

UCLA

UCLA Electronic Theses and Dissertations

Title

The Form of Law: Practical Principles and the Foundations of Kant's Moral Theory

Permalink

<https://escholarship.org/uc/item/3s91d8t5>

Author

Reckner, William Leland

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Form of Law:
Practical Principles and the Foundations
of Kant's Moral Theory

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Philosophy

by

William Leland Reckner

2017

© Copyright by

William Leland Reckner

2017

ABSTRACT OF THE DISSERTATION

The Form of Law:
Practical Principles and the Foundations
of Kant's Moral Theory

by

William Leland Reckner

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2017

Professor Barbara Herman, Chair

Immanuel Kant argued that morality requires us to act on principles that we can will as universal laws. However, there has always been profound disagreement about how to apply this requirement, and about why this demand should be morally fundamental. This dissertation offers new answers to these questions, by developing a deeper understanding of the “practical” principles that Kant wants us to be able to will as universal laws.

My primary thesis is that practical principles state three things: a reason to act, the end or goal that this reason requires us to accomplish, and the means that we must use to achieve that end. Several crucial lines of Kant's thought require this structure for practical principles, I argue. Primarily: practical principles, as principles for action, must also be causal principles, and Kant's views on causation require practical principles to have the structure I propose.

I use this structure to answer some venerable problems: first, why can't we will any principle whatsoever as a universal law, especially if we make each principle unique to each case? Because Kant requires causal relations to be rule-governed, I argue, in a way that prevents practical principles from being unique to each situation.

Next, why isn't morality satisfied when everyone could follow our principles? Under my interpretation, practical principles say how we are required to act. As rational beings, though, we cannot be required to contradict ourselves. Everyone could contradict themselves in order to follow a principle, but we cannot be required to do so. So I argue that "willing" our principles as universal laws is supposed to capture how principles cannot require us to contradict ourselves.

Last, why must we act on principles that we can will as universal laws? Kant answers: to act that way is to give laws to ourselves, and we must think of ourselves as laws to ourselves. That is autonomy. But if I give a law to myself, couldn't I release myself from that law, too? I conclude by using my interpretation to explain how laws that we give to ourselves can also impose genuine requirements on us.

The dissertation of William Leland Reckner is approved.

William Fred Bristow

C. T. Burge

Barbara Herman, Committee Chair

University of California, Los Angeles

2017

To all my teachers, especially Mom and Dad

Table of Contents

| | |
|---|-----|
| 1. Practical Principles and Laws of Freedom | 1 |
| 2. Freedom, Causation, and the Structure of Practical Principles..... | 38 |
| 3. The Structure of Practical Principles | 78 |
| 4. Alternative Views of Practical Principles | 117 |
| 5. Deriving the Formula of Universal Law..... | 155 |
| 6. An Empty Formalism? | 193 |
| 7. The Form of Law and the Idea of Freedom..... | 258 |
| Bibliography..... | 323 |

List of Abbreviations

Works by Kant

| | |
|----|---|
| CJ | <i>Critique of the Power of Judgment</i> |
| CP | <i>Critique of Practical Reason</i> |
| CR | <i>Critique of Pure Reason</i> , translated by Paul Guyer and Allen W. Wood |
| FI | First Introduction to the <i>Critique of the Power of Judgment</i> |
| G | <i>Groundwork of the Metaphysics of Morals</i> |
| JL | <i>Immanuel Kant's Logic</i> , edited by Gottlob Benjamin Jäsche |
| P | <i>Prolegomena to Any Future Metaphysics</i> |
| MF | <i>Metaphysical Foundations of Natural Science</i> |
| MM | <i>Metaphysics of Morals</i> |
| R | <i>Religion within the Boundaries of Mere Reason</i> |

Works by Hegel

| | |
|----|---|
| EL | <i>Encyclopedia of the Philosophical Sciences, Part I</i> |
| PR | <i>Elements of the Philosophy of Right</i> |
| PS | <i>Phenomenology of Spirit</i> |

Please see the Bibliography for full citations.

Acknowledgements

I have strived to include specific acknowledