conflict cost,  $c_1$ , must be high enough that his income from conflict is less than his income from not engaging in opportunism and remaining at the status quo:

$$Q_{1w}^{**} < g. \quad (34.9)$$

If  $c_1$  is too low, then actor 1's income from conflict is greater than his income from remaining at the status quo, so he is undeterrable. To calculate the value of  $c_1$  below which actor 1 is

---

<sup>4</sup>If any other settlement fraction were used, then one of the actor's would have to fight for longer than  $T^*$  to deter opportunism by the other. And the other would have to fight back for as long as the first actor is fighting simply to maximize his income. For, as shown in the conflict equilibrium derived in section 3.x, if one actor is fighting, then the other can maximize his income only by fighting back at his optimal allocation  $f_i^*$ .

undeterrable, set  $Q_{1w}^{**}$  equal to  $g$  and solve for  $c_1$ :

$$c_{1min,LI} = \frac{c_2}{2 + 2c_2(2p - 1)\frac{k_p^2}{g}}. \quad (34.10)$$

Actor 1 is deterrable as long as  $c_1 > c_{1min,LI}$ . Similarly, actor 2's marginal conflict cost,  $c_2$ , must be high enough that his income from conflict is less than his income from not engaging in opportunism.

$$Q_{2w}^{**} < g - b. \quad (34.11)$$

If  $c_2$  is too low, then actor 2's income from conflict is greater than his income from conceding  $b$  when he really can tell that the anchor moved due to natural forces. So he is undeterrable. To calculate the value of  $c_2$  below which actor 2 is undeterrable, set  $Q_{2w}^{**}$  equal to  $g - b$  and solve for  $c_2$ :

$$c_{2min,LI} = \frac{c_1}{2 - 2c_1(2p - 1)\frac{k_p^2}{g} - 4c_1b\frac{k_p^2}{g^2}}. \quad (34.12)$$

Actor 2 is deterrable as long as  $c_2 > c_{2min,LI}$ .

When deterrence conditions 34.10 and 34.12 are met, the following strategy combination is a Bayesian Perfect Nash Equilibrium.

**Actor 1' strategy.** When the anchor does not move due to natural forces, do not move it intentionally. When the anchor moves due to natural forces, claim  $b$ . If actor 2 initiates conflict over the claim, prolong the conflict either until actor 2 accepts the claim and concedes  $b$  or until time  $T^*$  has passed. Once the conflict has continued for  $T^*$  or longer, agree to settle the conflict by accepting  $p_s^*b$  from actor 2 rather than  $b$  (where  $0 < p_s^* < 1$ ).

**Actor 2's strategy.** If the anchor moves and actor 1 claims  $b$ , concede the claim when able to tell that the movement of the anchor was unintentional on actor 1's part ( $1 - w$ ). When unable to tell ( $w$ ), initiate conflict over the claim and prolong the conflict either until actor 1 abandons the claim and accepts the status quo distribution  $(g, g)$  or until time  $T^*$  has passed. Once the conflict has continued for  $T^*$  or longer, agree to settle the conflict by conceding  $p_s^*b$  to actor 1.

In this equilibrium the actor's incomes are:

$$Q_{1wb}^* = (1 - v)g + v(1 - w)(g + b) + (vw)g \quad (34.13)$$

$$Q_{2wb}^* = (1 - v)g + v(1 - w)(g - b) + (vw)(g - b). \quad (34.14)$$

There are three outcomes within this equilibrium. With probability  $1 - v$ , there is no natural movement of the anchor and actor 1 does not move it intentionally. With probability  $v(1 - w)$ , the anchor moves by  $b$  due to natural forces and actor 2 can tell that it was unintentional on actor 1's part. So he concedes  $b$  to actor 1. With probability  $vw$ , the anchor moves by  $b$  due to natural forces and actor 1 claims  $b$ , but actor 2 cannot tell that the anchor's movement was unintentional on actor 1's part. So there is a conflict that lasts for  $T^*$ . It leaves actor 1 with the same income he would have obtained if he had not made a claim ( $g$ ) and actor 2 with the same income he would have obtained if he had conceded actor 1's claim ( $g - b$ ). For each actor must fight for long enough to deter the other actor from making claims or opposing claims opportunistically. While this equilibrium deters opportunism, the cost of maintaining deterrence in this way is the cost of a conflict that lasts for  $T^*$ .

## 2.2 equilibria 2a and 2b: capitulation by one side or the other

Suppose that actor 1's marginal cost  $c_1$  is lower than the threshold value  $c_{1min}$ . Then actor 1 is undeterrable. He obtains a higher income from moving the anchor intentionally and claiming  $b$  than from remaining at the status quo – even if actor 2 fights forever over his claim. Hence actor 1 moves the anchor whenever he has the opportunity  $(1 - v)$ . In turn, actor 2 is better off conceding the claim than fighting over the issue, so he always concedes the claim. In this equilibrium the actor's incomes are:

$$Q_1^* = g + b \quad (34.15)$$

$$Q_2^* = g - b. \quad (34.16)$$

Now suppose that actor 2's marginal cost  $c_2$  is lower than the threshold value  $c_{2min}$ . Then actor 2 is undeterrable. He obtains a higher income from bluffing that he could not tell a movement of the anchor was unintentional on actor 1's part than from not bluffing –



even if actor 1 fights over it forever. Hence actor 2 bluffs that he could not tell whenever he has the opportunity  $(1 - w)$ . In turn, actor 1 is better off abandoning his original claim in response than fighting over it, so he always abandons his claim. In this equilibrium the actor's incomes are:

$$Q_1^* = g \quad (34.17)$$

$$Q_2^* = g. \quad (34.18)$$

### 3. With an adjudication system: A justice bargain that avoids costly conflict

Suppose there is an adjudication system that performs the two informational functions explained above. It distinguishes which movements of the anchor were intentional and which were unintentional on actor 1's part. And it distinguishes between the cases where actor 2 really *cannot* tell which movements were intentional and which were unintentional on actor 1's part, on one hand, and the cases where actor 2 really *can* tell. Now the two actors can strike a justice bargain. Actor 1 agrees that whenever he is found guilty of moving the anchor intentionally for his own gain, he will concede defeat and accept the status quo distribution without resistance  $(g, g)$ . In return, actor 2 agrees that whenever the anchor moves in actor 1's favor by  $b$  and it is proven to have been unintentional on actor 1's part, then actor 2 will concede a fraction  $p_j$  of the claim without resistance (where  $0 < p_j < 1$ ). In this equilibrium the actor's incomes are:

$$Q_{1j}^* = (1 - v)g + v(g + p_j b) = g + v p_j b \quad (34.19)$$

$$= (1 - v)g + v(1 - w)(g + p_j b) + v w(g + p_j b) \quad (34.20)$$

$$Q_{2j}^* = (1 - v)g + v(g - p_j b) = g - v p_j b \quad (34.21)$$

$$= (1 - v)g + v(1 - w)(g - p_j b) + v w(g - p_j b). \quad (34.22)$$

Equations 34.20 and 34.22 are provided for comparison with equations 34.13 and 34.14 (the incomes from the conflict equilibrium). The comparison shows that when the actors switch to the justice equilibrium, conflict is avoided and each actor saves his conflict costs. The

actors' savings are:

$$Q_{1s} = Q_{1j}^* - Q_{1wb}^* = \frac{d}{1-d}(p_j - (1-w))vb \quad (34.23)$$

$$Q_{2s} = Q_{2j}^* - Q_{2wb}^* = \frac{d}{1-d}(1-p_j)vb. \quad (34.24)$$

Actor 1's savings are greater than zero whenever  $b > 0$ ,  $v > 0$ , and  $w > 1 - p_j$  (condition JS1). Actor 2's savings are greater than zero whenever  $b > 0$ ,  $v > 0$ , and  $p_j < 1$  (condition JS2).

### 3.1 conditions for a justice bargain to be self-enforcing

Suppose that a justice bargain is in place but actor 1 reneges on his half of the bargain. He moves the anchor intentionally whenever he has the opportunity to do so  $(1-v)$ . And when the adjudication system finds him guilty of moving the anchor intentionally, he refuses to concede defeat and abandon his claim to  $b$ . Actor 2 can retaliate by refusing to concede  $b$  in the other cases – when the system finds a movement of the anchor to have been unintentional on actor 1's part  $(v)$ . Then actor 1's income from reneging and suffering retaliation is:

$$Q_{1mr}^* = (1-v)(g+b) + v(g). \quad (34.25)$$

His income from returning to compliance with the justice bargain is  $Q_{1j}^*$  (equation 34.19). He is better off returning to compliance if  $Q_{1j}^* > Q_{1mr}^*$ . This condition can be rewritten:

$$v > \frac{1}{1+p_j} = v_{1j}^*. \quad (34.26)$$

Thus, as long as the unintentional movements of the anchor are frequent enough, actor 1 can be deterred from reneging on the justice bargain without conflict.

When actor 1 reneges, does actor 2 have an incentive to retaliate by not fulfilling his half of the justice bargain? His income from retaliating in this way is:

$$Q_{2r}^* = (1-v)(g-b) + v(g). \quad (34.27)$$

His income from not retaliating is:

$$Q_{2cap}^* = (1 - v)(g - b) + v(g - p_j b) \quad (34.28)$$

He is better off retaliating if  $Q_{2r}^* > Q_{2cap}^*$ . This condition can be rewritten:  $vp_j b > 0$ .<sup>5</sup> Thus actor 2 is better off retaliating than capitulating if:

$$\left. \begin{array}{l} v > 0, \\ p_j > 0, \text{ and} \\ b > 0. \end{array} \right\} \quad (34.29)$$

In sum, the justice bargain is self-enforcing without conflict provided conditions 34.26 and 34.29 are met ( $v > v_{1j}^*$ ,  $p_j > 0$  and  $b > 0$ ). If these conditions are not met, then this kind of retaliation has no effect. Actor 1 will continue to renege on the justice bargain even if actor 2 persists in retaliating in this way (in which case their respective incomes are  $Q_{1mr}^*$  and  $Q_{2r}^*$ ).

Suppose that condition 34.26 is not met, so this kind of retaliation is ineffective. The only other way for actor 2 to retaliate is through conflict. Actor 1's income from reneging and suffering that form of retaliation is  $Q_{1w}^{**}$ . His income from returning to compliance with the justice bargain is still  $Q_{1j}^*$  (equation 34.19). He is better off returning to compliance if  $Q_{1j}^* > Q_{1w}^{**}$ . This condition can be rewritten:

$$Q_{1w}^{**} < g + vp_j b = Q_{1w}^{**\diamond}. \quad (34.30)$$

Thus actor 1 can be deterred from reneging on the justice bargain through conflict provided (1) he is not too strong, (2) the unintentional movements of the anchor are frequent enough, (3) the stakes in each movement of the anchor  $[b]$  are high enough, and (4) his share in the justice bargain  $[p_j]$  is large enough.

If actor 1 is too strong ( $c_1$  is too low), then he cannot be deterred from reneging on the

---

<sup>5</sup>Note that the larger  $v$  and  $b$  are, the greater is actor 2's incentive to retaliate.

justice bargain through conflict. To calculate the value of  $c_1$  below which he is undeterrable, set  $Q_{1w}^{**}$  equal to  $g + vp_jb$  and solve for  $c_1$ :

$$c_{1min,FI} = . \quad (34.31)$$

Actor 1 is deterrable as long as  $c_1 > c_{1min,FI}$ . However, if actor 1 is so strong that even condition 34.30 is not met, then retaliation through conflict has no effect either. Actor 1 will continue to renege on the justice bargain even if actor 2 persists in retaliating in this way (in which case their respective incomes are  $Q_{1w}^{**}$  and  $Q_{2w}^{**}$ ).

Note that  $c_{1min,FI} < c_{1min,LI}$ . That is, the lowest level of  $c_1$  at which actor 1 is still deterrable through conflict when there *is* an effective adjudication system is lower than the lowest level of  $c_1$  at which actor 1 is still deterrable through conflict when there is *not* an effective adjudication system. To put it another way, the upper bound of actor 1's strength beyond which he is undeterrable through conflict *with* an effective adjudication system is higher than the upper bound of actor 1's strength beyond which he is undeterrable through conflict *without* an effective adjudication system. Thus, there is a range of actor 1's strength within which he is too strong to be deterred through conflict without such a system, but not too strong to be deterred through conflict with such a system. The range is:  $c_{1min,FI} < c_1 < c_{1min,LI}$ .

This conclusion has some major implications. When actor 1's strength lies within this range, actor 2 will need help from a powerful third party to deter actor 1 by force in the absence of an effective adjudication system - but will not need help from a powerful third party in the presence of such a system. Thus, an effective adjudication system reduces the need for a powerful third party to aid in enforcing contracts by force. By implication, polities with effective adjudication systems are expected to have less enforcement assets than polities without effective adjudication systems. At the same time, moreover, polities with effective adjudication systems are expected to experience less domestic conflict than polities without effective adjudication systems. For with effective adjudication, contracts can be enforced without conflict whenever  $v > v_{1j}^*$ . They must be enforced through conflict only when  $v < v_{1j}^*$ . And even then, conflicts are necessary only when actor 1 reneges on his half of a justice bargain, not when anchors move due to natural forces. By contrast, without

effective adjudication, contracts must be enforced through conflict (if they are enforceable at all) not only when  $v < v_{1j}^*$  but also when  $v > v_{1j}^*$ . And even if no actor engages in opportunism because he is deterred by the threat of conflict, there will still be conflicts whenever anchors move due to natural forces. In sum, states with effective adjudication systems will have fewer enforcement assets *and* less domestic conflict than states without effective adjudication systems. This conclusion runs counter to the conventional wisdoms that domestic law and order emanate from a strong state – and when a state is observed to have domestic law and order, it must be because that state is strong.

Does actor 2 have an incentive to retaliate through conflict (when actor 1 reneges on his half of the justice bargain)? His income from retaliating in this way is  $Q_{2w}^{**}$ . His income from retaliating the other way when it has no effect is  $Q_{2r}^*$ . He is better off retaliating through conflict than retaliating the other way without effect if  $Q_{2w}^{**} > Q_{2r}^*$ . This condition can be written:

$$b > \frac{g - Q_{2w}^{**}}{1 - v} = b_{2jw}^*. \quad (34.32)$$

Thus actor 2 prefers retaliation through conflict provided he is not too weak [ $Q_{2w}^{**}$  is not too low] and the stakes in each movement of the anchor are high enough [ $b > b_{2jw}^*$ ]. Note that the higher the stakes are, the weaker he can be and still be better off retaliating than capitulating. But if condition 34.41 is not met, then actor 2 is better off retaliating in the other way, even if it is without effect, than retaliating through conflict.

In sum, there are four equilibria in the justice bargain enforcement game. The first equilibrium occurs when conditions 34.26 and 34.29 are met ( $v > v_1^*$ ,  $p_j > 0$  and  $b > 0$ ). Actor 2's threat to retaliate by not fulfilling his half of the bargain is sufficient to deter actor 1 from reneging on the justice bargain. So actor 1 never moves the anchor intentionally. The second equilibrium occurs when condition 34.26 is not met ( $v < v_{1j}^*$ ), but conditions 34.30 and 34.41 are met ( $Q_{1w}^{**} < Q_{1w}^{**\diamond}$  and  $b > b_{2jw}^*$ ). Actor 1 can only be deterred by actor 2's threat to retaliate through conflict. Actor 2 is better off retaliating in this way than retaliating the other way without effect. So he retaliates this way, and actor 1 never moves the anchor intentionally. The third equilibrium occurs when neither condition 34.26 nor condition 34.30 is met, but condition 34.41 is met. Actor 1 cannot be deterred by either

mode of retaliation. But actor 2 is better off retaliating through conflict (without effect) than retaliating by not fulfilling his half of the justice bargain (without effect). So actor 1 moves the anchor intentionally whenever he has the opportunity, and actor 2 retaliates through conflict (without effect). The fourth equilibrium occurs when none of conditions 34.26, 34.30 or 34.41 is met. Actor 1 cannot be deterred by either mode of retaliation. Actor 2 is better off retaliating by not fulfilling his half of the bargain (without effect) than retaliating through conflict (without effect). So actor 1 moves the anchor intentionally whenever he has the opportunity. And actor 2 retaliates by not fulfilling his half of the bargain (without effect).

## 4. Equilibrium refinements

This section develops five refinements of the justice bargain equilibrium concept:

1. ownership *versus* possession
2. summons screens defendant for positive valuation of justice bargain
3. civil sanctioning *versus* criminal sanctioning
4. verdict establishes a screen for intentionality in repeat offenses
5. questions of fact *versus* questions of law.

### 4.1 ownership *versus* possession

the anchor defines the share of the good to which each actor has *ownership* rights. When the anchor moves in actor 1's favor by  $b$ , the distribution of ownership rights changes by definition. However, it is not clear how the movement of the anchor affects the share of the good in each actor's *possession*. Before the anchor moves, actor 2 has possession of  $b$ . Does the movement of the anchor leave possession of  $b$  with actor 2 or transfer possession to actor 1? The results above assume that the movement leaves possession of  $b$  with actor 2. Hence he retains the option either to concede  $b$  to actor 1 or withhold it from him. This option is what enables actor 2 to retaliate if actor 1 reneges on the justice bargain: actor 2 can withhold  $b$  after the unintentional movements of the anchor (when the justice bargain would require him to concede  $b$ ). Logical consistency requires, however, that actor 2 be assumed to retain possession of  $b$  in all cases where the anchor moves – both the cases where actor 1

moved it intentionally and the cases where it moved due to natural forces beyond his control. Yet, if actor 2 retains possession in the cases where actor 1 moves the anchor intentionally and claims  $b$ , then why is this claim a problem for actor 2? The claim itself does not impose any material loss on actor 2 if he still retains *possession* of  $b$  after the claim is made. So why would he need to retaliate against actor 1 simply for claiming  $b$ ? Thus, even if actor 1 reneges on his half of the justice bargain – refusing to abandon his claim to  $b$  when found guilty of moving the anchor intentionally – it is *not necessary* for actor 2 to retaliate.

This objection can be addressed by making the opposite assumption about how a movement of the anchor affects possession of  $b$ . Suppose that the movement itself automatically transfers possession of  $b$  from actor 2 to actor 1. Now actor 2 suffers a material loss immediately when actor 1 moves the anchor intentionally; and retaliation is necessary. Logical consistency again requires, however, that possession of  $b$  be assumed to transfer automatically to actor 1 in all cases where the anchor moves (both the cases where actor 1 moved it intentionally and the cases where it moved unintentionally on his part). Now, actor 2 is *not able* to retaliate by withholding  $b$  following the unintentional movements of the anchor, because he cannot withhold what he no longer possesses.

In sum, the model as specified so far is logically untenable. Retaliation by actor 2 is either unnecessary or infeasible. These problems can be solved by making the following assumption (which is also more realistic than either of the previous assumptions<sup>6</sup>). With probability  $\phi$ , a movement of the anchor automatically transfers possession of  $b$  from actor 2 to actor 1 (regardless of what caused the anchor to move: natural causes or intentional actions by actor 1). With probability  $1 - \phi$ , a movement of the anchor leaves possession of  $b$  with actor 2 (again regardless of what caused the anchor to move). Now, suppose that actor 1 reneges on the justice bargain. He moves the anchor intentionally at the times when he has the opportunity to do so *and* possession of  $b$  will transfer to him automatically  $[\phi(1 - v)]$ . And when he is found guilty of it, he refuses to concede defeat voluntarily. Now actor 2 can retaliate by withholding  $b$  at the times when the anchor moved due to natural causes *and* it left possession of  $b$  with him  $[(1 - \phi)v]$ . This refinement of the model solves the problem:

---

<sup>6</sup>In the real world, the movement of an anchor causes possession to shift automatically in some cases but not in others. See the appendix for examples of both kinds that were formative in the emergence of the common law in medieval England.

now actor 2's retaliation is both necessary and feasible.

The terms of the justice bargain remain the same because it is defined in terms of ownership rights, not possession. When both actors comply with the bargain, therefore, their incomes are the same and their savings from switching to the justice bargain are the same. The conditions for the bargain to be self-enforcing are different, however. Actor 1's income from reneging on the bargain and suffering retaliation is:

$$Q_{1mr,\phi}^* = (1 - v) \left( (1 - \phi)g + \phi(g + b) \right) + v \left( (1 - \phi)g + \phi(g + b) \right). \quad (34.33)$$

His income from returning to compliance with the justice bargain is:

$$Q_{1j,\phi}^* = (1 - v) \left( (1 - \phi)g + \phi(g) \right) + v \left( (1 - \phi)(g + p_j b) + \phi(g + p_j b) \right). \quad (34.34)$$

Actor 1 is better off returning to compliance if  $Q_{1j,\phi}^* > Q_{1mr,\phi}^*$ . This condition can be rewritten:

$$v > \frac{\phi}{p_j} = v_{1j,\phi}^*. \quad (34.35)$$

Thus he can be deterred from reneging on the bargain without conflict as long as the unintentional movements of the anchor are frequent enough and possession does not shift automatically with movements of the anchor too frequently. If condition 34.35 is not met, however, then this kind of retaliation has no effect. Actor 1 will continue to renege on the justice bargain even if actor 2 persists in his retaliation (in which case their respective incomes are  $Q_{1mr,\phi}^*$  and  $Q_{2r,\phi}^*$ ).

Now reconsider actor 2's decision whether to retaliate, and if so, how. His income from not retaliating at all is:

$$\begin{aligned} Q_{2cap,\phi}^* &= (1 - v) \left( (1 - \phi)g + \phi(g - b) \right) + v \left( (1 - \phi)(g - p_j b) + \phi(g - b) \right) \\ &= (1 - \phi)g + \phi(g - b) - v(1 - \phi)p_j b. \end{aligned} \quad (34.36)$$



His income from retaliating by not fulfilling his half of the justice bargain is:

$$\begin{aligned} Q_{2r,\phi}^* &= (1-v) \left( (1-\phi)g + \phi(g-b) \right) + v \left( (1-\phi)g + \phi(g-b) \right) \\ &= (1-\phi)g + \phi(g-b). \end{aligned} \quad (34.37)$$

The difference is  $Q_{2r,\phi}^* - Q_{2cap,\phi}^* = v(1-\phi)p_jb$ . Thus actor 2 is better off retaliating than capitulating whenever:

$$\left. \begin{aligned} v &> 0, \\ p_j &> 0, \\ b &> 0, \text{ and} \\ \phi &< 1. \end{aligned} \right\} \quad (34.38)$$

This is the same as condition 34.29 except for the additional requirement that  $\phi < 1$ .

Suppose that condition 34.35 is not met, so this kind of retaliation is ineffective and the only other way for actor 2 to retaliate is through conflict. As before, actor 1 prefers compliance with the justice bargain to renegeing and suffering such retaliation if  $Q_{1j}^* > Q_{1w}^{**}$ . So condition 34.30 remains the one that must be met for retaliation through conflict to be effective. If it is not met, then actor 1 will continue to renege on the justice bargain even if actor 2 persists in retaliation through conflict (in which case their respective incomes are  $Q_{1w}^{**}$  and  $Q_{2w}^{**}$ ).

In turn, actor 2's income from retaliating through conflict is still  $Q_{2wb}^*$  (equation 34.14). So he prefers retaliating in this way to capitulating if  $Q_{2wb}^* > Q_{2cap,\phi}^*$ . This condition can be rewritten: [re-analyze]

$$b > \frac{(1-p_j)g - Q_{2w}^{**}}{\phi} = b_{2jw,\phi}^*. \quad (34.39)$$

As before, actor 2 prefers retaliation through conflict provided he is not too weak [ $Q_{2w}^{**}$  is not too low] and the stakes in each movement of the anchor are high enough [ $b > b_{2jw}^*$ ]. And

again, the higher the stakes are, the weaker he can be and still be better off retaliating than capitulating. But now an additional condition must be met for him to prefer retaliation. The frequency with which possession shifts automatically with shifts in the anchor must be high enough. Otherwise his losses from actor 1's opportunism are so infrequent that he prefers to accept the losses than to retaliate. Finally, as before, if condition 34.39 is not met, actor 2 is better off retaliating in the other way, even if it is without effect, than retaliating through conflict.

In sum, the justice bargain enforcement game has the same four equilibria as before. The only differences are that (1) condition 34.26 is replaced by condition 34.35, (2) condition 34.29 is replaced by condition 34.38, and (3) condition 34.41 is replaced by condition 34.39. The model is now more realistic because it recognizes that possession shifts automatically with movements of the anchor in some cases but not in others. Hence the model produces more observable implications that are consistent with the evidence on actual legal, economic and political systems (see later sections).

## 4.2 summons screens defendant for positive valuation of justice bargain

When conditions JS1, 34.26 and 34.29 are met, the justice bargain is self-enforcing without conflict. Whenever actor 1 is found guilty of moving the anchor intentionally, he has an incentive to concede defeat without resistance: he receives a higher income by doing so than he would receive by standing firm and suffering retaliation. By implication, actor 1 has an incentive to heed a summons to attend court to participate in the proceedings that will determine whether a movement of the anchor was intentional on his part. If the court finds it was unintentional then he receives  $b$  (which obviously gives him an incentive to attend court). But even if he is found guilty of moving the anchor intentionally, he is still better off conceding defeat without resistance. For as long as condition JS1 is met, he loses less from such concessions over time than he gains from actor 2's concessions in the other cases when the anchor moved due to natural forces.

The incentive to comply with a court summons and a guilty verdict without resistance is fundamental. It means that the law enforcement authorities do not need to expend any resources to *coerce* actor 1 into court or into compliance with court verdicts *against his*

*will*. This is the practical significance of the switch from the conflict equilibrium to the justice equilibrium. The state can secure compliance with law without using force or even threatening to use force. By avoiding conflict altogether in these cases, the justice bargain avoids the costs of conflicts and investments in power assets needed to engage in conflicts. This welfare gain is only possible for contracts and laws that meet conditions JS1, JS2, 34.26 and 34.29. For other contracts and laws the deterrence of opportunism requires engaging in conflict (where conflict is defined as one side imposing costs on the other against his will and the other side resisting by imposing costs back). And this kind of conflict requires costly investments in raising relative power.

The court summons game is displayed in Figure 2. It is a variant of the anchor shifting game. Now, when the anchor moves for any reason and actor 1 claims  $b$ , actor 2 chooses whether to issue a court summons<sup>7</sup> or concede  $b$ . If he issues a summons, then actor 1 chooses whether to heed the summons or ignore it. Regardless of his choice, the court renders a verdict on whether the anchor's movement was intentional or unintentional on his part. If it is found unintentional, then actor 2 concedes  $b$ . If it is found intentional, then actor 1 chooses whether to comply with the verdict without resistance, in which case he abandons his claim and accepts the status quo distribution of  $(g, g)$ , or to flout the verdict. If he flouts it, then actor 2 chooses whether to capitulate, in which case he concedes  $b$ , or to retaliate. If he retaliates, he chooses whether to retaliate by engaging in conflict or by not fulfilling his half of the justice bargain (e.g. withholding  $b$  after movements of the anchor are found unintentional). Either way, actor 1 then chooses whether to back down and comply with the guilty verdict after all, or to persist in his reneging strategy.

The only difference between the path where actor 1 heeds the summons and the path where he ignores the summons is in the actors' relative power when actor 2 retaliates by engaging in conflict. If actor 1 heeds the summons and comes to court, then he suffers a power disadvantage in the event that he flouts a guilty verdict and is engaged in conflict by

---

<sup>7</sup>This can mean a few things in practice, depending on the type of legal system and enforcement method in operation. It may mean the plaintiff serves notice on the defendant himself. It may mean the plaintiff purchases a *writ* that a court officer then serves on the defendant. The substantive significance is that the defendant is *called* to court by the issuance of a summons rather than coerced into court by force. If he heeds the call, he is coming to court voluntarily, without resistance, rather than being forced into court against his will by the threat or use of force.

actor 2, because his presence in court and physical proximity to actor 2 make him easier to target and engage in conflict.

There are four equilibria in this game. The first equilibrium occurs when the justice bargain is self-enforcing without conflict (under the conditions provided above). Actor 1 heeds the summons every time. He receives  $b$  when the movement of the anchor is found unintentional on his part. When it is found intentional, he concedes defeat without resistance. The second equilibrium occurs when the justice bargain is self-enforcing only through conflict. Actor 1 complies with the justice bargain but ignores the summons every time. By avoiding the loss of relative power that he would suffer from attending court, he narrows the range of relative power over which actor 2 can obtain his compliance with the justice bargain by retaliating through conflict – and widens the range of relative power over which such retaliation is ineffective and he can therefore get away with his opportunism.

The third equilibrium occurs when the justice bargain is not self-enforcing by either method of retaliation on actor 2's part, but he is better off retaliating through conflict (without effect) than the other way (without effect). Actor 1 reneges on the justice bargain and ignores the summons every time. Actor 2 retaliates through conflict without effect. The fourth equilibrium occurs when the justice bargain is not self-enforcing by either method of retaliation on actor 2's part, but he is worse off retaliating through conflict (without effect) than the other way (without effect). Actor 1 reneges on the justice bargain and ignores the summons every time. Actor 2 retaliates by not fulfilling his half of the justice bargain (without effect).

Four points emerge from the court summons game. First, the summons screens actor 1 for his valuation of a justice bargain. He heeds the summons only if he expects to receive a net gain from a justice bargain enforced without conflict (equilibrium 1). He ignores the summons if his income would be lower under such a justice bargain than it would be in any of the other outcomes (equilibria 2, 3 and 4).

Second, when actor 1 complies with a summons and a guilty verdict without resistance, he does so *not* because he fears forceful retaliation by actor 2 or a coercive state. Rather, he does so because he fears losing access to the main benefit he receives from the justice system: the chance to gain  $b$  in situations where the anchor's movement was unintentional

on his part. It is actor 2's threat to deny this benefit that induces actor 1 to comply with the summons and the verdict without resistance. A threat to retaliate through conflict is *not necessary* to induce actor 1 to comply without resistance. Even if actor 2 never makes such a threat – or fails to carry out such threats once made – actor 1 will still comply without resistance.

Third, when actor 1 complies with a summons and a guilty verdict without resistance, it is wrong to infer that he complies because the state is strong in coercive assets. Rather, one should infer that he complies because the state has an effective adjudication system – one that performs the two informational functions studied here. Later I provide statistical evidence that states with higher ratings on the “rule of law” have lower coercive capacity (other things equal). This pattern is exactly what one would expect to observe in a world where compliance with the law results more from an effective adjudication system than from the coercive capacity of the state (at least for some laws). A state may in fact be strong in coercive assets, but this does not necessarily mean that it is *economic* for the state to use such assets to coerce compliance with court summonses and verdicts. It may not be economic. Empirically, states with strong coercive capacity do not necessarily use it to enforce the law – perhaps precisely because it is not economic. And if it is not, then one cannot infer from the *existence* of state coercive capacity that it *explains* why actor 1 complies with court summonses and verdicts in cases where he does.

Fourth, when actor 1 cannot be deterred by actor 2's threat to withhold the main benefit he receives from the justice system, actor 1 may or may not be deterred by the alternate threat of retaliation through conflict. But either way, that alternate threat will not induce him to heed court summonses. On the contrary, he is better off not heeding a summons to avoid incurring the power disadvantage that would result if he attends court (as explained above). To put it another way, there is a range of relative power within which he is deterrable through conflict *if* he heeds the summons and attends court, but *not* if he ignores the summons and stays away from court (thereby retaining his power advantage). Within this range, he can get away with opportunism only by staying away from court. If he attends court, then he would be better off resuming compliance with the justice bargain after suffering retaliation than engaging in opportunism and continuing to suffer retaliation. This

is why actor 1 ignores the summons whenever he is undeterrable by the method of actor 2 withholding the main benefit he receives from the justice system.

### 4.3 civil sanctioning *versus* criminal sanctioning

When actor 2 retaliates by not fulfilling his half of the justice bargain, the practical effect is that actor 1 loses the chance to gain  $b$  when the anchor moves unintentionally on his part. That is, he loses the right to receive favorable variances from the terms of existing contracts and laws when such variances are occasioned by natural forces beyond his control. In actual legal systems, this loss takes the form of being denied access to the justice system as a plaintiff as long as there are previous judgments outstanding against him as a defendant. Such denial of justice may be termed “civil outlawry” – and it is one of the main ways that defendants are penalized for ignoring adverse verdicts in civil cases.<sup>8</sup>

When actor 2 retaliates by engaging in conflict, by contrast, the practical effect is that whenever actor 1 moves the anchor intentionally and ignores a summons to court to answer for it, he is engaged in conflict. And he must fight back simply to maximize his income (or more aptly, to minimize his income loss).<sup>9</sup> In some legal systems, this method of enforcement takes the form of actor 1 being branded “fair game” for actor 2 to impose costs upon, by force, “on sight”. This stigma may be termed “criminal outlawry” – and historically it was one of the main ways that defendants were penalized for ignoring summonses to court in clear-cut cases of wrongdoing.

Actual legal systems have long distinguished between civil outlawry – denial of access to the justice system – and criminal outlawry – being branded fair game for private reprisals or public enforcement actions. This distinction is captured by the summons game as the difference between equilibrium 1 and equilibrium 2. The model thus explains *why it is* that civil outlawry can be effective without the use of force by the plaintiff or the state authorities, while criminal outlawry can be made effective only through the use of force: the feasibility of justice bargains in the first instance, their infeasibility in the second instance.

---

<sup>8</sup>This observation itself suggests that the theory of justice bargains has something to it.

<sup>9</sup>He is also engaged in conflict whenever the anchor moves due to natural forces and he claims  $b$ . See footnote X above.

The model thereby illuminates some key differences between criminal law and civil law. First, the criminal law allows few exceptions or variances due to natural forces beyond the defendant's control. In terms of the model,  $v < v_1^*$  in equilibrium 2, so a justice bargain is not possible even when there *are* circumstances beyond the defendant's control that might be thought to warrant an exception or a variance (that is, even when  $v > 0$ ). The civil law, by contrast, allows many exceptions and variances - to the point that it can appear to be a continuous elaboration over time of new categories of exceptions and variances). Second, the criminal law is enforced primarily by the use of force against the defendant's resistance. Hence criminal enforcement authorities amass the power assets needed to wield force. By contrast, the civil law is often enforceable merely by the threat of civil outlawry (at least for the laws and contracts on which a justice bargain is possible<sup>10</sup>). In these cases, civil enforcement agencies may rely simply on their authority to deny guilty parties access to the justice system to enforce verdicts.

#### 4.4 verdict establishes a screen for intentionality in repeat offenses

Suppose that actor 1 takes an action that has two effects: an intended effect and an unintended effect. Suppose the intended effect has nothing to do with the anchor, but the unintended effect is to move the anchor in his favor by  $b$ . Because actor 2 stands to lose  $b$  from this movement of the anchor, he may initiate a dispute over it.<sup>11</sup> The task of the adjudicator is to infer whether the movement of the anchor was due to the intentional actions of actor 1. The key question is: Did actor 1 *know* that his action would produce that effect? If the adjudicator has direct evidence on what actor 1 knew before taking the action, then intentionality can be inferred directly from the evidence. But if not, then the adjudicator must screen actor 1 for intentionality indirectly.

To establish the screen, the adjudicator issues a settlement in which:

1. the two actors are directed to split the difference in the dispute (so actor 1 gains  $\frac{1}{2}b$  and actor 2 loses  $\frac{1}{2}b$ );

---

<sup>10</sup>For laws and contracts on which a justice bargain is not possible, the civil law is enforced through coercion against the will of the guilty party (albeit sometimes non-physical coercion).

<sup>11</sup>See the appendix for examples of this kind of dispute that were formative in the emergence of the common law in medieval England.

2. a new clause is inserted into the existing contract or law to prohibit the questionable action; and
3. actor 1 is informed of the new clause.

Actor 2 can agree to accept this settlement without fear that he is giving actor 1 an incentive to repeat the questionable action. For once actor 1 has been informed of the new clause that prohibits the action, he can be presumed to *know* that the action has an adverse effect on actor 2. So repetition of it may be presumed to be intentional on his part. If he repeats it and the anchor moves in his favor again, therefore, the adjudicator can find him guilty of moving the anchor intentionally. He will be obligated to concede defeat without resistance rather than reaping a gain from repeating his action. So actor 1 gains no incentive for future opportunism from his gain in the current settlement. In this way, the new clause screens actor 1 for intentionality in the case of repeat offenses.

An adjudication system that performs these three services in the course of settling disputes will produce a growing body of new contract clauses and case law over time. Suppose, however, that the adjudication system only performs the first service: directing the two actors to split the difference in the dispute. This type of system will not produce a growing body of new contract clauses or case law over time. It will not deter actors from moving anchors intentionally either. For when an actor repeats a questionable behavior, the adjudicator will not have any previous settlements with new clauses on which to rely in assessing what the actor *knew* about the effects of his behavior before acting. So the adjudicator will have no basis for concluding that the actor repeated the behavior intentionally, and thus no basis for finding him guilty of doing it intentionally (with the attendant obligation to concede defeat). Hence the actor will gain from repeating the behavior rather than having to pay a penalty for it. So he will have an incentive to repeat it. When the adjudication system only performs the first service, therefore, it does not distinguish between intentional and unintentional movements of the anchors. So justice bargains are impossible. In this environment the only way to deter opportunism is by engaging in conflict.

#### 4.5 uncontracted contingencies: questions of fact *versus* questions of law

The contract specifies that when the anchor moves from  $b$  to  $b'$ , the distribution shifts in actor 1's favor by  $\delta$ . Suppose the anchor moves in a way that is not specified in the



contract, however. Then there is no way to tell from the contract what the distribution of the good should be. How can the actors agree on a distribution of the good without conflict? To analyze this question, suppose that the anchor is composed of two components,  $b_i$  and  $b_j$ . The two components always move together in the actors' experience. So they do not distinguish between the two in their minds or in the contract. The contract simply specifies that if "the anchor" moves "from  $b$  to  $b'$ ", then the distribution of the good shifts in actor 1's favor by  $\delta$ . Yet in fact, when the anchor moves in this way, it is actually shifting from  $(b_i, b_j)$  to  $(b'_i, b'_j)$ .

What happens if the two components start to move independently? Consider an example from medieval England that was formative in the emergence of the Common Law. The contract specifies that when "the tenant" dies, then "his first son" will become "the tenant". the anchor in this contract is "the tenant". Yet in fact, the anchor has two components: (1) the rightful tenant and (2) his expectant heir during his lifetime. The two components usually move together: When "the tenant" dies, the first component shifts to his son and the second component shifts to that son's son. In the terms of the contract, "the anchor" shifts from [(the tenant, his first son)] to ([the tenant's first son],[that son's first son]).

Now suppose the second component shifts but the first does not. That is, the expectant heir dies while the rightful tenant is still alive, and the expectant heir is replaced by *his* expectant heir. In the contract's terms, the anchor shifts from [(the rightful tenant, his first son)] to ([the rightful tenant],[ his first son's first son]). Now it is not clear who should inherit the land after the current tenant dies: his grandson or his *second* son? This problem was only solved "after bitter struggles between grandchildren and their uncles". Such disputes were settled through conflict based on relative power: "Before the [issue] was settled, a vestige of the older feudal world lingered on: whoever succeeded in gaining seisin [possession of the land] kept it."<sup>12</sup>

The opposite problem could also arise: the first component could shift while the second did not. Consider the same contract: When "the tenant" dies, his first son will become "the tenant". This anchor actually has two components: (1) the vassal and (2) his wife. Suppose the vassal dies before the wife does. Who should inherit the land, the wife or the son?

---

<sup>12</sup>Baker 2002: 267.

To analyze these problems, suppose that the anchor has two components that move independently. Each component shifts from  $b_x$  to  $b'_x$  due to natural forces with probability  $v$ . This development produces two new positions of the anchor that are not specified in the contract:  $(b'_i, b_j)$  and  $(b_i, b'_j)$ . How can the actors agree on a distribution of the good without conflict in these contingencies? Suppose they write a new contract clause that defines one of these positions as “a movement of the anchor to  $(b'')$ ” and the other as “no movement of the anchor”. Then the probability that the anchor moves from  $b$  to  $b''$  remains  $v$ . So neither actor gains or loses anything by handling these new contingencies in this way. Their incomes remain as given in equation X.

This method can be generalized. Suppose the anchor has  $n$  components. Each component shifts independently from  $b_x$  to  $b'_x$  due to natural forces with probability  $v$ . This development produces  $2^n - 2$  new positions of the anchor (compared to the initial situation where the anchor has only two components that always move together). If the actors write a new contract clause that defines half of these new positions as “a movement of the anchor to  $(b'')$ ” and the other half as “no movement of the anchor”, then the probability that the anchor moves from  $b$  to  $b'$  remains  $v$ . Again neither actor gains or loses by defining new contingencies in this way. Hence neither has an incentive to oppose defining them in this way. Thus, the world may become very complex – in the sense that each anchor becomes differentiated into more and more attributes that vary independently of each other – and the actors can still divide the good in a stable manner without conflict.

When the problem of uncontracted contingencies arises due to natural forces, it is relatively easy to solve. The problem is this: When new “states of the world” arise all the time and no two of them are perfectly identical, how can actors agree on which states to label as “a movement of the anchor” and which states to label as “no movement of the anchor”? This problem can be solved simply by the convention of alternating between one label and the other with each new contingency that arises (or more generally, assigning one label  $z$  percent of the time and the other label  $100 - z$  percent of the time). This convention is a rule for deciding which of the latest developments is contractually relevant - in the sense that it means the anchor has moved - and which is not. In short, it is a method for reaching consensus on what the contractually relevant facts are. Neither actor has an incentive to

oppose this method of deciding which facts in the world are relevant. Hence it is a positive sum game in that it avoids conflict over uncontracted contingencies.

The situation is different when new contingencies arise from the intentional actions of one actor. Suppose that actor 1 discovers new ways to move the anchor intentionally that are not specified in the contract. What happens if the two actors handle these novel situations in the same way? To analyze this question, suppose that initially actor 1 has only one way to move the anchor intentionally. Call it action  $a$ . The contract defines action  $a$  as intentionally moving the anchor from  $b$  to  $b'$  and prohibits action  $a$ . Now suppose that actor 1 discovers that he can move the anchor in two different ways: action  $a_x$  and action  $a_y$ . By taking action  $a_x$  he moves the anchor from  $b_i$  to  $b'_i$ . By taking action  $a_y$  he moves the anchor from  $b_j$  to  $b'_j$ . Again this development produces two new contingencies that are not specified in the contract: (1) he takes action  $a_x$  without taking action  $a_y$ , thereby shifting the anchor from  $(b_i, b_j)$  to  $(b'_i, b_j)$ ; and (2) he takes action  $a_y$  without taking action  $a_x$ , thereby shifting the anchor from  $(b_i, b_j)$  to  $(b_i, b'_j)$ . Should either or both of these actions be prohibited as “intentionally moving the anchor”? Consider another example from medieval England that was formative in the emergence of the Common Law.

When a tenant died, his lord gained valuable rights over his land. These “feudal incidents” included: (1) the right to receive the agricultural profits from the land from the time the tenant died until the time that the heir assumed possession of it [“primer seisin”]; and (2) the right to control both the land and the heir if the heir was a minor [“wardship”]. The lord stood to receive the agricultural profits for a decade or more if the heir was a minor. These profits included the value of any services owed to the deceased tenant by his sub-tenants.

The primary anchor in this contract is “the tenant’s status: alive or dead”. It appears to have one dimension of variation - on which it shifts due to natural forces. In contractual terms, when the anchor shifts from “alive” to “dead”, the lord receives the feudal incidents. Yet in fact, this anchor has a second component that tenants could shift intentionally and independently to reduce the value of the incidents owed to lords. This component is: “the tenant’s life expectancy is finite”. As long as this component did not change, the lord would eventually receive the incidents. A tenant could shift this component intentionally, however, by donating a part of his land to a church. Then the church could become the tenant of

this parcel. Since a church never died, “the tenant’s life expectancy is *infinite*”. So the lord would never receive the feudal incidents on this parcel.

Another anchor in the contract is: “the agricultural profits from the land from the time the tenant died until the time the heir assumed possession of it”. This anchor also has a second component that tenants could shift intentionally and independently to reduce the value of incidents owed to lords: “the value of any services owed to the deceased tenant by his sub-tenants”. A tenant could sell a part of his land to a sub-tenant in return for a large cash payment and a commitment to perform some trivial service of no value. Then “the value of the services owed to the tenant [seller]” would be little or nothing. Again the lord would never receive the feudal incidents on this parcel.

Suppose the actors try to handle these contingencies by the same method as described above. They write a new contract clause that labels one of the two new actions as “intentionally moving the anchor” and therefore prohibited, but the other as “not intentionally moving the anchor” and therefore permitted. Now, whenever actor 1 has the opportunity to move the anchor intentionally ( $1 - v$ ), he can take one of the actions alone and gain  $(1 - v)\delta$ . If actor 2 agrees to handle such novel contingencies in this manner, then he will lose  $(1 - v)\delta$ .

This conclusion generalizes too. Suppose that actor 1 develops  $n$  ways to move the anchor from  $b$  to  $b'$  and he can take each of these actions independently of the others. This situation produces  $2^n$  new positions of the anchor (depending on which of the  $n$  actions that actor 1 take). Suppose that actor 2 agrees to new contract clauses that define half of these new actions as “intentionally moving the anchor” and “prohibited”, but the other half as “not intentionally moving the anchor” and “not prohibited”. Then actor 1 can take all  $2^n$  actions one after the other. And after a new clause of one type or the other is written in each case, the outcome will be that he gains  $\delta$  at actor 2’s expense half of the time. Thus actor 2 would lose  $\frac{1}{2}(1 - v)\delta$  by agreeing to handle novel contingencies in this way. He therefore has a strong incentive to oppose this method. If actor 2 is to avoid losing anything to actor 1’s opportunism, actor 2 would have to secure new contract clauses that define *all* of actor 1’s new actions as “intentionally moving the anchor” and “prohibited”.

In sum, when uncontracted contingencies arise from the intentional actions of one actor, the problem is not so easy to solve. The problem is: How can one actor detect *every* novel

way that another actor might discover to move the anchor intentionally for his own gain, and prohibit each of these opportunistic actions against the will of the opportunist? This is not simply a question of reaching consensus on what the relevant facts are. It is a question of what the laws should be. The opportunist has an incentive to oppose new laws written for this purpose. The actor harmed by that opportunism has an incentive to secure such laws. Hence it is a zero sum game.

## 4.6 appendix

Suppose that actor 1 takes an action that has the effect of moving the anchor by  $b$ , but this effect was not his intention. For example, suppose that a farmer who is tilling his field removes a large amount of soil and dumps it into the creek that marks the boundary between his property and a neighboring property. As a result the creek shifts course onto the neighbor's land, thereby shifting the boundary. The farmer's sole intention in removing the soil from his field was to prepare the field for planting. The unintended effect though was to shift the boundary to his advantage.

If the existing property agreement does not contain a clause that prohibits such dumping, then a dispute will arise between the two farmers. An adjudicator cannot possibly know the first farmer's real intention in dumping the soil into the creek. But suppose the adjudicator settles the dispute by inserting a new clause into the property agreement that prohibits such dumping – and informs the farmer of the new clause. Then the adjudicator will be able to know the farmer's intention if he dumps soil into the creek *again*. For once the farmer has been informed that the settlement includes the new clause, he may be presumed to know that his soil dumping has an adverse effect on his neighbor. From this point forward, further dumping on his part may be presumed to be intentional. Thus the new clause screens the farmer for intentionality in the case of repeat offenses.

$$\begin{aligned}
 Q_{2w}^{**} - (g - b) &> v(1 - p_j)b \\
 Q_{2w}^{**} &> g - b(1 - v(1 - p_j)).
 \end{aligned}
 \tag{34.40}$$

Does actor 2 have an incentive to retaliate through conflict? His income from retaliating in this way is  $Q_{2w}^{**}$ . He is better off retaliating than capitulating if  $Q_{2w}^{**} > Q_{2cap}^*$ . This condition can be written:<sup>13</sup>

$$b > \frac{g - Q_{2w}^{**}}{1 - v(1 - p_j)} = b_{2j}^*. \quad (34.41)$$

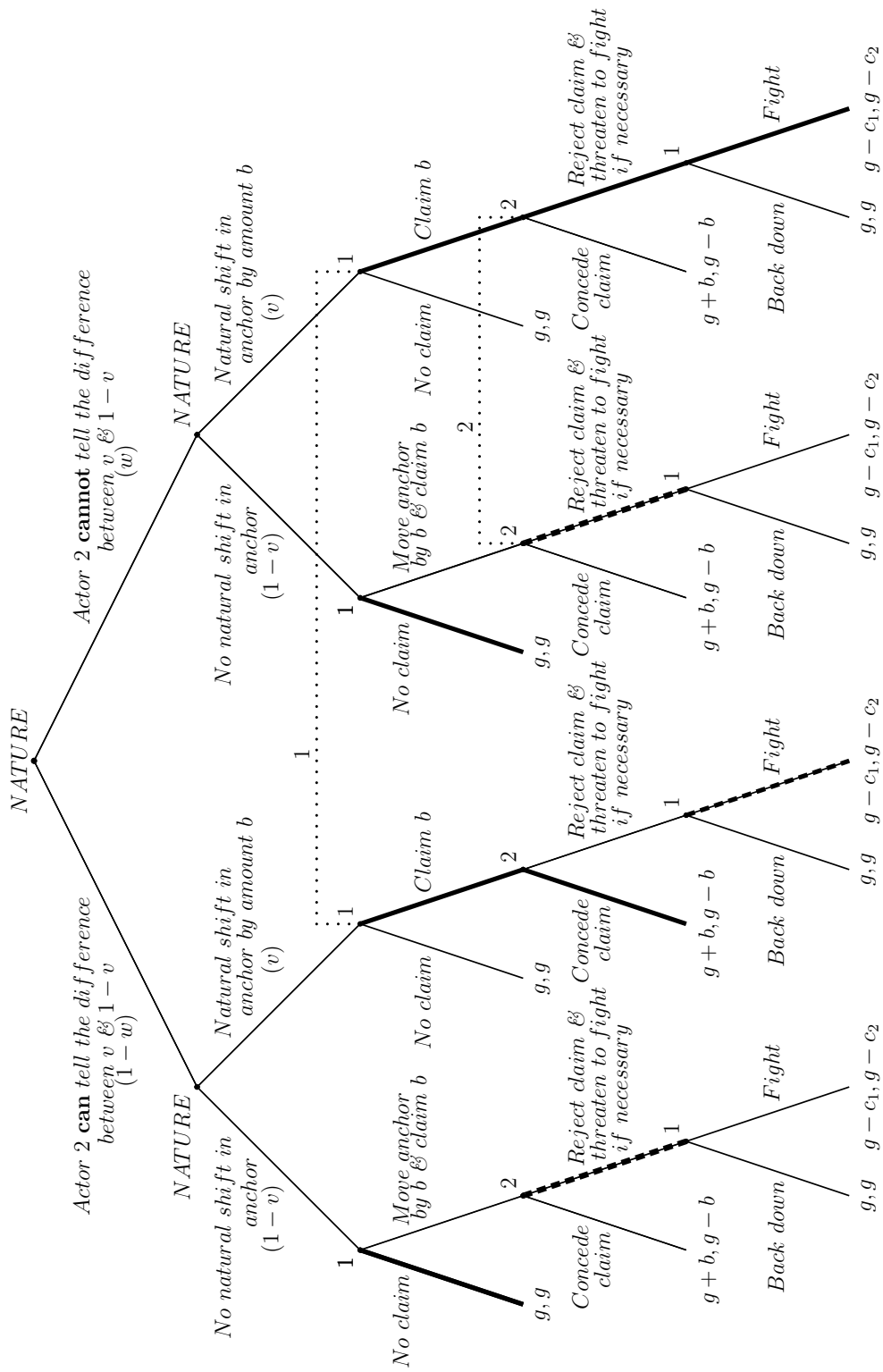
Thus actor 2 prefers retaliation provided (1) he is not too weak and (2) the stakes in each movement of the anchor  $[b]$  are high enough. The higher the stakes are, the weaker he can be and still be better off retaliating than capitulating.

In sum, there are three equilibria in the justice bargain enforcement game. The first equilibrium occurs when conditions 34.26 and 34.29 are met ( $v > v_1^*$ ,  $p_j > 0$  and  $b > 0$ ). Actor 2's threat to retaliate by not fulfilling his half of the bargain is sufficient to deter actor 1 from reneging on the bargain. Actor 2 is better off retaliating in this way than capitulating. So actor 1 never moves the anchor intentionally. The second equilibrium occurs when condition 34.26 is not met ( $v < v_{1j}^*$ ), but conditions 34.30 and 34.41 are met ( $Q_{1w}^{**} < Q_{1w}^{**\diamond}$  and  $b > b_{2j}^*$ ). Actor 1 cannot be deterred by actor 2's threat not to fulfill his half of the justice bargain. But his threat to retaliate through conflict is sufficient. And he is better off retaliating in this way than capitulating. So actor 1 never moves the anchor intentionally. The third equilibrium occurs when neither condition 34.26 nor condition 34.41 is met. Actor 1 cannot be deterred by actor 2 not fulfilling his half of the justice bargain. And actor 2 prefers capitulating than retaliating through conflict. So actor 1 moves the anchor intentionally whenever he has the opportunity, and actor 2 capitulates.

## 5. Figures for this chapter

---

<sup>13</sup>Again, the larger  $b$  is, the more likely that actor 2 is better off retaliating. But now, the larger  $v$  is, the more likely that actor 2 is *not* better off retaliating.



•Equilibrium behavior is shown in bold lines. Off-the-equilibrium-path behavior is shown in dashed lines. Information sets are in dotted lines

•Figure 36.1 : The contractual anchor shifting game

## CHAPTER 35

# A formal model of international security competition and hegemonic war

### 1. definition of the system constraint

The inspiration for the model is Waltz's notion that the international system imposes a constraint on state policy. If the statesman does not heed this constraint, then the state will incur significant costs that impair its ability to defend itself (Waltz 1979). The model specifies the most fundamental constraints on state policy posed by international military competition. Six constraints jointly generate the system constraint. For simplicity I assume that the international system is composed of two states.

**Territory constraint.** The system contains a fixed amount of territory. The more territory one state gets, the less territory is available for the other state.

**Relative power constraint.** If the two states cannot agree to divide the total territory peacefully, then the only other way to divide it is by war. When a war occurs, the outcome depends on the relative power of the two states.

**Income constraint.** Each state must generate enough income from civilian economic production to be able to defend its territory in a war. If national income falls by too much, then there will not be enough tax revenue to feed and pay the soldiers needed to defend the territory. Policy makers then would be at risk of losing office – either by being conquered by a foreign army or by being overthrown from within for failing to defend the country. Policy makers therefore are constrained to avoid policy mistakes that would lose too much national income compared to the maximum that is possible given current technology and labor productivity. Put another way, the system forces policy makers to adopt policies that move the state toward the maximum possible national income, even if they never actually



achieve the maximum. The greater is the income loss from a policy error (compared to the national income-maximizing policy), the stronger is the incentive for policy makers to avoid that error. For the greater is the income loss, the greater will be the risk of not having enough tax revenue to feed and pay the soldiers to defend the land.

**Productivity constraint.** Each state's income is constrained by the productivity of its labor force in two functions: securing territory by military force and producing goods and services in the civilian economy.

**Population constraint.** Each state has a fixed population size in the short run. The most basic decision that leaders make, therefore, is how to divide the population between the military function and the civilian function. Each state faces a fundamental trade-off. The more men it allocates to the military, the less men it will have available to work in the civilian economy. If it allocates too many men to the military, then its economy will be starved of civilian labor as a factor of production and its national income will be lower than it could have been with more civilian labor. On the other hand, the less men it allocates to the military, the less territory it will have as an input to civilian economic production. If it allocates too few men to the military, then its economy will be starved of land as a factor of production and its national income will be lower than it could have been with more land. Leaders thus engage in constrained income-maximization due to the presence of a population constraint and a security constraint. The country's income can only be enjoyed if it is secured from foreign military threats. Therefore enough men must be devoted to the military to secure the land, people and income. The error of devoting too few men to the military results in leaving the country undefended against foreign armies. At the same time, however, enough income must be generated by the civilian economy to insure that there is enough tax revenue to feed and pay the soldiers. The error of devoting too many men to the army results in a loss of current national income needed to supply the soldiers. Either error could cost the state so much as to impair its ability to defend itself.

**Status quo constraint.** If a state's income from peace at the current distribution of territory and labor productivity falls below its expected income from war, then it is better off going to war.<sup>1</sup>

---

<sup>1</sup>Hypothetically, there should always be a way for the two states to avoid war by splitting the difference

## 2. assumptions

To formalize this definition of international security competition for analytical purposes, I assume that each state has a total population of  $N_i$  (where  $i = 1, 2$ ). Each state divides its total population between a military labor pool and a civilian labor pool. Let  $A_i$  be the size of the military labor pool and  $L_i$  be the size of the civilian labor pool for state  $i$ . The population constraint for state  $i$  is then:  $N_i = A_i + L_i$ . The total amount of territory in the system is  $T_t$  (the territory constraint). The total is divided between the two states according to relative army size (the relative power constraint). If the two armies are the same size, the two states split the territory equally. If one state's army is larger, this state gets more than half of the territory and the other state gets less than half. Force superiority garners territory in direct proportion to the superiority. Each state's share of the total territory is thus:

$$p_1 = .5 + \frac{A_1 - A_2}{T_t} \quad (35.1)$$

$$p_2 = .5 - \frac{A_1 - A_2}{T_t} = .5 + \frac{A_2 - A_1}{T_t} = 1 - p_1 \quad (35.2)$$

Each share varies between 0 and 1 by assumption.<sup>2</sup> The two shares sum to 1. Multiplying each state's share  $p_i$  by the total territory gives each state's amount of territory secured by military force.

$$T_i = p_i T_t \quad \text{for } i = 1, 2 \quad (35.3)$$

---

peacefully in their disputes. Even if one state's income from peace has fallen below its expected income from war, there should still be a way for the two states to split the difference peacefully and avoid war. The standard explanation of war, and costly conflict in general, is that it results from a bargaining breakdown due to uncertainty about relative power in an environment of shifting power (Fearon 1995). In other research reported elsewhere (Koppel 1997, 2007, 2011), I provide a model in which bargaining breaks down due to a novel cause that is independent of relative power. Even when there is certainty about relative power and no shifts in power, bargaining can still break down due to this novel cause, resulting in costly conflict. In the present article, I simply assume that bargaining has broken down due to this cause. [add]

<sup>2</sup>The unit of measurement of the territory can always be rescaled to insure that each state's share of the territory varies between 0 and 1. As long as the scaling that is chosen remains constant throughout the analysis, it makes no difference to the substantive conclusions what scaling is chosen.

Each state's national income is a function of its civilian economic production using two inputs: territory secured by military force ( $T_i$ ) and civilian labor ( $L_i$ ). Given the population and security constraints, each state faces the trade-off mentioned above. The more men it devotes to the army to secure territory, the less men it will have left to produce goods and services in the civilian economy. But the less men it devotes to the army to secure territory, the less territory it will have as an input to civilian economic production. This trade-off is expressed by assuming that each state's total national income ( $Q_i$ ) is a function of both its income from territory and its income from civilian labor.

$$\begin{aligned} Q_i &= Q_{Ti} + Q_{Li} \\ &= (aT_i - kT_i^2) + (bL_i - kL_i^2) \quad \text{for } i = 1, 2 \end{aligned} \quad (35.4)$$

The parameter  $a$  represents the productivity of territory as an input to civilian economic production. The parameter  $b$  represents the productivity of civilian labor in producing economic goods and services. There are diminishing returns to both factors (as reflected by the squared terms). The parameter  $k$  represents the rate of diminishing returns to each factor. The parameters  $a$ ,  $b$  and  $k$  constitute the productivity constraint mentioned above.<sup>3</sup>

### 3. optimal army size and equilibrium income in a war

State 1's national income function is graphed in Figure 35.1.<sup>4</sup> The  $x$ -axis represents the division of labor between military and civilian use, where military labor is counted from

---

<sup>3</sup>The function is assumed to be additively separable in the two factors of production: territory ( $T_i$ ) and labor ( $L_i$ ). This assumption is only made to enable the derivation of a linear marginal product schedule for each factor of production. It is not made to suggest that the economy could produce at least some income using only one factor of production. In reality, the economy requires at least some of both factors to produce any income at all. I capture this reality by using only "interior solutions" in the analyses that follow. These are solutions in which the optimal army size is greater than zero and less than the state's total population, and thus the economy receives at least some of both factors as inputs.

This approach assumes there is additive separability between the two factors of production, rather than complementarity. This approach thus relies only on the basic trade-off between one factor of production and the other. The more of one factor that the state uses, the less of the other factor it will have available to use. All of the model's observable implications are derived from this basic trade-off. Any other functional form that also has this basic trade-off will also bear these same implications, whether it has complementarities or not. Therefore, the substantive conclusions of this model would follow equally well from a function with complementarities, as long as it has this basic trade-off.

<sup>4</sup>The figure assumes that the parameters values are:  $a = b = 600$ ;  $k = 2$ ;  $N_1 = N_2 = 100$ ;  $T_k = 100$ .

the left and civilian labor is counted from the right. The  $y$ -axis represents income. The contribution of each factor to national income is shown by a dashed line. The vertical sum of the two factor contributions is state 1's total national income,  $Q_1$ , shown by a solid line. The horizontal tangent line indicates the maximum possible income level given current technology and labor productivity. The division of labor between military and civilian use at this point,  $A_1^*$ , is the division that maximizes national income.<sup>5</sup>

Because there are diminishing returns to each factor, the marginal productivity of labor declines in each use. That is, the more men a state allocates to each use, the lower is the increment to national income from adding another man to that use. This is evident in Figure 35.1: the slope of each dashed line diminishes as more men are added to that use. Figure 35.2 graphs the diminishing slope (value) of each factor's contribution as a function of the number of men in that use. The  $x$ -axis still represents the division of labor between military and civilian use. The  $y$ -axis now represents the marginal *increment* to national income from adding another man to that use (e.g. the marginal product). The marginal product of military labor declines as more men are added to the military (moving from left to right). The marginal product of civilian labor declines as more men are added to the civilian economy (moving from right to left).

The optimal allocation of men between the two uses is the point where the marginal products are equal in both uses. This occurs at the intersection of the two marginal product lines. Again the optimal division of labor is  $A_1^*$ . At this point, which I call the optimal army size, the state is maximizing its national income. If the state were at any other allocation point, then it could increase national income by reallocating men to this point, because the marginal gain from moving toward this point exceeds the marginal loss.

**This is the point toward which policy makers are forced by the international system. Any shift in one or both marginal product lines that changes the state's *optimal* army size will force policymakers to adjust the state's *actual* army size – and any policy that depends on it.**

The optimal army size can be obtained mathematically by deriving each state's best

---

<sup>5</sup>This conclusion assumes that state 2 is also at its income-maximizing division of labor,  $A_2^*$ . A full analysis using best response functions is provided below.

response function and solving for the intersection of the two functions. Consider state 1. Its best response function is obtained by deriving the marginal products of military and civilian labor, setting them equal to each other and solving for the army size,  $A_1$ . The marginal products of military and civilian labor are:

$$\begin{aligned} mp_{A_1} &= \frac{\partial Q_1}{\partial A_1} = a - 2kT_1 \\ &= a - 2kp_1T_k \\ &= a - 2k(.5T_k + A_1 - A_2) \end{aligned} \quad (35.5)$$

$$mp_{L_1} = \frac{\partial Q_1}{\partial L_1} = b - 2kL_1 \quad (35.6)$$

The last equation represents the civilian labor allocation,  $L_1$ , counting from right to left. Using the population constraint, the equation can be rewritten in terms of the military labor allocation,  $A_1$ , counting from left to right:

$$mp_{L_1} = \frac{\partial Q_1}{\partial L_1} = -b + 2k(N_1 - A_1) \quad (35.7)$$

Setting equation 35.5 equal to equation 35.7 and solving for  $A_1$ , the result is state 1's best response function:

$$A_1 = \frac{a - b}{4k} + \frac{N_1 + A_2}{2} - \frac{T_k}{4} \quad (35.8)$$

Because both states have the same population constraint, income function and power function ( $p_i$ ), State 2's best response function is symmetrical:

$$A_2 = \frac{a - b}{4k} + \frac{N_2 + A_1}{2} - \frac{T_k}{4} \quad (35.9)$$

Substituting  $A_2$  into state 1's best response function and solving for  $A_1$ , we obtain state 1's optimal army size in competitive equilibrium:

$$A_{1w}^* = \frac{a - b}{2k} + \frac{2N_1 + N_2}{3} - \frac{T_k}{2} \quad (35.10)$$

Using the population constraint, state 1's civilian workforce in equilibrium is:

$$L_{1w}^* = N_1 - A_{1w}^* \quad (35.11)$$

State 2's optimal army size and civilian workforce are symmetrical:

$$A_{2w}^* = \frac{a - b}{2k} + \frac{2N_2 + N_1}{3} - \frac{T_k}{2} \quad (35.12)$$

$$L_{2w}^* = N_2 - A_{2w}^* \quad (35.13)$$

Notice in equation 35.5 that the marginal productivity of military labor ( $mpA_i$ ) is an increasing function of the productivity of territory as an input to civilian production ( $a$ ). The higher  $a$  is, the higher is the marginal product of military labor, and hence the higher is the optimal army size ( $A_{1w}^*$ ). In the following analysis, therefore, I refer to  $a$  as the productivity of military labor, or simply *military productivity*, and the parameter  $b$  as the productivity of civilian labor, or simply *civilian productivity*.

Substituting  $A_{1w}^*$ ,  $A_{2w}^*$  and  $L_{1w}^*$  into equations 35.1, 35.3 and 35.4, the result is state 1's equilibrium income in a war:

$$Q_{1w}^*(A_{1w}^*, A_{2w}^*) = \frac{b^2 - a^2}{4k} + \frac{2(a - kT_t)(N_1 - N_2)}{3} - \frac{2k(N_1 - N_2)^2}{9} + T_t \left( a - \frac{kT_t}{2} \right) \quad (35.14)$$

State 2's war income is symmetrical. A comparative statics analysis of these results is contained in the appendix. These results assume that the war is continuous. There is never any peace and never any peace dividend.

#### 4. armed peace equilibrium

The two states can achieve peace by agreeing to stop fighting and respect the division of territory that was produced by the war (or would be produced by a war). Once these agreements are reached, each state can transfer some of its men from the military into the civilian economy and thereby reap a peace dividend: more civilian labor produces a higher

national income. The transferred men are the *registered reserves* – men who work in the civilian economy but must be mobilized into the army for the state to reach its optimal army size. The higher is a state's productivity of civilian labor,  $b$ , the larger is the peace dividend and the greater is the incentive for policy makers to reach peace agreements. Since national security requires income to feed and pay soldiers, peace can be beneficial for national security – if only to maximize the size of the treasury available for the next war.<sup>6</sup>

How many men can each state transfer into the civilian economy in peacetime and yet still maintain deterrence of the other state? To put it the other way around, how many men must each state leave in the army at all times - in the barracks and ready to fight at a moment's notice – to maintain deterrence? To answer this question, three additional features of the international system must be considered. First, each state usually has the capacity to mobilize at least some of its registered reserves from the economy into the military secretly – before the other state can notice it and mobilize its own reserves in response. A secret mobilizer then can launch a sudden attack using its temporary force advantage and gain more territory than it would gain otherwise (e.g. without a secret mobilization). Second, the sudden attacker usually enjoys a defense advantage *after* reaping its excess territorial gains: it can keep at least some its gains through defensive actions and emplacements even after the opponent mobilizes fully to its optimal army size. Third, military men are trained primarily to do military jobs. Hence when the registered reserves are released into the civilian economy, they are not as productive as the full-time civilian workers.<sup>7</sup> When all three considerations are present, each state must maintain at least some of its men in the army at all times to maintain deterrence of sudden attacks. The amount that each must maintain depends on the degree to which secret mobilization is possible, the degree of the defense advantage and the degree of specificity of labor skills to military work.

---

<sup>6</sup>Peace is even more essential if there are more than two states in the system, and the other states are reaching peace agreements with each other and growing their economies as a result. Such dynamics are beyond the scope of the present analysis which assumes there are only two states in the system. In other research, I extend the framework to a system with more than two states (Koppel 2009, 2010, 2012).

<sup>7</sup>More precisely, labor skills are a finite resource because they are a form of intellectual capital. A worker can only absorb so much training and no more. Hence the more military labor skills he embodies, the less civilian labor skills he can embody (other things being equal). There may be some complementarities between the two types of skills (e.g. economies of scope). But even then, there is still an underlying trade-off that limits the total amount of skills a single worker can embody.

To demonstrate this result, let us assume that each state divides its army into two parts in peacetime: the part that remains in the barracks ready to fight at a moment's notice and the part that is transferred into the civilian economy. Let the parameter  $t$  represent the fraction that is transferred into the civilian economy, where  $0 \leq t \leq 1$ . Call this the *transfer fraction*. Then  $1 - t$  is the fraction that remains in the army at all times: the *deterrence fraction*. The number of men released into the economy is then  $tA_{iw}^*$ . These are the *registered reserves*. The number of men remaining in the army at all times is  $(1 - t)A_{iw}^*$ . This is the *deterrence force level*.

Suppose that state 1 mobilizes its registered reserves secretly and launches a sudden attack based on its temporary force advantage. Suppose that it retains a fraction  $f_1$  of its excess territorial gains even after state 2 mobilizes fully to its optimal army size. This fraction is a function of two variables: state 2's transfer fraction,  $t_2$ , and a technology parameter,  $0 \leq \theta \leq 1$ , which represents the technological determinants of a state's defense advantage after completing a sudden attack.

$$f_1(t_2, \theta) = t_2\theta \quad (35.15)$$

For a given value of  $\theta$ , the higher is state 2's transfer fraction,  $t_2$ , the higher is the fraction of excess gains that state 1 retains. This assumption reflects a basic fact of international security competition. The more men state 2 releases into the civilian economy during peacetime, the fewer men it has ready to fight immediately in response to a sudden attack by state 1, and the more state 1 can capture assets quickly that increase its capacity to defend its excess gains later, after state 2 mobilizes fully.

Suppose that  $f_1 = 0$ . Then state 1 cannot retain any of its excess gains from its temporary force advantage after state 2 mobilizes fully. Then its income from a sudden attack is simply the equilibrium income from war derived earlier, because state 1 loses all of its excess gains after state 2 mobilizes fully:

$$Q_{1w}^*(A_{1w}^*, A_{2w}^*) \quad (35.14)$$

Now suppose  $f_1 = 1$ . Then state 1 can retain all of its excess gains from its temporary force



advantage after state 2 mobilizes fully. Then its income from a sudden attack is simply the income that it reaps while it has a temporary force advantage, because it retains all of these excess gains:

$$Q_1(A_{1w}^*, (1 - t_2) A_{2w}^*) \quad (35.16)$$

If  $0 < f_1 < 1$ , then state 1's income from a sudden attack is a linear combination of these two incomes:

$$Q_{1w}^{SA} = \left[ f_1 \cdot Q_1(A_{1w}^*, (1 - t_2) A_{2w}^*) \right] + \left[ (1 - f_1) \cdot Q_{1w}^*(A_{1w}^*, A_{2w}^*) \right] \quad (35.17)$$

This income depends on state 2's transfer fraction,  $t_2$ , and the technology parameter,  $\theta$ . The greater is state 2's transfer fraction, the fewer soldiers it has ready in the barracks to fight at a moment's notice, and the higher will be state 1's income from a sudden attack. Similarly, the greater is the technology parameter, the greater is the fraction of excess gains that state 1 retains after state 2 mobilizes fully, and the higher will be state 1's income. Figure 35.3 displays state 1's income from launching a sudden attack as a function of state 2's transfer fraction for different values of the technology parameter.

Is state 1 better off launching a sudden attack than remaining at peace at the current distribution of territory? This depends on its income from remaining at peace. When a state at peace releases its registered reserves into the civilian economy, they are less productive than full-time civilian workers because of the specificity of military training to military work. The rate of productivity loss as men are released into the economy is not constant, however. The first reservists to be released are almost as productive as full-time civilian workers, while the last reservists released are the least productive compared to full-time civilian workers. To capture this fact, let the parameter  $0 \leq \omega \leq 1$  represent the degree to which military labor skills are transferable into the civilian economy (e.g. fungible). As the state releases reservists from the military into the civilian economy, the value of  $\omega$  is assumed to decline from 1 to  $\omega_{min}$ . Thus  $\omega_{min}$  is the fungibility of labor skills of the *last* reservist released into the economy as the transfer fraction rises to its equilibrium level of  $t^*$  (to be defined below). Specifically, as the value of  $t$  increases from 0 to  $t^*$ , each reservist is  $\omega_{min}^t$  times as productive

as a full-time civilian worker. This expression can be called the *productivity deflator*. To illustrate how it varies, suppose that  $\omega_{min} = 1$ . Then military training is completely fungible and all of the reservists released into the economy are just as productive as full-time civilian workers. Conversely, if  $\omega_{min} = 0.1$ , then military training is almost completely non-fungible and the last reservist released into the economy has almost no productivity. The men released before him are decreasingly productive compared to full-time civilian workers. Figure 35.4 shows how the productivity deflator changes in value as men are released from the army into the civilian economy for different values of  $\omega_{min}$ . The  $x$ -axis represents the state's transfer fraction  $t$ . As reservists are released into the economy and  $t$  increases from 0 to 1, the productivity deflator declines in value from 1 to  $\omega_{min}$ .

The productivity difference between reservists and full-time civilians produces, in effect, a gap between the number of men released from the army and the number of men that enter the civilian workforce effectively. The number who leave the army is:

$$t_i A_{iw}^* \quad (35.18)$$

This number increases as the transfer fraction  $t_i$  increases from 0 to 1. Deflating this number by the productivity deflator, the result is the number of men who effectively enter the civilian workforce:

$$\omega_{min}^{t_i} t_i A_{iw}^* \quad (35.19)$$

When  $\omega_{min} = 1$ , the number of men effectively entering the civilian workforce is the same as the number who leave the army:  $t_i A_{iw}^*$ . The lower  $\omega_{min}$  is, the less quickly the number of men entering the workforce increases as men are released from the army. Figure 35.5 shows the number of men leaving the army and the number effectively entering the workforce as the transfer fraction increases from 0 to 1 for different values of  $\omega_{min}$ . The higher the transfer fraction is, the greater is the number of men leaving the army ( $\omega_{min} = 1$ ) and the number of men entering the workforce ( $\omega_{min} < 1$ ). The lower  $\omega_{min}$  is, however, the more slowly the number of men entering the workforce increases compared to the number leaving the army.

After all of the registered reserves arrive in the civilian economy, the number of workers

in the civilian economy is:

$$L_{ip}^* = L_{iw}^* + \omega_{min}^{t_i} t_i A_{iw}^* \quad (35.20)$$

Substituting this number into equation 4, the result is state 1's income from remaining at peace at the current distribution of territory:

$$Q_{1p}^* = Q_1(T_{1w}^*, L_{1p}^*) \quad (35.21)$$

This income depends on state 1's transfer fraction maintained in peacetime,  $t_1$ , and the fungibility of military training to civilian work,  $\omega_{min}$ . The higher  $t_1$  is, the greater is state 1's income from peace at the current distribution of territory. The lower  $\omega_{min}$  is, however, the more slowly this income increases as reservists are released from the army into the civilian economy.

Now we can compare state 1's income from launching a sudden attack to its income from remaining at peace at the current distribution of territory. Figure 35.6 displays both incomes as a function of the peacetime transfer fraction maintained by both states.<sup>8</sup> I assume that both states maintain the same transfer fraction, because both have the same population constraint, income function and power function. It is clear that state 2 must maintain its transfer fraction below  $t = 0.8$  to insure that state 1's income from peace is higher than its income from launching a sudden attack (assuming that state 1 is also maintaining its transfer fraction below 0.8). If state 2 increases its transfer fraction above this level (while state 1 maintains its fraction below this level), then peace is not stable because state 1 is better off launching a sudden attack. I call this critical level the *maximum stable transfer fraction* ( $t^*$ ). This fraction can be calculated for different values of  $\theta$  and  $\omega_{min}$  by setting state 1's income from launching a sudden attack (eq. 35.17) equal to its income from peace at the current distribution of territory (eq. 35.21) and solving for  $t$ . The calculation assumes that both states maintain the same transfer fraction in equilibrium. Figure 35.7 plots the solution for the maximum stable transfer fraction in equilibrium,  $t^*$ , for different levels of  $\theta$

---

<sup>8</sup>The technology parameter,  $\theta$ , is assumed to be 0.8. The skill fungibility parameter,  $\omega_{min}$ , is assumed to be 0.5.

and  $\omega_{min}$ . The  $x$ -axis represents the technology parameter,  $\theta$ . The  $y$ -axis represents  $t^*$ . The solid lines show  $t^*$  for different levels of the skill fungibility parameter,  $\omega_{min}$ . The dashed line is added to show that a maximum stable deterrence fraction of 0.9 is consistent with many different combinations of the technology and skill fungibility parameters. The higher is the sudden attacker's ability to retain its excess gains ( $\theta$ ), the lower the other state must keep its transfer fraction to keep the peace. On the other hand, the higher is the fungibility of military labor skills into the civilian economy ( $\omega$ ), the higher is the sudden attacker's income from remaining at peace, and therefore the higher the other state can move its transfer fraction and still keep the peace.

Empirical observation suggests that states only maintain a small fraction of their optimal fighting armies at full readiness during peacetime. In the analyses that follow, therefore, I assume that both states maintain a transfer fraction of 0.9, which implies a deterrence fraction of 0.1. Thus each state is assumed to maintain 10% of its optimal army size at full readiness in peacetime. I use this value simply because it is a small fraction and a round number. It is also consistent with many different combinations of the technology and skill fungibility parameters. Hence there is no need to know what these parameter values are in fact. Any plausible combination of them is sufficient to obtain the analytical results and observable implications derived below.

## 5. Figures for this chapter

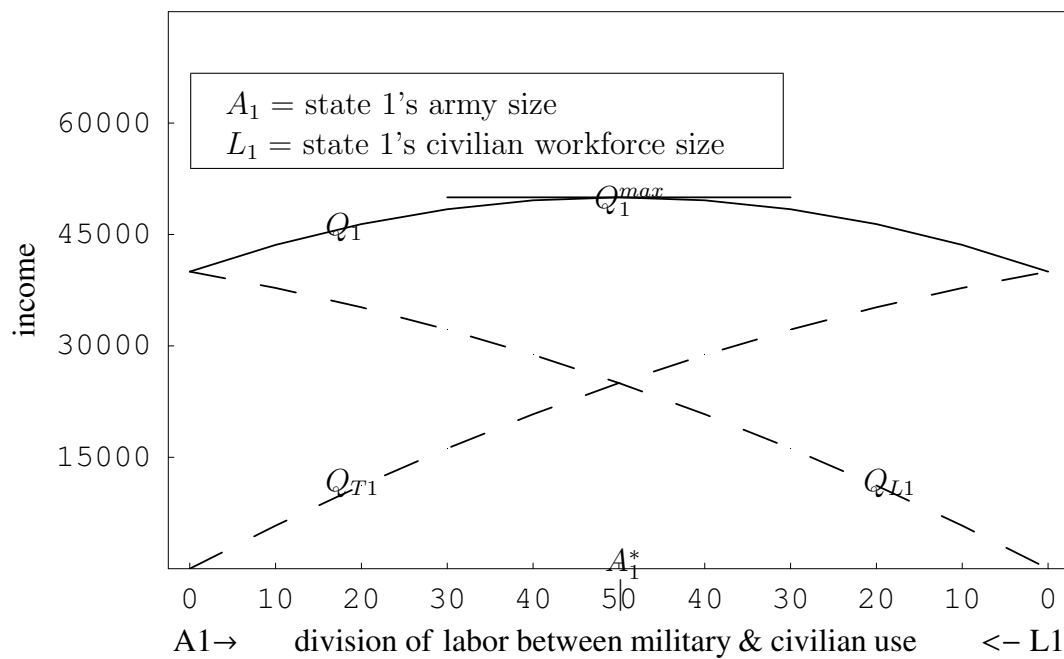


Figure 35.1: State 1's total income ( $Q_1$ ) and contributions of factors  $T_1$  and  $L_1$  to total income

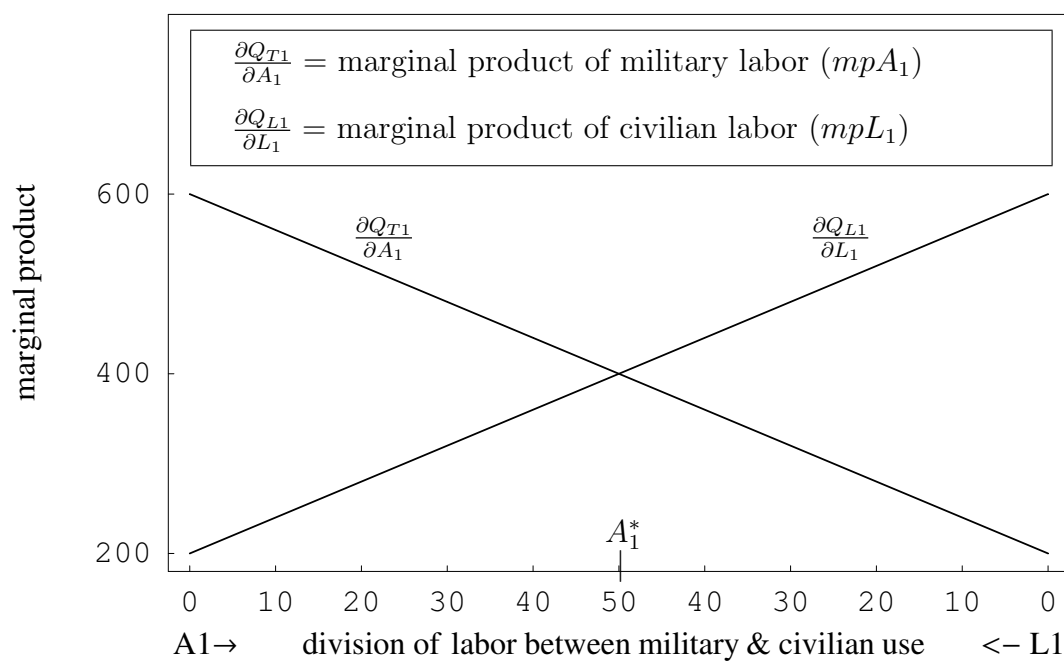


Figure 35.2: State 1's marginal products of military and civilian labor

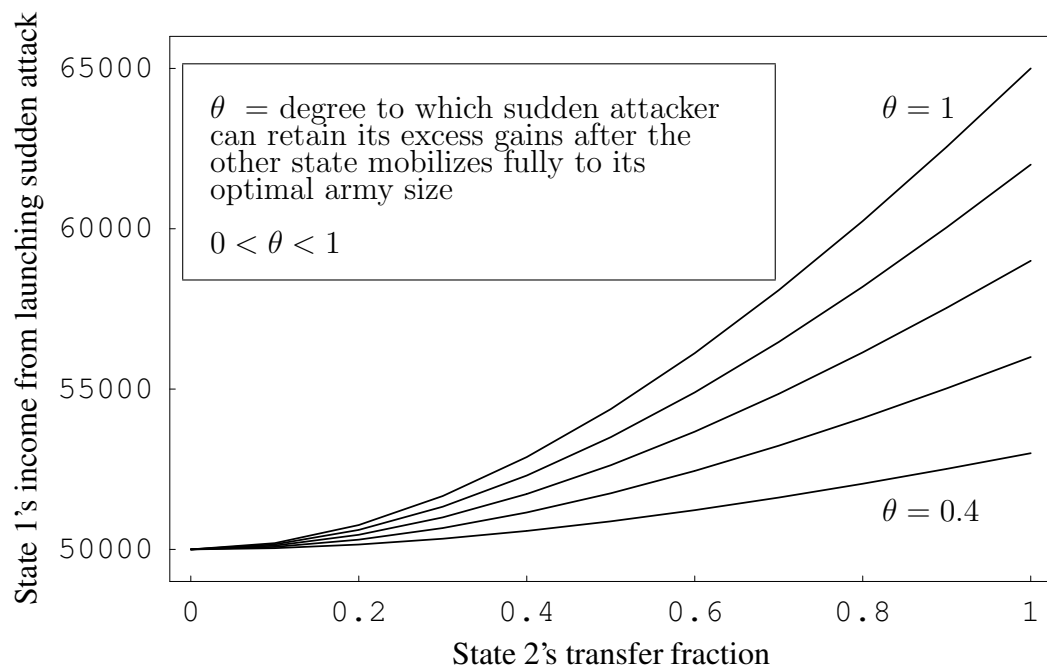


Figure 35.3: State 1's income from launching a sudden attack as a function of State 2's transfer fraction

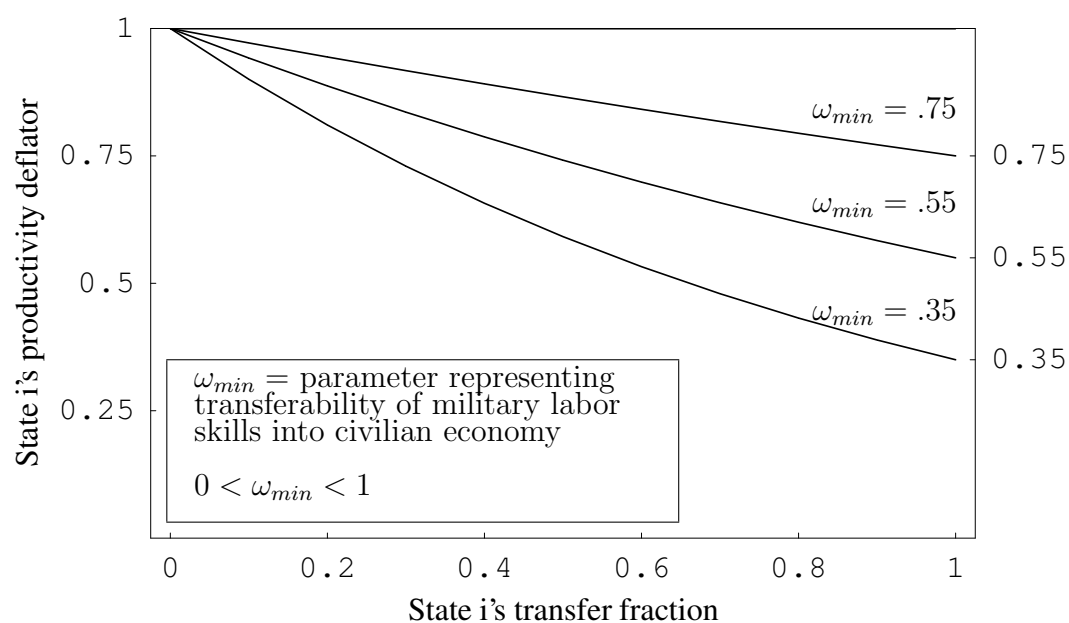


Figure 35.4: State i's productivity deflator

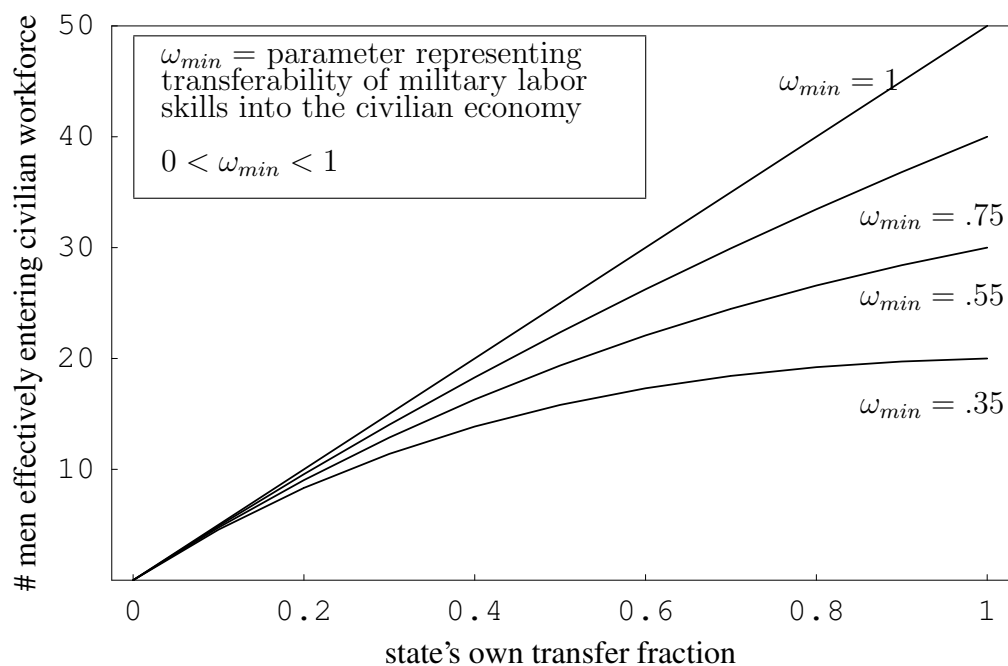


Figure 35.5: number of men effectively entering civilian workforce as a function of the state's own transfer fraction

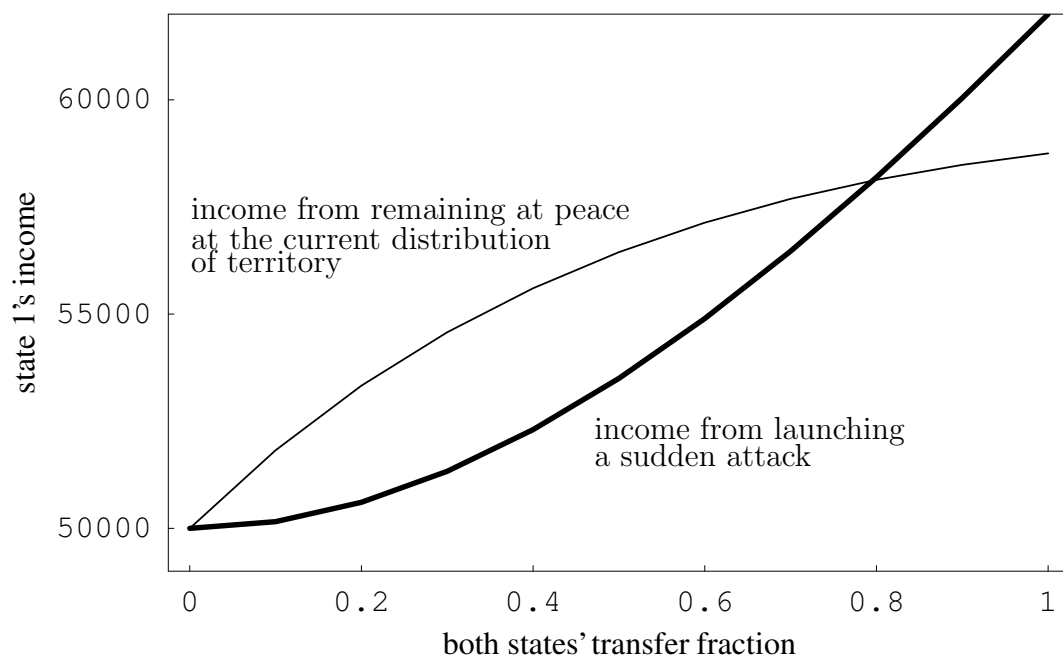


Figure 35.6: State 1's income from peace and income from launching a sudden attack as a function of both states' transfer fraction

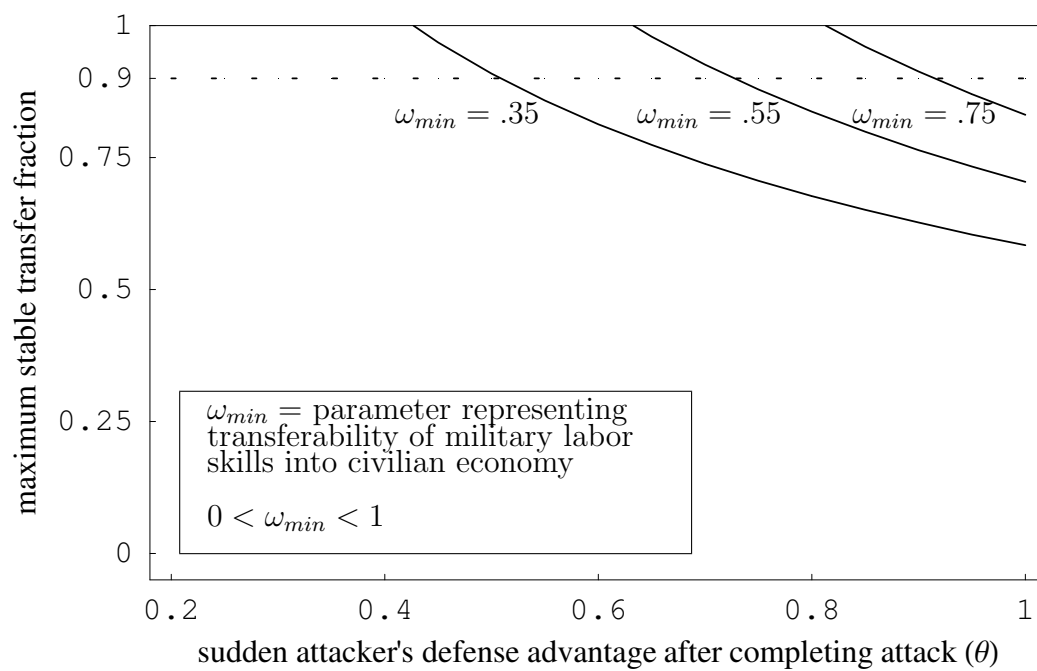


Figure 35.7: Maximum stable transfer fraction as a function of sudden attacker's defense advantage after completing attack



## CHAPTER 36

### A formal model of how enforcement induces a competition in raising relative power to punish and evade punishment

#### 1. basic setup: the mutual gains from exchange game

There are two states, 1 and 2. Each state has a policy variable  $\kappa_i$  that it can increase in exchange for a commitment from the other state to do the same (for  $i \in 1, 2$ ). The status quo levels of the policy variables prior to any exchange are normalized to zero ( $\kappa_i = \kappa_j = 0$ ). When state  $i$  increases its policy to  $\kappa_i > 0$ , it incurs a cost of  $\kappa_i$  and produces a benefit for both state  $i$  and state  $j$  of  $b_i\sqrt{\kappa_i - m}$ . The parameter  $b_i$  is the rate at which the change in  $\kappa_i$  produces benefits for both states. The parameter  $m$  is the minimum increase in  $\kappa_i$  necessary to produce any benefit at all. This parameter is held constant throughout the analysis. Thus, when state  $i$  increases  $\kappa_i$  above zero, each state's benefit is

$$\delta y_i = \delta y_j = \begin{cases} b_i\sqrt{\kappa_i - m} & \text{for } \kappa_i > m \\ 0 & \text{for } \kappa_i \leq m. \end{cases} \quad (36.1)$$

The assumption that  $m > 0$  captures a general feature of most beneficial actions. If the benefit from the action is not large enough, then the action is not worth taking at all. if  $b_i < 2\sqrt{m}$ , then state  $i$ 's total benefit from unilaterally increasing its policy above  $m$  is always less than its total cost of doing so (e.g. no matter how much or little it increases  $\kappa_i$ ). Hence the increase is not worth making. The larger the parameter  $m$  is, the greater the benefit rate  $b_i$  must be for state  $i$  to find it worthwhile to increase its policy unilaterally from the status quo ( $\kappa_i = 0$ ). Also, note that while each state's increase in its policy produces

a benefit for both states, the benefit is *not* a public good. It is rival in consumption and excludable.<sup>1</sup>

If both states increase their policies above  $m$ , then state  $i$ 's income is

$$y_i = b_i\sqrt{\kappa_i - m} + b_j\sqrt{\kappa_j - m} - \kappa_i. \quad (36.2)$$

Thus  $y_i$  measures whether state  $i$ 's *total* benefit from the mutual policy adjustment exceeds its *total* cost. When its total benefit exceeds its total cost,  $y_i > 0$ . When its total benefit is less than its total cost,  $y_i < 0$ .

If state  $i$  unilaterally increases  $\kappa_i$  above  $m$ , while state  $j$  sets  $\kappa_j$  at or below  $m$ , then state  $i$ 's income reduces to

$$y_{iu} = b_i\sqrt{\kappa_i - m} - \kappa_i. \quad (36.3)$$

If  $b_i > 2\sqrt{m}$ , then  $y_{iu} > 0$ . In this case state  $i$  has an interest in increasing  $\kappa_i$  above  $m$  even if state  $j$  does not reciprocate by increasing its policy above  $m$ . Conversely, if  $b_i < 2\sqrt{m}$ , then  $y_{iu} < 0$ . In this case state  $i$  only receives a net benefit from increasing its policy above  $m$  if state  $j$  reciprocates.<sup>2</sup> I assume that  $b_i$  is below this critical threshold for both states to capture the essence of the international cooperation problem. Each state needs the other state to reciprocate its policy change at least to some extent or else it has no incentive to change its policy from the status quo level ( $\kappa_i = 0$ ). In short, neither state can solve the substantive problem at hand alone. Let the critical threshold be denoted  $b_{i,max} = 2\sqrt{m}$ .

Assume that both states increase their policies above  $m$ . Each state faces a choice whether to increase its policy to the level that maximizes its own income or the level that maximizes

---

<sup>1</sup>State  $i$ 's increase in  $\kappa_i$  produces a *total* benefit of  $2b_i\sqrt{\kappa_i - m}$ . This total is split evenly between the two states (by assumption). It could be split unevenly. However it is split, though, the more of the benefit that one state receives, the less of it is available for the other state. So the benefit is rival in consumption. To put this another way, the benefit is exhaustible in supply. If one state reduces its policy by  $\delta\kappa_i$ , to both states' detriment, then the other state can replace its loss only by incurring a replacement cost of  $\delta\kappa_j$ . The benefit is also excludable. If state  $i$  reduces its policy from the socially optimal level to the privately optimal level (as shown below), then state  $j$  can exclude state  $i$  from some measure of benefit in retaliation by reducing its policy  $\kappa_j$  in response. State  $j$  pays a cost of its own to exclude in this way. As long as its gain from excluding in this way exceeds its cost of doing so, however, there is excludability.

<sup>2</sup>See appendix X for the proof, which relies on the optimal value of  $\kappa_i$  derived below.

the sum of both states' incomes. First suppose each state increases its policy to the level that maximizes its own income. It increases  $\kappa_i$  to the level at which its own marginal benefit from the increase is equal to its own marginal cost from the increase. The marginal benefit schedule is:  $\frac{b_i}{2\sqrt{\kappa_i - m}}$ . The marginal cost schedule is 1. Setting these schedules equal and solving for  $\kappa_i$ , state  $i$ 's optimal policy level is

$$\kappa_{ip}^* = \frac{b_i^2}{4} + m. \quad (36.4)$$

The subscript  $p$  denotes *privately* optimal, as distinct from *socially* optimal. The greater is state  $i$ 's policy productivity parameter  $b_i$ , the more state  $i$  increases its policy. The greater is the minimum increase  $m$  necessary to produce any benefit at all, the more state  $i$  increases its policy. When each state increases its policy to the level that maximizes its own income, state  $i$ 's income is

$$y_{ip}^* = \frac{b_i^2}{4} + \frac{b_j^2}{2} - m. \quad (36.5)$$

Each state's income rises with both the productivity of its own policy and the productivity of the other state's policy. Its income rises faster with the productivity of the other state's policy than with the productivity of its own, however. So each state has a greater interest in improving the other state's policy productivity than in improving its own.

There is a minimum threshold of  $b_i$  below which state  $i$  gains nothing from increasing its policy above  $m$  (e.g.  $y_i < 0$ ) even if state  $j$  reciprocates. This threshold is

$$b_{i,min} = \sqrt{4m - 2b_j^2}. \quad (36.6)$$

When both states' policies are equally effective ( $b_i = b_j$ ), this minimum threshold reduces to

$$b_{min} = 2\sqrt{\frac{m}{3}}. \quad (36.7)$$

Unless otherwise noted, I assume that  $b_{i,min} < b_i < b_{i,max}$  for  $i \in 1, 2$ . Figure X displays state 1's income  $y_{1p}^*$  for different values of  $b_1$  and  $b_2$ .

When each state increases its policy to the level that maximizes its own income, the sum

of the two states' incomes is

$$y_p^* = y_{ip}^* + y_{jp}^* \quad (36.8)$$

$$= \frac{3(b_i^2 + b_j^2)}{4} - 2m. \quad (36.9)$$

When both policies are equally effective, this sum is greater than zero only if  $b > b_{min}$ .

Now suppose that each state increases its policy to the level that maximizes the sum of the two states' incomes. The sum of their incomes is

$$y_s = 2b_i\sqrt{\kappa_i - m} + 2b_j\sqrt{\kappa_j - m} - \kappa_i - \kappa_j. \quad (36.10)$$

Now each state increases its policy  $\kappa_i$  to the point that the *social* marginal benefit from the increase is equal to the *social* marginal cost of the increase. The social marginal benefit schedule is:  $\frac{b_i}{\sqrt{\kappa_i - m}}$ . The social marginal cost schedule is 1. Setting these schedules equal and solving for  $\kappa_i$ , the socially optimal level of state  $i$ 's policy is

$$\kappa_{is}^* = b_i^2 + m. \quad (36.11)$$

This level is higher than the privately optimal level. If each state adopts the socially optimal level, then *the two states maximize the mutual gains that are available from the exchange of policy adjustments*. Then state  $i$ 's income is

$$y_{is}^* = b_j^2 - m. \quad (36.12)$$

Now each state's income rises only with the productivity of the *other* state's policy, not with the productivity of its own policy. So each state has an interest in improving the other state's policy productivity but not its own (!).

Now the sum of the two states' incomes is

$$y_s^* = y_{is}^* + y_{js}^* \quad (36.13)$$

$$= b_i^2 + b_j^2 - 2m. \quad (36.14)$$

This sum is higher than the sum when both states adopt the privately optimal policy levels. The total gain from the mutual shift to the socially optimal levels is

$$y_{s-p}^* = y_s^* - y_p^* \quad (36.15)$$

$$= \frac{b_i^2 + b_j^2}{4}. \quad (36.16)$$

This gain is greater than zero whenever  $b_i > 0$  for  $i \in 1, 2$ .

When both states switch from the privately optimal policy levels to the socially optimal levels, each state's income gain is:

$$y_{is}^* - y_{ip}^* = \frac{b_j^2}{2} - \frac{b_i^2}{4}. \quad (36.17)$$

This gain is greater than zero as long as  $b_i < \sqrt{2}b_j$ . When this condition holds for both states, each state has an incentive to reach an agreement with the other to raise its policy from the privately optimal level to the socially optimal level.

## 2. cheating and detection

Suppose the two states reach an agreement to adopt the socially optimal levels. Then their incomes are

$$y_{1s}^* = y_1(\kappa_{1s}^*, \kappa_{2s}^*) = b_2^2 - m$$

$$y_{2s}^* = y_2(\kappa_{1s}^*, \kappa_{2s}^*) = b_1^2 - m.$$

Suppose state 1 cheats on the agreement by reducing  $\kappa_1$  to the privately optimal level, while state 2 maintains  $\kappa_2$  at the socially optimal level. Then their incomes are

$$y_{1x1}^* = y_1(\kappa_1^*, \kappa_{2s}^*) = \frac{b_1^2}{4} + b_2^2 - m \quad (36.18)$$

$$y_{2x1}^* = y_2(\kappa_1^*, \kappa_{2s}^*) = \frac{b_1^2}{2} - m. \quad (36.19)$$

By cheating state 1 gains  $\frac{b_1^2}{4}$  and state 2 loses  $\frac{b_1^2}{2}$ .

If state 1's cheating is not detected, then state 2 is not able to punish.<sup>3</sup> So state 1 can continue to cheat and reap this gain. Hence state 1 has an incentive to invest resources in evading detection. If state 1's cheating is detected, by contrast, then state 2 is able to punish. To do so, state 2 reduces its policy from the socially optimally level to the privately optimal level.<sup>4</sup> Then state 1 is worse off than if it had not cheated at all.<sup>5</sup> If state 1 knows it will be detected and punished, therefore, then it will not cheat in the first place. So state 2 avoids the loss of  $\frac{b_1^2}{2}$  by being able to punish. Hence state 2 has an incentive to invest resources in detecting cheating.

Assume that state 2 invests an amount of resources  $d_2$  in detection at a cost of  $\frac{c_2 d_2^2}{2}$ . In turn, state 1 invests an amount of resources  $e_1$  in evading detection at a cost of  $\frac{c_1 e_1^2}{2}$ . The probability that state 1's cheating is detected,  $d$ , is determined by state 2's investment in detection relative to state 1's investment in evasion:

$$d = \frac{d_2}{e_1 + d_2}. \quad (36.20)$$

The more state 2 invests in detection, the higher the detection probability is. The more state 1 invests in evasion, the lower the detection probability is. If state 1 invests nothing in evasion, then state 2 can raise the detection probability to one by investing even the smallest amount in detection. Conversely, if state 2 invests nothing in detection, then state 1 can reduce the detection probability to zero by investing even the smallest amount in evasion. As long as each state invests something, the range of the detection probability is  $0 < d < 1$ . This range captures the main case that the neoliberal institutionalist paradigm claims to

---

<sup>3</sup>One might object to the assumption that state 1's cheating might not be detected at all, so state 2 cannot punish at all. One might argue that, in the real world, state 2 always notices the cheating eventually. For the present analysis, it is sufficient to assume that a long delay in detection is equivalent to no detection at all. This corresponds to a common problem in the real world: punishing is economic to do if the cheater is detected quickly, but uneconomic to do if not. For example, X.

<sup>4</sup>State 2 is better off punishing in this way than allowing the cheating to continue unpunished, because  $y_{2p}^* > y_{2x1}^*$ . The difference is:  $y_{2p}^* - y_{2x1}^* = \frac{b_2^2}{4}$ .

<sup>5</sup>As shown above,  $y_{1p}^* < y_{1s}^*$ . The difference  $y_{1s}^* - y_{1p}^*$  is shown in equation 36.17.

explain.<sup>6</sup> Now the incomes are

$$y_{1d} = d y_{1s}^* + (1 - d) y_{1x1}^* - \frac{c_1 e_1^2}{2} \quad (36.21)$$

$$y_{2d} = d y_{2s}^* + (1 - d) y_{2x1}^* - \frac{c_2 d_2^2}{2}. \quad (36.22)$$

State 1 invests in evasion to the point that the marginal benefit from the investment is equal to the marginal cost. The marginal benefit schedule is:  $e_1 y_{1x1}$ . The marginal cost schedule is:  $e_1 y_{1s} + c_1 e_1$ . Setting these schedules equal and solving for  $e_1$ , the result is actor 1's reaction curve (e.g. his optimal investment in evasion as a function of state 2's investment in detection):

$$e_1(d_2) = \frac{b_1 \sqrt{d_2} - 2d_2 \sqrt{c_1}}{2\sqrt{c_1}}. \quad (36.23)$$

State 2's reaction curve is obtained in the same way. State 2 is assumed to invest in detection to the point that the marginal benefit equals the marginal cost. The marginal benefit schedule is:  $d_2 y_{2s}$ . The marginal cost schedule is:  $d_2 y_{2x1} + c_2 d_2$ . Setting these schedules equal and solving for  $e_1$ , the result is actor 2's reaction curve (e.g. his optimal investment in detection as a function of state 1's investment in evasion):

$$d_2(e_1) = \frac{b_1 \sqrt{2d_1} - 2d_1 \sqrt{c_2}}{2\sqrt{c_2}}. \quad (36.24)$$

To obtain state 1's equilibrium investment in evasion, substitute equation 36.24 into equation

---

<sup>6</sup>To see this, consider the alternative cases where the detection probability is either zero or one. If the detection probability is zero after the optimal investments in detection and evasion are made, then state 1 expects to be able to cheat all the time and state 2 expects never to be able to detect and punish successfully. So state 2 has *no* incentive to invest resources in detection. And no institutions for the purpose of detection are expected to be observed. Conversely, if the detection probability is one after the optimal investments are made, then state 2 expects always to be able to detect and punish successfully. So state 1 expects never to be able to cheat successfully. Hence state 1 has *no* incentive either to invest resources in evasion or to cheat. And institutions that detect actual cheating are not expected to be observed. Thus the case where  $0 \leq d \leq 1$  is the only case in which institutions are expected to come into existence *and* to be observed detecting actual cheating. These are the two main empirical observations that the neoliberal institutionalist paradigm claims to explain.

36.23 and solve for  $e_1$ :

$$e_1^* = \frac{b_1^2 c_2}{2(2c_1 + c_2)^2}. \quad (36.25)$$

In turn, to obtain state 2's equilibrium investment in detection, substitute equation 36.23 into equation 36.24 and solve for  $d_2$ :

$$d_2^* = \frac{b_1^2 c_1}{(2c_1 + c_2)^2}. \quad (36.26)$$

Each state's optimal investment *increases* as its own investment cost *decreases*.<sup>7</sup> Each state's investment cost is an opportunity cost: the larger the state's resource base is, the lower this opportunity cost is. This cost is thus an indicator of the state's underlying resource power. The greater its resource power is, the lower its investment cost is and the more resources it invests in detection or evasion. Hence the relative capacity of the two states to detect or evade detection is a function of their relative resource power.<sup>8</sup>

At the same time, each state must *make* its investment of  $d_2^*$  or  $e_1^*$  if it is to increase its capacity to detect or evade detection into line with its underlying resource power. If either state fails to make its investment, then it will fail to maximize its income. That is, it will receive a lower income than it can receive by investing. Thus when *either* state has the option to invest in its power to detect or evade detection, the other state is *forced* to invest in its own countervailing power simply to maximize its income. This competition to raise the relative power to detect or evade detection is unavoidable as long as state 1's power to cheat successfully depends on state 2's power to detect cheating (and vice versa).

When each state invests at the optimal level, the equilibrium detection probability is

$$d^* = \frac{d_2^*}{e_1^* + d_2^*} \quad (36.27)$$

$$= \frac{2c_1}{2c_1 + c_2}. \quad (36.28)$$

---

<sup>7</sup>Conversely, each state's optimal investment *decreases* as the other state's investment cost *decreases*.

<sup>8</sup>Each state's resource base is assumed to include all factors of production: land, labor, capital, technology, institutional and social capital.



Figure 36.1 displays this probability for different values of the investment costs  $c_1$  and  $c_2$ . The probability increases as state 2's investment cost declines, but decreases as state 1's investment cost declines. In other words, the probability increases with state 2's resource power but decreases with state 1's resource power (as defined above). So the equilibrium detection probability is a function of relative resource power – provided each state makes its optimal investment in detection or evasion.

At the same time, the equilibrium detection probability is a measure of the *effective terms of trade* between the two states in their exchange of policy commitments. That is, it measures the *actual* distribution of the mutual gains from the exchange policy commitments, as opposed to the formal distribution of the gains codified in the agreement (e.g. an even split). The higher the equilibrium detection probability is, the greater is the fraction of the time that state 2 receives income  $y_{2s}^*$ . Conversely, the lower the detection probability is, the greater is the fraction of the time that state 2 receives the lower income  $y_{2x1}^*$ . Thus state 2's income rises with increases in  $d^*$  but falls with decreases in  $d^*$ . Conversely, state 1's income falls with increases in  $d^*$  but rises with decreases in  $d^*$ . Since  $d^*$  is a function of relative resource power, it follows logically that the distribution of the gains from the exchange of policy commitments is also a function of relative resource power. In short, *the benefit from cooperation is distributed according to relative resource power* – again provided each state makes its optimal investment in detection or evasion.

This conclusion is evident from the two states' equilibrium incomes. These incomes are obtained by substituting  $d_2^*$ ,  $e_1^*$  and  $d^*$  into equations 36.21 and 36.22.

$$y_{1d}^* = (b_2^2 - s) + b_1^2 \left( \frac{c_2^2}{4(2c_1 + c_2)^2} \right) \quad (36.29)$$

$$y_{2d}^* = (b_1^2 - s) - b_1^2 c_2 \left( \frac{4c_1 + c_2}{2(2c_1 + c_2)^2} \right). \quad (36.30)$$

It is evident from these expressions that state 1's income rises with  $c_2$ , while state 2's income declines with  $c_2$ . Figures 36.2 and 36.3 display these incomes for different values of the investment costs  $c_1$  and  $c_2$ . Each state's income increases as its own investment cost falls, but decreases as the other state's investment cost falls. Thus each state's income increases with its own resource power but decreases with the other state's resource power.

Now compare each state's income in this detection equilibrium with its income in the private maximization equilibrium discussed earlier (where each state adopts the privately optimal policy level). Figure 36.4 displays state 1's incomes in the two equilibria for different values of the investment costs  $c_1$  and  $c_2$ . Figure 36.5 displays state 2's incomes similarly. In each figure the state's income from the detection equilibrium is shaded more lightly than its income from the private maximization equilibrium. State 1 receives a higher income in the detection equilibrium than in the private maximization equilibrium for all values of  $c_1$  and  $c_2$ . State 2 receives a higher income in the detection equilibrium than in the private maximization equilibrium provided  $c_2 < c_1$  (roughly). Thus, if this condition is met, each state has an incentive to enter an agreement with the other to raise its policy from the privately optimal level to the socially optimal level. These incentives exist even though both states know that cheating will occur. State 1 has an incentive to *commit* to raise its policy to the socially optimal level even knowing that it will not *fulfill* its commitment a significant portion of the time  $(1 - d^*)$ . State 2 has an incentive to *commit* to raise its policy to the socially optimal level in return even knowing that state 1 will not *fulfill* its commitment a significant portion of the time. Moreover, each state has an incentive to enter the agreement even knowing that the agreement will induce a wasteful competition in raising relative power that may dissipate a significant portion of the gains from the agreement. State 2 has an incentive to enter the agreement even knowing that it will make costly efforts to detect cheating – yet these efforts will be unsuccessful some of the time and hence wasted. So state 1 will be able to cheat successfully at these times. In turn, state 1 has an incentive to enter the agreement even knowing that it will make costly efforts to evade detection – yet these efforts will be unsuccessful some of the time and hence wasted. So it will be unable to cheat at these times.

Now compare each state's income in the detection equilibrium with its income in the social maximization equilibrium (where each state adopts the socially optimal policy level). Figures 36.6 and 36.7 display each state's incomes in the two equilibria for different values of the investment costs  $c_1$  and  $c_2$ . In each figure the state's income from the social maximization equilibrium is shaded more lightly than its income from the detection equilibrium. State 1 receives a higher income in the detection equilibrium than in the social maximization equilibrium, while state 2 receives a lower income in the detection equilibrium than in the social maximization equilibrium. Thus, if the two states are at the social maximization

equilibrium with no cheating and no investments in detection or evasion, then state 1 has an incentive to start cheating (since  $y_{1x1} > y_{1s}$ ). State 2 then has an incentive to start investing in detection (since  $y_{2d} > y_{2x1}$ ). And state 1 then has an incentive to start investing in evasion simply to maximize its income (as explained above). These reactions shift the states to the detection equilibrium.

State 1's income gain from this switch is  $y_{1d}^* - y_{1s}^*$ , while state 2's income loss is  $y_{2s}^* - y_{2d}^*$ . State 1's gain is less than state 2's loss, however. This is evident by comparing the sum of the two states' incomes in each equilibrium. Figure 36.8 shows the sum for the detection equilibrium in the medium shade and the sum for the social maximization equilibrium in the lighter shade. Figure 36.9 shows the joint loss from switching to the detection equilibrium, which may be denoted  $y_{s-d}$ . This is the total loss from the two states' competition to raise relative power. It is instructive to express this loss as a percentage of the potential social gain from the mutual commitment to switch from the privately optimal policy levels to the socially optimal levels (equation 36.15):

$$100 \left( \frac{y_{s-d}^*}{y_{s-p}^*} \right) \quad (36.31)$$

Figure 36.9 displays this percentage for different values of the investment costs  $c_1$  and  $c_2$ . The total loss ranges from 10 to 60 percent of the potential social gain when an agreement is feasible ( $c_2 < c_1$ ). Thus the competition to raise relative power can dissipate more than half of the potential gains from cooperation.

This conclusion underestimates the magnitude of the problem. For this model assumes that only one state invests in opportunism and evasion, and only the other state invests in detection and punishment. If both states invest in opportunism and evasion, and both states invest in detection and punishment, then the loss would be larger. Potentially, the loss could dissipate all of the potential gains from cooperation.

### 3. Figures for this chapter

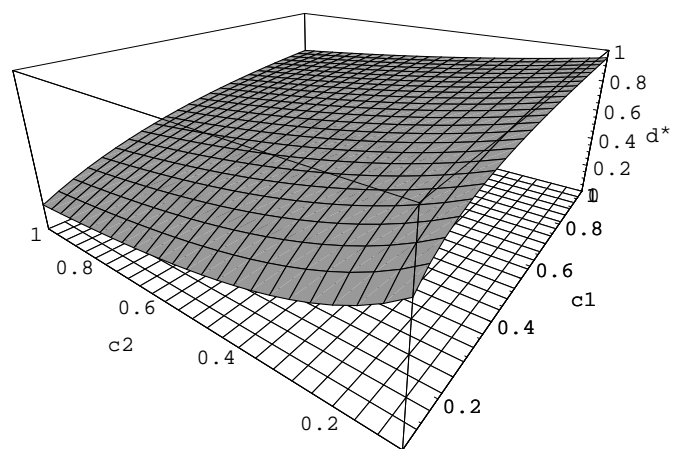


Figure 36.1: Equilibrium detection probability

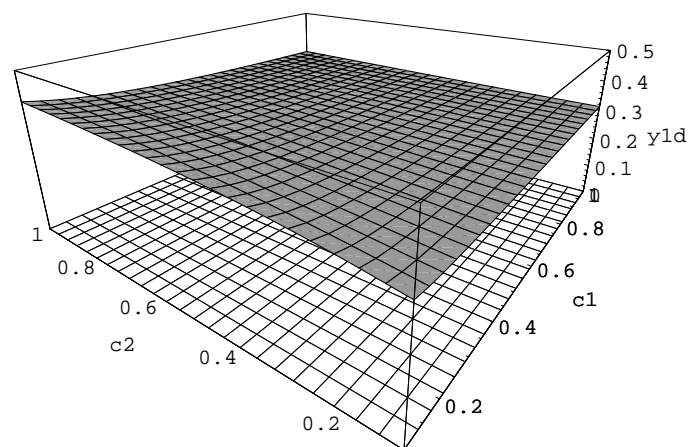


Figure 36.2: State 1's income from the detection equilibrium

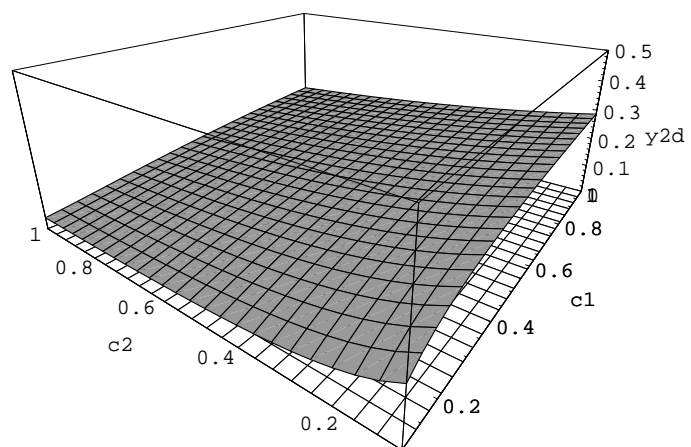


Figure 36.3: State 2's income from the detection equilibrium

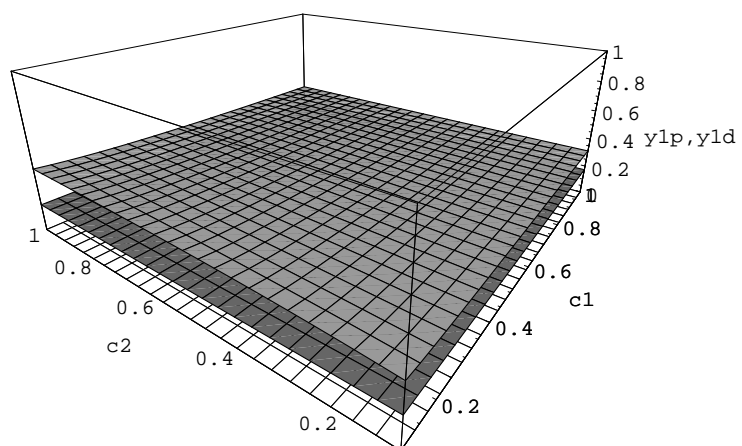


Figure 36.4: State 1's incomes from the detection equilibrium and the private maximization equilibrium

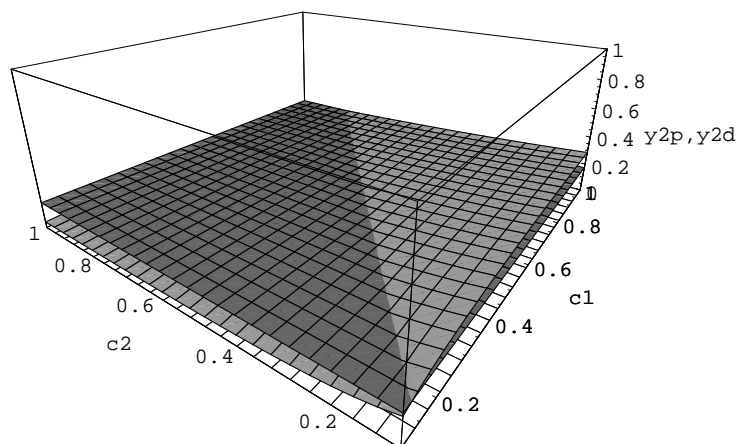


Figure 36.5: State 2's incomes from the detection equilibrium and the private maximization equilibrium



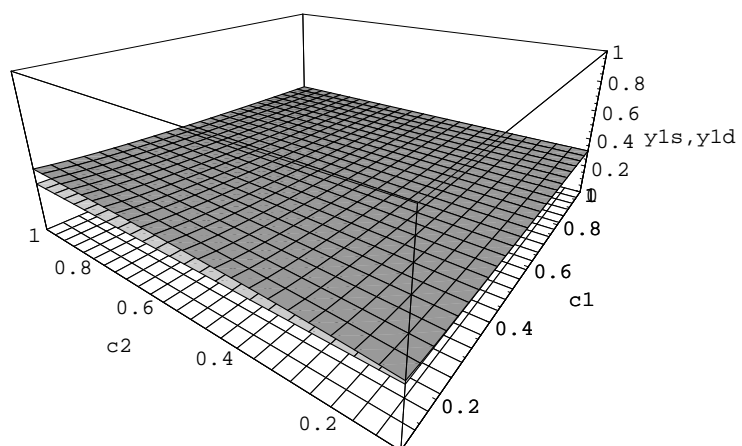


Figure 36.6: State 1's incomes from the detection equilibrium and the social maximization equilibrium

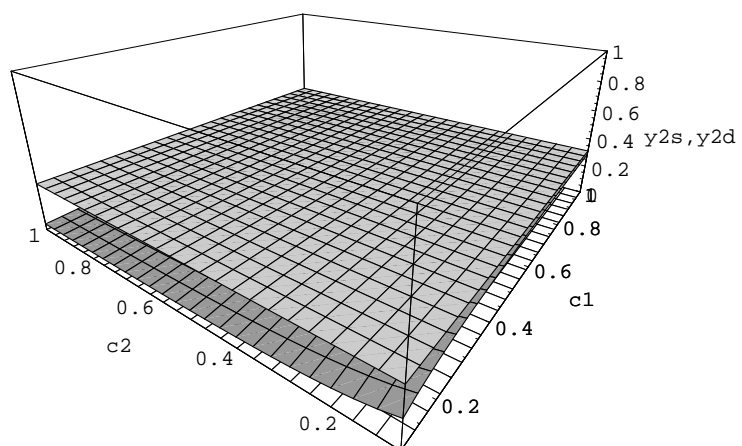


Figure 36.7: State 2's incomes from the detection equilibrium and the social maximization equilibrium

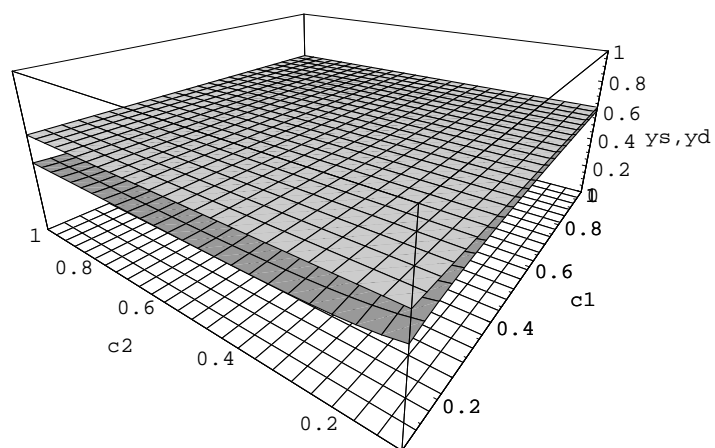


Figure 36.8: Sum of the two states' incomes in the detection equilibrium and in the social maximization equilibrium

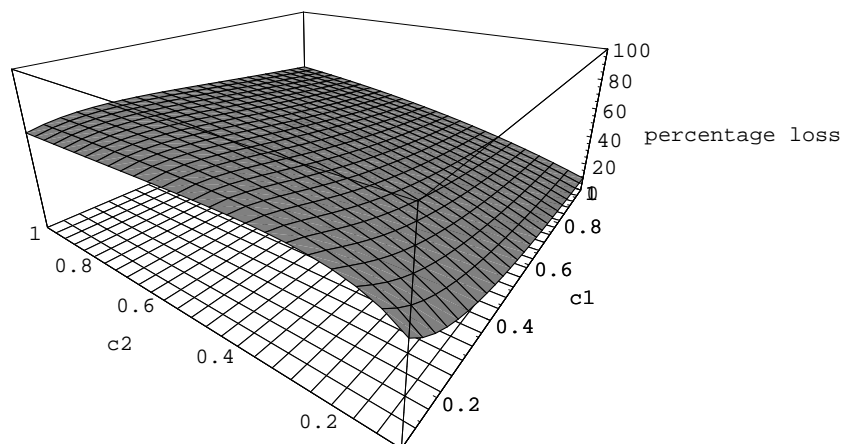


Figure 36.9: Joint income loss from the detection equilibrium as percent of potential joint gain from the social maximization equilibrium

## CHAPTER 37

### Original causal assumptions of this study: Quotations from my earliest studies of the topic

The theory is composed of seven hard core assumptions. Together these assumptions provide a general explanation of conflict and settlement in European history.

**Assumption 1.** Actors divide valuable assets in two different ways. Sometimes actors divide assets according to agreements whose terms vary with changes in the actors' relative power. However, this induces a competition in raising relative power that dissipates all of the gains from the agreements. To avoid this problem, actors divide assets according to agreements whose terms do *not* vary with changes in their relative power. Such an agreement ensures that neither actor can shift the distribution of the asset in his favor by increasing his relative power. The agreement thus avoids costly competitions in raising relative power. I call such agreements anchor term contracts, because the terms refer to anchors in the real world that do not move even when relative power shifts. These are incomplete contracts in that they do not cover all possible contingencies that might arise in the future.<sup>1</sup>

---

<sup>1</sup>The idea of an incomplete contract goes back at least to Klein, Crawford, and Alchian (1978). The idea of a contract whose terms are invariant to shifts in relative power is due to Rowe (1989b). I first invoked these two ideas in my earliest studies of this dissertation topic. In Koppel (1997) and Koppel (1998a), I wrote:

"If the distribution of property rights in a cooperative relationship were determined by the balance of physical power to appropriate rights from others, actors would invest resources in the physical means to overcome others' protective barriers. The ensuing competition in such investments would dissipate more resources than the combined gains of all actors from it. To avoid this wasteful competition, cooperating actors agree to a division of property rights that is not determined by the balance of physical power.... [They] stabilize the terms arbitrarily with respect to the instantaneous value of relative physical power among them. This stability makes the terms a function of historical precedent as well as of other factors."

In Koppel (1998b) I wrote:

"[U]nrestricted rivalry [among states] causes incumbent [governments] and their supporters to face continuous economic losses. States therefore reach quantity restrictions to avoid this 'destructive competition'. These explicit or implicit contracts include prohibitions on interfering in domestic politics, spheres of influence, arms control, and especially, agreement not to pursue

**Assumption 2.** Every action has at least two effects: an intended effect and an unintended effect. Although the intended effect may not violate any anchor term contracts (explicit or implicit), the unintended effect may do so.<sup>2</sup>

**Assumption 3.** When an action has an unintended secondary effect on another actor,

---

distribution of surplus by *continuous* adjustment according to relative military power.

"[S]tates have an incentive to reach inter-subjective understandings about the distribution of wealth in peace (i) to stabilize it arbitrarily with respect to relative power (to avoid wasteful competitions in accumulating bargaining power and using it).... The resulting stability and adjustments make international order 'intersubjective' in that it is a function of historical precedent...."

In Koppel (1998c) I wrote:

"[T]he incentive to avoid wasteful rivalries in accumulating bargaining power leads states to stabilize the distribution of gains from cooperation somewhat arbitrarily with respect to relative power – and thus intersubjectively."

In Koppel (2000b) I wrote:

"[R]elative power does not determine the distribution of gains from mutual restraints on action. For if it did, then actors would *continually* have an incentive to spend resources to enhance their relative power to gain a larger share of the cooperative rents. Without any restraints on action whatever – *independent* of relative power – such rivalry for shares would dissipate all of the cooperative rents! Thus the principal function of rule-following behavior is precisely to specify the distribution of rights to actions, and hence to resources, that does *not* shift with every change in actors' relative power.... [However] actors' mutual [contractual] expectations about permissible actions do not cover *all* feasible actions.... [There are] lacunae in their mutual [contractual] understandings.... [R]ights regarding actions are not fully specified." (Italics in original.)

In Koppel (1997) and Koppel (1998a), I attributed these ideas to Rowe and Klein:

"The model [of hegemonic war] presented here synthesizes and extends five other models from the economics literature that capture critical aspects of international relations.... Rowe (1989b) offers a model of the wage contract in which labor and management strike multi-year bargains, rather than re-contracting more frequently, in order to avoid the wasteful competition in investments in bargaining power that would ensue if the wage rate were adjusted continuously. The series of long-term bargains reflect both parties' interpretations of the 'fair share' of the joint rents due to each party according to a single, implicit perpetual contract. This 'social consensus' is reinterpreted in each bargain in the series by reference to proxy variables that instrument for it and, importantly, are not manipulable by either side in efforts to increase bargaining power....

"Klein (1980, 1981, 1996) develops a general model of vertical relationships in which contract terms are self-enforcing because they ensure both parties a positive share of the rents from continuing cooperation that exceeds the one-time gain from cheating and then having the contract terminated. When unforeseeable, and hence uncontracted, exogenous shifts in costs or demands change parties' expected gains such that exchange at the agreed contract terms no longer satisfies this condition, the relationship is thrown out of the self-enforcing range and becomes susceptible to holdups of one party by another."

<sup>2</sup>As I wrote in Koppel (2006a, p. 19):

"Every action has at least two effects. The primary effect is the one that is intended by the actor. The secondary effect is either not known to the actor or, if known, is not among his or

it may create a problem of legal incompatibility. This is a situation where one contract term gives one actor the legal right to continue its action, but another contract term gives the other actor the legal right to be free of the negative secondary effect of that action. The legal incompatibility causes a breakdown in common knowledge about the legal meaning of the contract terms. When a conflict erupts due to a legal incompatibility, each side must destroy the value of the asset at stake in the dispute to ensure the other side does not gain on net from the dispute, or else it will have an incentive to create similar disputes intentionally in the future.<sup>3</sup>

---

her reasons for taking the action. The problem arises when the primary effect does not violate any other actor's interests or any existing agreements between the actors (formal or informal) – but the secondary effect does.”

I changed the word "state" to the word "actor" for generality.

<sup>3</sup>I first proposed this assumption in my earlier studies of this dissertation topic. In Koppel (1998b) I wrote:

”Hegemonic wars result when changes in one state’s domestic political contract cause it to violate one of these [international] agreements and thereby induce unrestricted [international] rivalry.... [T]he historical data reveal that the last four hegemonic wars were caused not by power shifts but by domestic upheavals that gave one state a domestic political contract *incompatible* with the interests of other states.... [W]hen an irreconcilable dispute over surplus arises, states cannot capture the gains from exchange and avoid the costs of war simply by splitting the difference and agreeing to a new contract. They must destroy the surplus in dispute lest each state become vulnerable to repeated violations from which it eventually loses all of its rights.”

In Koppel (1998c) I wrote:

”The last three hegemonic wars resulted when a domestic upheaval caused one state to violate another’s international rights under [these] implicit contracts...due to the deep structure of incompatibility in rights brought by [the] domestic upheaval.... [T]he states could not simply split the difference in the dispute [peacefully] and capture the remaining gains from cooperation by a new agreement, as this would render the losing state vulnerable to repeated violations in the same way.... [I]nternational order could only be restored by destroying the violator’s gains – through war.... The violated state has an incentive to respond with unrestricted rivalry in the military dimension of competition – war – until the amount of surplus [at stake] in the dispute is destroyed. This ensures that the violator does not gain and have incentive for further violations. After the war, consensus is restored by mutual agreement to another rent-sharing contract.”

In Koppel (1999d) I wrote:

”[W]ars that occur between peer states due to large disputes (e.g. in which the gap between the *inconsistent claims* staked by the two sides is large) will be highly destructive. For each state must destroy a large amount of the other’s resources to deter future violations of its interests.”

In Koppel (2000b) I wrote:

”In an environment where actors’ mutual [contractual] expectations about permissible actions do not cover *all* feasible actions, they sometimes violate each other’s interests unintentionally. Because these lacunae in their mutual [contractual] understandings can be exploited opportunistically, however, the only way for the violated party to deter opportunism is to destroy the resources the violator gained from its action (and then some). However, if the violator acted naively (*believing it was within current [contractual] understandings*), it will

**Assumption 4.** In such a conflict, each side's strategy is to increase its relative power at the other side's expense on all available margins. These strategies induce an unrestricted competition in raising relative power. Each side invests resources in increasing its power up to the point that the marginal benefit from the last unit of investment is just equal to the marginal cost. In equilibrium each side loses more than it gains from engaging in the competition, so it suffers a net loss.<sup>4</sup>

---

feel violated and hence feel it necessary to destroy the *punisher's* gains to deter opportunism! In sum, when rights regarding actions are not fully specified, resources must be destroyed to deter opportunism, and naive violators will be intransigent and fight back." (First, third and fourth italics in original; second italics added for emphasis.)

In Koppel (2006a) I wrote:

"The main assumption of the model is that every foreign policy has at least two effects on the country toward which it is directed. The primary effect is the one that is intended by the policymakers. The secondary effect is either not known to the policymakers or, if known, is not among their reasons for enacting the policy. The problem arises when the primary effect does not violate the other country's interests or any existing agreements between the governments (formal or informal) – but the secondary effect it does. In this situation the foreign government may respond non-cooperatively in backlash against the harmful secondary effect. If the *home government is not aware that its earlier action had caused harm* or, though aware, does not understand that the foreign government's backlash was motivated by this harm, the home government may conclude that the foreign government is untrustworthy (in that it acts unilaterally at the home government's expense). In this case the home government will respond forcefully to punish the backlash and break the foreign government's resistance rather than exercising forbearance for the backlash."

In Koppel (2006b) I wrote:

"Each crisis occurred because the British government was unaware that its imperial policy was inflicting harmful secondary effects on the weaker state. When this state erupted in backlash against these effects, therefore, the British government mistook the backlash as a sign of unprovoked hostility and responded forcefully."

In Koppel (2010b) I wrote:

"When a spillover problem arises between two states, it causes a dispute over which state is obligated to bear the costs of solving the problem. Is the state suffering the spillover effects obligated to accept them? Or is the state producing the spillover effects obligated to stop its activity that is producing them as a byproduct? This is fundamentally a dispute over rights to income. One state is defending its right to be free of the spillover effects because they reduce its income. The other state is defending its right to continue the useful activity that is producing the spillover effects as a byproduct. For if it were required to stop that activity to stop the spillovers, then it would lose the income that activity produces.... Once war breaks out due to this novel cause, the two states will not be able to come to terms peacefully until this cause is remedied – no matter how long they fight or how certain they become about each other's power.... And as long as this novel cause is not remedied, the two states will not be able to reach a stable peace settlement no matter how widely or repeatedly relative power shifts back and forth during the war."

Underlining and italics added for emphasis in the above quotes, except where noted. Also see Koppel (2002b).

<sup>4</sup>The idea of a wasteful competition in raising relative power is due to Rowe (1989b). I first cited Rowe's idea in my earlier studies of this dissertation topic. See quotations in the previous three footnotes. In the field of applied microeconomics, models of "cutthroat competition" identify the conditions under which all



**Assumption 5.** To settle such a conflict, each side must agree to another anchor term contract (see assumption 1 above). Settlements based on anchor term contracts will be stable. Settlements not based on anchor term contracts (e.g. settlements whose terms vary with changes in actors' relative power) will be unstable.

**Assumption 6.** The enforcement of anchor term contracts requires "justice bargains". In this type of bargain, each side commits that it will accept the guilty verdict and pay the penalty voluntarily whenever it is convicted of moving an anchor in its favor intentionally for its own gain. In return for complying with this commitment, each side receives its "variable benefit" whenever an anchor moves in its favor due to random forces beyond its control. If it reneges on its commitment, however, then its variable benefit will be withheld by the other side. This threat enables the contract to be enforced without reliance on costly, power-based enforcement assets. In other words, it enables "anchor-based" enforcement rather than "power-based" enforcement.<sup>5</sup>

**Assumption 7.** To avoid the recurrence of conflict, actors need institutions that have  


---

firms suffer net losses from unrestricted price competition. I first applied this idea to international military competition in Koppel (1997) and Koppel (1998a):

"The production of destructive force is characterized by a U-shaped cost curve due to two factors. First, the cost of a given capacity is fixed because it does not change with the fraction of it utilized in actually supplying destruction. Second, the cost of extracting military resources from the economy and society rises to form a hard capacity constraint due to both the economic production possibilities frontier and the potential for domestic political rivalry if taxes or human costs are raised too high.... [U]nrestricted competition [between two great powers] in the supply of destructive force produces continuous net losses for both powers. The process of war raises both sides' cost curves beyond the point where average cost equal [average] gains.... Since great powers cannot reap net gains by war with other great powers, they have an incentive to avoid the destructive competition that characterizes war through quantity restrictions on the actual supply of military force. An agreement to supply zero quantities characterizes a state of peace.... [Great powers] therefore have incentive to reach explicit or implicit agreements that limit [their competition in] policy areas that affect their relative costs of producing destructive force [such as limits on] foreign interference in domestic politics [and] military clashes over economic access to less developed states.... [Great powers] have the incentive to maintain [military] capacity to deter violations [of these agreements] and to be able to punish them if they occur.... The function of great power war...is to punish violations of one power's international rights at the hands of another by destroying the violator's prospective gains. Such punishment limits losses from the violation and deters future violations."

Also see Koppel (1998b). I provide citations to the economic literature on cutthroat competition in Section 6.1 below. Economic models of conflict assume that each side invests in a conflict technology until marginal benefit equals marginal cost. I first cited and applied these models in Koppel (2000b, 2002a, 2002b, 2009a, 2010a, 2010b). See citations to that economic literature in Part 2 of this study.

<sup>5</sup>This enforcement mechanism is original in this study, as far as I am aware. I first characterized this enforcement mechanism in **Koppel11**.

the capacity to distinguish between intentional and unintentional movements of the anchors specified in the anchor term contracts. The function of these institutions is not only to detect intentional opportunism, but also to manage unintentional disputes that arise from unexpected contingencies and legal incompatibilities (by reforming the existing contracts and forming new ones).<sup>6</sup>

---

<sup>6</sup>While many studies of international and domestic institutions assume that they play an adjudicative role, I am not aware of any studies that define adjudication in terms of the concepts of anchor term contracts and justice bargains (as I have defined them here). See M. R. Garfinkel, McBride, and S. Skaperdas (2012) for a model in which power-based deterrence and "governance based on norms" are substitutes rather than complements. I am not aware of any other studies beside theirs and mine that make this assumption. More generally, institutions for "contract management" are studied in different subfields of political science under different names. For example, see the "managerial" school of thought on international institutions. CITE Chayes.

## Bibliography

- Abernethy, David (2000). *The dynamics of global dominance : European overseas empires, 1415-1980*. New Haven: Yale University Press.
- Amato, J. and H. Shin (2006). “Imperfect common knowledge and the information value of prices”. In: *Economic Theory* 27.1.
- Anderton, C. H. (1989). “ARMS-RACE MODELING - PROBLEMS AND PROSPECTS”. In: *Journal of Conflict Resolution* 33.2.
- Arnold, Benjamin (1985). *German knighthood, 1050-1300*. Oxford [Oxfordshire]; New York: Clarendon Press ; Oxford University Press.
- Asch, Ronald G. (1997). *The Thirty Years War : the Holy Roman Empire and Europe, 1618-48*. New York: St. Martin’s Press.
- Bagwell, Kyle and Robert W. Staiger (1999). “An Economic Theory of GATT”. In: *The American Economic Review* 89.1.
- (2002). *The economics of the world trading system*. Cambridge, Mass.: MIT Press.
- Barnhart, Michael A. (1987). *Japan prepares for total war : the search for economic security, 1919-1941*. Ithaca: Cornell University Press.
- Barzel, Yoram (1997). *Economic analysis of property rights*. Cambridge; New York: Cambridge University Press.
- Bates, Robert H. et al (1998). *Analytic narratives*. Princeton, N.J.: Princeton University Press.
- Beales, Derek Edward Dawson (2009). *Joseph II*. Vol. 2. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.
- Bell, P. M. H. (2007). *The origins of the Second World War in Europe*. 3rd. Harlow ; New York: Longman.
- Benson, John and Takao Matsumura (2001). *Japan, 1868-1945 : from isolation to occupation*. Harlow, England ; New York: Longman an imprint of Pearson Education Limited.
- Bhagwati, J., R. Brecher, and T.N. Srinivasan (1984). “Directly Unproductive Activities and economic theory”. In: *Neoclassical Political Economy*. Ed. by D. Colander. Cambridge, MA: Ballinger.
- Bireley, Robert (2014). *Ferdinand II, Counter-Reformation emperor, 1578-1637*.

- Bittlingmayer, George (1982). "Decreasing Average Cost and Competition: A New Look at the Addyston Pipe Case". In: *The Journal of Law & Economics* 25.2.
- (1985). "Did Antitrust Policy Cause the Great Merger Wave?" In: *The Journal of Law & Economics* 28.1.
- Blackbourn, David (2003). *History of Germany, 1780-1918 : the long nineteenth century*. Malden, MA: Blackwell Pub.
- Blackbourn, David and Geoff Eley (1984). *The peculiarities of German history : bourgeois society and politics in nineteenth-century Germany*. Oxford [Oxfordshire]; New York: Oxford University Press.
- Blanning, T. C. W. (1986). *The origins of the French revolutionary wars*. London; New York: Longman.
- (1996). *The French Revolutionary Wars, 1787-1802*. London; New York; New York: Arnold ; Distributed exclusively in the USA by St. Martin's Press.
- Brawley, Mark R. (1993). *Liberal leadership : great powers and their challengers in peace and war*. Ithaca, N.Y.: Cornell University Press.
- Brightwell, Peter (1979). "The Spanish Origins of the 'Thirty Years' War". In: *European Studies Review* 9.4.
- (1982). "Spain, Bohemia and Europe, 1619-21". In: *European Studies Review* 12.4.
- Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer (1985). "Multimarket Oligopoly: Strategic Substitutes and Complements". In: *Journal of Political Economy* 93.3.
- Carnegie, Allison (2014). "States Held Hostage: Political Hold-Up Problems and the Effects of International Institutions". In: *American Political Science Review* 108.1.
- Carr, Raymond (1982). *Spain, 1808-1975*. Oxford: Clarendon Press.
- Carter, Alice (1968). "Britain as a European power from her Glorious Revolution to the French Revolutionary War". In: *Britain and the Netherlands in Europe and Asia*. Ed. by J.S. Bromley and E.H. Kossmann. London: Macmillan.
- Chan, S. and B. F. Tessen (2009). "Relative Decline: Why Does It Induce War or Sustain Peace?" In: *Systemic Transitions: Past, Present, and Future*.
- Claeys, Gregory (2001). *The Chartist movement in Britain, 1838-1850*. London; Brookfield, Vt.: Pickering & Chatto.

- Clapham, J. H. (1899). *The causes of the war of 1792*. Cambridge: University Press.
- Clark, George (1954). "The Character of The Nine Years War, 1688-1697". In: *Cambridge Historical Journal* 11.2.
- (1970). "The Nine Years War, 1688-1697". In: *The New Cambridge Modern History. Volume 6, The Rise of Great Britain and Russia, 1688-1715/25*. Ed. by J.S. Bromley.
- Cooley, Alexander and Hendrik Spruyt (2009). *Contracting states : sovereign transfers in international relations*. Princeton: Princeton University Press.
- Copeland, Dale C. (2000). *The origins of major war*. Ithaca: Cornell University Press.
- Cronin, Bruce (1999). *Community under anarchy : transnational identity and the evolution of cooperation*. New York: Columbia University Press.
- Crowley, James B. (1966). *Japan's quest for autonomy; national security and foreign policy, 1930-1938*. Princeton, N.J.: Princeton University Press.
- Croxton, Derek (1999). "The Peace of Westphalia of 1648 and the Origins of Sovereignty". In: *The International History Review* 21.3.
- Croxton, Derek and Geoffrey Parker (2009). "'A swift and sure peace': the Congress of Westphalia 1643-1648". In: *The Making of Peace : Rulers, States, and the Aftermath of War*. New York: Cambridge University Press.
- Dalton, Paul (1994). *Conquest, anarchy, and lordship : Yorkshire, 1066-1154*. Cambridge [England] ; New York, NY: Cambridge University Press.
- Demsetz, Harold (1967). "Toward a Theory of Property Rights". In: *The American Economic Review* 57.2.
- Downs, G. W. and D. M. Roake (1994). "CONFLICT, AGENCY, AND GAMBLING FOR RESURRECTION - THE PRINCIPAL-AGENT PROBLEM GOES TO WAR". In: *American Journal of Political Science* 38.2.
- Doyle, William (2002). *The Oxford history of the French Revolution*. Oxford; New York: Oxford University Press.
- Dwyer, Philip G. (2008). *Napoleon : the path to power*. New Haven, CT: Yale University Press.
- Elman, Colin (2004). "Extending Offensive Realism: The Louisiana Purchase and America's Rise to Regional Hegemony". In: *The American Political Science Review* 98.4.

- Epstein, Joshua M. (2006). *Generative social science : studies in agent-based computational modeling*. Princeton: Princeton University Press.
- (2014). *Agent zero : toward neurocognitive foundations for generative social science*.
- Epstein, Joshua M. et al. (1996). *Growing artificial societies : social science from the bottom up*. Washington, D.C. Cambridge, Mass. ; Brookings Institution Press ; MIT Press.
- Evans, Robert John Weston (1979). *The making of the Habsburg monarchy, 1550-1700 : an interpretation*. Oxford New York: Clarendon Press ; Oxford University Press.
- Evera, Stephen C. Full publication date Spring van (1998). “Offense, Defense, and the Causes of War”. In: *International Security* 22.4.
- Farrell, Henry and Abraham Newman (2016). “The new interdependence approach: Theoretical development and empirical demonstration”. In: *Review of International Political Economy*.
- Fearon, James D. (1995). “Rationalist Explanations for War”. In: *International Organization* 49.3.
- Feinstein, C. H., Peter Temin, and Gianni Toniolo (1997). *The European economy between the wars*. Oxford [England] ; New York: Oxford University Press.
- Findlay, Ronald (1996). “Toward a model of territorial expansion and the limits of empire”. In: *The political economy of conflict and appropriation*. Ed. by Michelle Garfinkel and Stergios Skaperdas. New York: Cambridge University Press.
- Forrest, Alan I. (1995). *The French Revolution*. Oxford, U.K.; Cambridge, Mass., U.S.A.: Blackwell.
- Fritzsche, Peter (2008). *Life and death in the Third Reich*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Garfinkel, M. R., M. McBride, and S. Skaperdas (2012). “Governance and Norms as Determinants of Arming”. In: *Revue D Economie Politique* 122.2.
- Garfinkel, Michelle R. (1990). “Arming as a Strategic Investment in a Cooperative Equilibrium”. In: *The American Economic Review* 80.1.
- Garrard, John (2002). *Democratisation in Britain : elites, civil society, and reform since 1800*. Houndmills, Basingstoke, Hampshire; New York: Palgrave.
- Garrard, John, Vera Tolz, and Ralph White (2000). *European democratization since 1800*. Houndmills, Basingstoke, Hampshire; New York: Macmillan ; St. Martin’s Press.

- Gibbs, G.C. (1969). "The revolution in foreign policy". In: *Britain after the Glorious Revolution: 1689-1714*. Ed. by G. Holmes. London: St. Martin's Press.
- Gideon, Rose (1998). "Neoclassical Realism and Theories of Foreign Policy". In: *World Politics* 51.1.
- Gilpin, Robert (1981). *War and change in world politics*. Cambridge ; New York: Cambridge University Press.
- Gindely, Antonín and Andrew Ten Brook (1884). *History of the thirty years' war*. New York: G. P. Putnam's.
- Glaser, Charles L. (1994). "Realists as Optimists: Cooperation as Self-Help". In: *International Security* 19.3.
- Goddard, Stacie E. (2006). "Uncommon Ground: Indivisible Territory and the Politics of Legitimacy". In: *International Organization* 60.1.
- Goertz, Gary and James Mahoney (2012). *A tale of two cultures : qualitative and quantitative research in the social sciences*. Princeton, N.J.: Princeton University Press.
- Goldstein, Joshua S. (1988). *Long cycles : prosperity and war in the modern age*. New Haven: Yale University Press.
- Greengrass, Mark (1991). *Conquest and coalescence : the shaping of the state in early modern Europe*. New York: E. Arnold : Distributed in the USA by Routledge, Chapman and Hall.
- Grossman, Sanford J. and Oliver Hart (1986). "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". In: *Journal of Political Economy* 94.4.
- Gulick, Edward Vose (1955). *Europe's classical balance of power; a case history of the theory and practice of one of the great concepts of European statecraft*. Ithaca: Cornell University Press for the American Historical Association.
- (1957). "The final coalition and the Congress of Vienna, 1813-1815". In: *The New Cambridge Modern History. Volume 9, War and Peace in an Age of Upheaval, 1793-1830*. Ed. by C.W. Crawley. New York: Cambridge University Press.
- Hanneman, Mary L. (2001). *Japan faces the world, 1925-1952*. Harlow, England ; New York: Longman.
- Haslam, Jonathan (2002). *No virtue like necessity : realist thought in international relations since Machiavelli*. New Haven: Yale University Press.

- Hassner, R. E. (2003). “‘To halve and to hold’: Conflicts over sacred space and the problem of indivisibility”. In: *Security Studies* 12.4.
- Hensel, Paul R. and Sara McLaughlin Mitchell (2005). “Issue indivisibility and territorial claims”. In: *GeoJournal* 64.4.
- Hermalin, B., A. Katz, and R. Craswell (2007). “The law and economics of contracts”. In: *Handbook of Law and Economics*. Ed. by M. Polinsky and S. Shavell. London: Elsevier.
- Hidden, John (1993). *Germany and Europe, 1919-1939*. 2nd. London ; New York: Longman.
- Hirschman, Albert O. (1945). *National power and the structure of foreign trade*. Berkeley and Los Angeles: University of California Press.
- Hirshleifer, Jack (1989). “Conflict and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success”. In: *Public Choice* 63.2.
- (1991). “The Technology of Conflict as an Economic Activity”. In: *The American Economic Review* 81.2.
- (1995). “Theorizing about conflict”. In: *Handbook of Defense Economics*. Ed. by Keith Hartley and Todd Sandler. New York: Elsevier.
- Hochedlinger, Michael (2003). *Austria’s wars of emergence : war, state and society in the Habsburg monarchy, 1683-1797*. Harlow: Longman.
- Hudson, John (2010). “Feud, Vengeance and Violence in England c. 900-1200”. In: ed. by B. Tuten.
- Hyams, Paul (2001). “Feud and the State in Late Anglo-Saxon England”. In: *Journal of British Studies* 40.
- Ikenberry, G. John (2001). *After victory : institutions, strategic restraint, and the rebuilding of order after major wars*. Princeton: Princeton University Press.
- Illing, P. (2009). “The Brabant Revolution and the Western Question (1787-1790)”. In: *Dutch Crossing-Journal of Low Countries Studies* 33.1.
- Intriligator, M. D. and D. L. Brito (1976). “FORMAL MODELS OF ARMS RACES”. In: *Journal of Peace Science* 2.1.
- (1984). “CAN ARMS RACES LEAD TO THE OUTBREAK OF WAR”. In: *Journal of Conflict Resolution* 28.1.
- (1986). “ARMS RACES AND INSTABILITY”. In: *Journal of Strategic Studies* 9.4.



- Intriligator, Michael D. (1975). "Strategic Considerations in the Richardson Model of Arms Races". In: *Journal of Political Economy* 83.2.
- Iriye, Akira (1997). *Japan and the wider world : from the mid-nineteenth century to the present*. London ; New York: Longman.
- James, Patrick (1995). "Structural Realism and the Causes of War". In: *Mershon International Studies Review* 39.2.
- Jervis, Robert (1982). "Security Regimes". In: *International Organization* 36.2.
- (1985). "From Balance to Concert: A Study of International Security Cooperation". In: *World Politics* 38.1.
- Jones, J. R. (1978). *Country and court : England, 1658-1714*. Cambridge, Mass.: Harvard University Press.
- (1968). "English attitudes to Europe in the seventeenth century". In: *Britain and the Netherlands in Europe and Asia*. Ed. by J.S. Bromley and E.H. Kossmann. London: Macmillan.
- (1980). *Britain and the world 1649-1815*. Brighton, Sussex Atlantic Highlands, N.J.: Harvester Press ; Humanities Press.
- Kaiser, David E. (1990). *Politics and war : European conflict from Philip II to Hitler*. London: Tauris.
- Kaplan, Benjamin J. (2007). *Divided by faith : religious conflict and the practice of toleration in early modern Europe*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Katz, Michael (1989). "Vertical Contractual Relations". In: *Handbook of Industrial Organization, Vol. 1*. Ed. by Richard Schmalensee and Robert Willig. New York: North-Holland.
- Kennedy, Paul M. (1987). *The rise and fall of the great powers : economic change and military conflict from 1500 to 2000*. 1st. New York, NY: Random House.
- Keohane, Robert O. (1984). *After hegemony : cooperation and discord in the world political economy*. Princeton, N.J.: Princeton University Press.
- Kindleberger, Charles P. (1989). "Commercial policy between the wars". In: *The Cambridge economic history of Europe, Volume 8. The industrial economies: The development of economic and social policies*. Ed. by Peter Mathias and Sidney Pollard.
- Kindleberger, Charles Poor (1973). *The world in depression, 1929-1939*. Berkeley: University of California Press.

- King, Gary, Robert O. Keohane, and Sidney Verba (1994). *Designing social inquiry : scientific inference in qualitative research*. Princeton, N.J.: Princeton University Press.
- Kissinger, Henry (1957). *A world restored; Metternich, Castlereagh and the problems of peace, 1812-22*. Boston: Houghton Mifflin.
- Klein, Benjamin (1996). "Why hold-ups occur: The self-enforcing range of contractual relationships". In: *Economic Inquiry* 34.3.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978). "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process". In: *The Journal of Law & Economics* 21.2.
- Klein, Benjamin and Kevin M. Murphy (1988). "Vertical Restraints as Contract Enforcement Mechanisms". In: *The Journal of Law & Economics* 31.2.
- Klemperer, P. (1995). "COMPETITION WHEN CONSUMERS HAVE SWITCHING COSTS - AN OVERVIEW WITH APPLICATIONS TO INDUSTRIAL-ORGANIZATION, MACROECONOMICS, AND INTERNATIONAL-TRADE". In: *Review of Economic Studies* 62.4.
- Knorr, Klaus (1947). "Economics and International Relations: A Problem in Teaching". In: *Political Science Quarterly* 62.4.
- (1957). "The Concept of Economic Potential for War". In: *World Politics* 10.1.
- Koppel, William E. (1997). "Strategy on the margin: How great powers respond to major rivals' domestic upheavals". Unpublished Work. Manuscript submitted to dissertation advisor, Department of Political Science, UCLA.
- (2007). "Shifting power, disputes and war: A limited information model". Unpublished Work. Manuscript, Department of Political Science, UCLA.
- (2008). "Disputes, relative power and bargains that avoid war: A limited information model". Unpublished Work. Manuscript, Department of Political Science, UCLA.
- (2010b). "A new theory of hegemonic war". Unpublished Work. Discussion paper submitted to doctoral dissertation committee for oral examination, Department of Political Science, UCLA.
- (2006b). "Abstract of: Hegemonic backlash and domestic accountability: The causes and domestic political consequences of Britain's imperial policy failures in the 19th century". Unpublished Work. E-mail message to dissertation advisor and other professors, Department of Political Science, UCLA.

- Koppel, William E. (1998b). "Achieving security cooperation with states in transition: How great powers respond to major rivals' domestic upheavals". Unpublished Work. Fellowship proposal submitted to dissertation advisor, UCLA, and UC Institute on Global Conflict & Cooperation.
- (1998a). "An application of core theory to the analysis of hegemonic war: How great powers respond to major rivals' domestic upheavals". Unpublished Work. Manuscript submitted to dissertation advisor, Department of Political Science, UCLA.
- (1999d). "Discriminating between motives for war: Net gains maximization following a favorable power shift versus Net loss minimization following an exogenous income loss". Unpublished Work. Final term paper, course 233C, Department of Political Science, UCLA.
- (2002b). "Domestic upheaval and hegemonic war: How secessions, revolutions and depressions lead to aggression and expansion". Unpublished Work. Fellowship proposal submitted to dissertation advisor, UCLA, and UC Institute on Global Conflict & Cooperation (awarded).
- (2006a). "Hegemonic backlash and domestic accountability: The causes and domestic political consequences of Britain's imperial policy failures in the 19th century". Unpublished Work. Manuscript, Department of Political Science, UCLA.
- (2000b). "Power, institutions and social change: A new theory of hegemonic war and the evolution of issue dimensions in international relations". Unpublished Work. Fellowship proposal submitted to dissertation advisor, UCLA, and UC Institute on Global Conflict & Cooperation.
- (1998c). "Strategy on the margin: How Great Powers Respond to Major Rivals' Domestic Upheavals". Unpublished Work. Fellowship proposal submitted to dissertation advisor, UCLA, and Committee on International Peace & Security, Social Science Research Council.
- Krainin, C. (2017). "Preventive War as a Result of Long-Term Shifts in Power". In: *Political Science Research and Methods* 5.1.
- Krasner, Stephen D. (1976). "State Power and the Structure of International Trade". In: *World Politics* 28.3.

- Kugler, Jacek and Douglas Lemke (1996). *Parity and war : evaluations and extensions of The war ledger*. Ann Arbor, Mich.: University of Michigan Press.
- Kupchan, Charles (1994). *The vulnerability of empire*. Ithaca: Cornell University Press.
- Kupchan, Charles A. and Clifford A. Kupchan (1991). "Concerts, Collective Security, and the Future of Europe". In: *International Security* 16.1.
- Kydd, Andrew (1997). "Sheep in sheep's clothing: Why security seekers do not fight each other". In: *Security Studies* 7.1.
- Labs, Eric (1997). "Beyond victory: Offensive realism and the expansion of war aims". In: *Security Studies* 6.4.
- Lakatos, Imre (1970). "Falsification and the Methodology of Scientific Research Programmes". In: 4. Ed. by Alan Musgrave.
- Lake, David A. (1993). "Leadership, Hegemony, and the International Economy: Naked Emperor or Tattered Monarch with Potential?" In: *International Studies Quarterly* 37.4.
- (2007). "Escape from the State of Nature: Authority and Hierarchy in World Politics". In: *International Security* 32.1.
- Lane, Frederic (1942). "The economic meaning of war and protection". In: *Journal of social philosophy and jurisprudence* 7.3.
- (1950). "Oceanic Expansion: Force and Enterprise in the Creation of Oceanic Commerce". In: *The Journal of Economic History* 10.
- (1958). "Economic Consequences of Organized Violence". In: *The Journal of Economic History* 18.4.
- Langford, Paul (1976). *Modern British Foreign Policy: The eighteenth century, 1688-1815*. London: A. and C. Black.
- Langlois, Richard N. (1986a). *Economics as a process : essays in the new institutional economics*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.
- (1986b). "Rationality, institutions, and explanation". In: *Economics as a process : essays in the new institutional economics*. Ed. by Richard N. Langlois. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.
- Latsis, Spiro J. (1980). *Method and appraisal in economics*. 1st paperback. Cambridge ; New York: Cambridge University Press.

- Layne, Christopher (1993). "The Unipolar Illusion: Why New Great Powers Will Rise". In: *International Security* 17.4.
- (2006). "The Unipolar Illusion Revisited: The Coming End of the United States' Unipolar Moment". In: *International Security* 31.2.
- Lee, Marshall M. and Wolfgang Michalka (1987). *German foreign policy, 1917-1933 : continuity or break?* Leamington Spa [Warwickshire] ; New York New York: Berg ; Distributed exclusively in the US and Canada by St. Martin's Press.
- Leidy, M. P. and R. W. Staiger (1985). "ECONOMIC-ISSUES AND METHODOLOGY IN ARMS-RACE ANALYSIS". In: *Journal of Conflict Resolution* 29.3.
- Leitz, Christian (2004). *Nazi foreign policy, 1933-1941 : the road to global war*. London ; New York: Routledge.
- Lesaffer, Randall (2004). *Peace treaties and international law in European history : from the late Middle Ages to World War One*. Cambridge, UK ; New York: Cambridge University Press.
- Levy, Jack S. (1985). "Theories of General War". In: *World Politics* 37.3.
- (1986). "Organizational Routines and the Causes of War". In: *International Studies Quarterly* 30.2.
- (1987). "Declining Power and the Preventive Motivation for War". In: *World Politics* 40.1.
- (2008). "Preventive War and Democratic Politics". In: *International Studies Quarterly* 52.1.
- Lipson, Charles (1994). "Is the future of collective security like the past?" In: *Collective security beyond the Cold War*. Ed. by George W. Downs. Ann Arbor: University of Michigan Press.
- Lord, Evelyn (2004). *The Stuarts' secret army : English Jacobites, 1689-1752*. Harlow, England ; New York: Pearson Longman.
- Lynch, John (1981). *Spain under the Habsburgs*. Oxford, England: Basil Blackwell.
- Macartney, C. A. (1978). *The House of Austria : the later phase, 1790-1918*. Edinburgh: University Press.

- MacHardy, Karin Jutta (2003). *War, religion and court patronage in Habsburg Austria : the social and cultural dimensions of political interaction, 1521-1622*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave.
- Maland, David (1980). *Europe at war, 1600-1650*. London: MacMillan.
- McKay, Derek and H. M. Scott (1983). *The rise of the great powers, 1648-1815*. London ; New York: Longman.
- McPhee, Peter (2002). *The French Revolution, 1789-1799*. Oxford [England] ; New York: Oxford University Press.
- Mearsheimer, John J. (2001). *The tragedy of Great Power politics*. 1st. New York: W.W. Norton.
- Miller, Steven E. (1985). *Military strategy and the origins of the First World War : an International security reader*. Princeton, N.J.: Princeton University Press.
- Mirowski, Philip (1988). *Against mechanism : protecting economics from science*. Totowa, N.J.: Rowman Littlefield.
- (1989). *More heat than light : economics as social physics, physics as nature's economics*. Cambridge ; New York: Cambridge University Press.
- Mitzen, Jennifer (2013). *Power in concert : the nineteenth-century origins of global governance*. Chicago: University of Chicago Press.
- Modelski, George (1987). *Exploring long cycles*. Boulder London: L. Rienner Publishers ; Frances Pinter (Publishers).
- Modelski, George C. (1978). “The Long Cycle of Global Politics and the Nation-State”. In: *Comparative Studies in Society and History* 20.2.
- Morgan, T. C. and K. N. Bickers (1992). “DOMESTIC DISCONTENT AND THE EXTERNAL USE OF FORCE”. In: *Journal of Conflict Resolution* 36.1.
- Morgenthau, Hans Joachim (1978 [1948]). *Politics among nations : the struggle for power and peace*. 5th. New York: Knopf.
- Morris, William Alfred (1927). *The medieval English sheriff to 1300*. Manchester London, New York, etc.: The University Press; Longmans, Green co. ltd.
- Mortimer, Geoff (2015). *The origins of the Thirty Years War and the revolt in Bohemia, 1618*.

- Narang, Vipin and Rebecca M. Nelson (2009). “Who Are These Belligerent Democratizers? Reassessing the Impact of Democratization on War”. In: *International Organization* 63.2.
- Niou, Emerson and Peter C. Ordeshook (1986). “A Theory of the Balance of Power in International Systems”. In: *The Journal of Conflict Resolution* 30.4.
- (1987). “Preventive War and the Balance of Power: A Game-Theoretic Approach”. In: *The Journal of Conflict Resolution* 31.3.
- Nish, Ian (2002). *Japanese foreign policy in the interwar period*. Westport, CT: Praeger Publishers.
- Nordmann, Claude (1976). “Louis XIV and the Jacobites”. In: *Louis XIV and Europe*. Ed. by Ragnhild Hatton. Columbus, OH: Ohio State University Press.
- Organski, A. F. K. (1958). *World politics*. New York: Knopf.
- Organski, A. F. K. and Jacek Kugler (1980). *The war ledger*. Chicago: University of Chicago Press.
- Osiander, Andreas (1994). *The states system of Europe, 1640-1990 : peacemaking and the conditions of international stability*. Oxford Oxford ; New York: Clarendon Press ; Oxford University Press.
- Overy, R. J. and Andrew Wheatcroft (1989). *The road to war*. London: Macmillan London.
- Pages, Georges (1970). *The Thirty Years War, 1618-1648*. New York: Harper.
- Parker, Geoffrey (1988). *The military revolution : military innovation and the rise of the west, 1500-1800*. Cambridge [England] ; New York: Cambridge University Press.
- Parker, Geoffrey and Simon Adams (1997). *The Thirty Years’ War*. 2nd [rev.] London ; New York: Routledge.
- Patch, William L. (1998). *Heinrich Brüning and the dissolution of the Weimar Republic*. Cambridge, U.K. ; New York: Cambridge University Press.
- Peattie, Mark R. (1975). *Ishiwara Kanji and Japan’s confrontation with the West*. Princeton, N.J.: Princeton University Press.
- Pirrong, Stephen Craig (1989). *Market structure when the core is empty: Horizontal restrictions, vertical restrictions and the coordination of production*. Manuscript.
- (1992). “An Application of Core Theory to the Analysis of Ocean Shipping Markets”. In: *The Journal of Law & Economics* 35.1.
- Polišenský, Josef V. (1971a). *The thirty years war*. Berkeley: University of California Press.

- Polišenský, Josef V. (1971b). *The thirty years war*. Berkeley: University of California Press.
- Posen, Barry (1984). *The sources of military doctrine : France, Britain, and Germany between the world wars*. Ithaca: Cornell University Press.
- (1993). “The security dilemma and ethnic conflict”. In: *Survival: Global Politics and Strategy* 35.1.
- Powell, Robert (1993). “Guns, Butter, and Anarchy”. In: *The American Political Science Review* 87.1.
- (2002). “Bargaining theory and international conflict”. In: *Annual Review of Political Science* 5.
- (2006). “War as a Commitment Problem”. In: *International Organization* 60.1.
- Ragin, Charles C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.
- Rasmusen, Eric (1989). *Games and information : an introduction to game theory*. Oxford, UK; New York: Basil Blackwell.
- Richards, D. et al. (1993). “GOOD TIMES, BAD TIMES, AND THE DIVERSIONARY USE OF FORCE - A TALE OF SOME NOT-SO-FREE AGENTS”. In: *Journal of Conflict Resolution* 37.3.
- Richardson, Louise (1999). “The Concert of Europe and Security Management in the Nineteenth Century”. In: *Imperfect unions : security institutions over time and space*. Ed. by Helga Haftendorn, Robert O. Keohane, and Celeste A. Wallander. Oxford; New York: Oxford University Press.
- Ringrose, David R. (1996). *Spain, Europe, and the "Spanish miracle", 1700-1900*. Cambridge; New York, NY, USA: Cambridge University Press.
- Roberts, Stephen (2003). *The People's Charter : democratic agitation in early Victorian Britain*. London: Merlin.
- Roider, Karl A. (1987). *Baron Thugut and Austria's response to the French Revolution*. Princeton, N.J.: Princeton University Press.
- Ross, Steven T. (1973). *Quest for victory; French military strategy, 1792-1799*. South Brunswick: A. S. Barnes.
- Rothwell, Victor (2001). *The origins of the Second World War*. Manchester ; New York New York: Manchester University Press ; Distributed exclusively in the USA by Palgrave.
- Rowe, Nicholas (1989b). *Rules and institutions*. Ann Arbor: University of Michigan Press.



- Savage, G. (1998). "Favier's heirs: The French Revolution and the secret-du-roi (Louis XV, diplomacy, policy)". In: *Historical Journal* 41.1.
- Savic, I. and Z. C. Shirkey (2009). "Trust in the balance: Asymmetric information, commitment problems and balancing behavior". In: *Journal of Theoretical Politics* 21.4.
- Schroeder, Paul W. C. (1992). "Did the Vienna Settlement Rest on a Balance of Power?" In: *The American Historical Review* 97.3.
- Schroter, Harm (1996). "Cartelization and decartelization in Europe, 1870-1995: Rise and decline of an economic institution". In: *Journal of European Economic History* 25.1.
- Schweller, Randall L. (1996). "Neorealism's status-quo bias: What security dilemma?" In: *Security Studies* 5.3.
- (2003). "The progressiveness of neoclassical realism". In: *Progress in International Relations: Appraising the Field*. Ed. by Colin Elman and Miriam Elman. Cambridge, MA: MIT Press.
- Schweller, Randall L. C. (1994). "Bandwagoning for Profit: Bringing the Revisionist State Back In". In: *International Security* 19.1.
- (2004). "Unanswered Threats: A Neoclassical Realist Theory of Underbalancing". In: *International Security* 29.2.
- Sheehan, James J. (1989). *German history, 1770-1866*. Oxford [England] New York: Clarendon Press ; Oxford University Press.
- Sheehan, Michael (1996). *The balance of power : history and theory*. London ; New York: Routledge.
- Shubert, Adrian and José Alvarez Junco (2000). *Spanish history since 1808*. London; New York: Arnold ; Co-published in the United States of America by Oxford University Press.
- Simmons, Beth A. (1994). *Who adjusts? : domestic sources of foreign economic policy during the interwar years*. Princeton, N.J.: Princeton University Press.
- Simms, Brendan (2007). *Three victories and a defeat: The rise and fall of the first British Empire, 1714-1783*. New York: Basic Books.
- Sinnreich, Richard Hart (2009). "In search of military repose: the Congress of Vienna and the making of peace". In: *The making of peace : rulers, states, and the aftermath of war*. Ed. by Williamson Murray and Jim Lacey. Cambridge [England] ; New York: Cambridge University Press.

- Sjostrom, William (1989). "Collusion in Ocean Shipping: A Test of Monopoly and Empty Core Models". In: *Journal of Political Economy* 97.5.
- Skaperdas, Stergios (1992). "Cooperation, Conflict, and Power in the Absence of Property Rights". In: *The American Economic Review* 82.4.
- (1996). "Contest Success Functions". In: *Economic Theory* 7.2.
- Slantchev, Branislav (2005). "Territory and commitment: The Concert of Europe as a self-enforcing equilibrium". In: *Security Studies* 14.4.
- Smith, Angel (2000). "The corporatist threat and the overthrow of the Spanish Second Republic". In: *European democratization since 1800*. Ed. by John Garrard, Vera Tolz, and Ralph White. Houndmills, Basingstoke, Hampshire; New York: Macmillan ; St. Martin's Press.
- Snell, John L. and Hans A. Schmitt (1976). *The democratic movement in Germany, 1789-1914*. Chapel Hill: University of North Carolina Press.
- Snidal, Duncan (1985). "The Limits of Hegemonic Stability Theory". In: *International Organization* 39.4.
- Snyder, Jack L. (1991). *Myths of empire : domestic politics and international ambition*. Ithaca, N.Y.: Cornell University Press.
- T.N. Srinivasan, ed. (1997). "Special issue on "the determinacy paradox"". In: *Economics and Politics* 9.3.
- Telser, Lester G. (1987). *A theory of efficient cooperation and competition*. Cambridge; New York: Cambridge University Press.
- Tooze, Adam (2006). *The wages of destruction : the making and breaking of the Nazi economy*. London ; New York: Allen Lane.
- Turner, Henry Ashby (1985). *German big business and the rise of Hitler*. New York: Oxford University Press.
- Turner, Ralph V. (1968). *The king and his courts; the role of John and Henry III in the administration of justice, 1199-1240*. Ithaca, N.Y.: Cornell University Press.
- Van Evera, Stephen (1984). "The Cult of the Offensive and the Origins of the First World War". In: *International Security* 9.1.
- (1999). *Causes of war : power and the roots of conflict*. Ithaca: Cornell University Press.
- Walt, Stephen M. (1987). *The origins of alliances*. Ithaca: Cornell University Press.

- Walt, Stephen M. C. (1989). "The Case for Finite Containment: Analyzing U.S. Grand Strategy". In: *International Security* 14.1.
- Waltz, Kenneth N. (1979). *Theory of international politics*. 1st. New York: McGraw-Hill.
- (2000). "Structural Realism after the Cold War". In: *International Security* 25.1.
- Wangermann, Ernst (1969). *From Joseph II to the Jacobin trials: government policy and public opinion in the Habsburg dominions in the period of the French Revolution*. 2nd. London: Oxford U.P.
- Waswo, Ann (1996). *Modern Japanese society, 1868-1994*. Oxford ; New York: Oxford University Press.
- Webb, Michael C. and Stephen D. C. Krasner (1989). "Hegemonic Stability Theory: An Empirical Assessment". In: *Review of International Studies* 15.2.
- Weber, Katja (2000). *Hierarchy amidst anarchy : transaction costs and institutional choice*. Albany, N.Y.: State University of New York Press.
- Webster, Charles K. (1925). *The foreign policy of Castlereagh, 1815-1822 : Britain and the European alliance*. London: G. Bell and Sons.
- Wedgwood, C. V. (1999). *The Thirty Years' War*. London: Folio Society.
- Wendt, Alexander (1999). *Social theory of international politics*. Cambridge, UK; New York: Cambridge University Press.
- Whaley, Joachim (2012b). *Germany and the Holy Roman Empire*. Vol. 2. Oxford: Oxford University Press.
- Williamson, Oliver E. (1975). *Markets and hierarchies, analysis and antitrust implications : a study in the economics of internal organization*. New York: Free Press.
- (1979). "Transaction-Cost Economics: The Governance of Contractual Relations". In: *The Journal of Law & Economics* 22.2.
- (1985). *The economic institutions of capitalism : firms, markets, relational contracting*. New York London: Free Press ; Collier Macmillan.
- Wilson, Peter H. (2008). *The Thirty Years War : Europe's Tragedy*. Cambridge, MA: Harvard University Press.
- Wohlforth, William C. C. (1994). "Realism and the End of the Cold War". In: *International Security* 19.3.
- (1999). "The Stability of a Unipolar World". In: *International Security* 24.1.

- Wohlforth, William C. C. (2007). "Unipolar Stability: The Rules of Power Analysis". In: *Harvard International Review* 29.1.
- (2008). "Realism". In: *Oxford Handbook of International Relations*. Ed. by Christian Reus-Smit and Duncan Snidal. New York: Oxford University Press.
- Zakaria, Fareed (1992). "Realism and Domestic Politics: A Review Essay". In: *International Security* 17.1.
- (1998). *From wealth to power : the unusual origins of America's world role*. Princeton, N.J.: Princeton University Press.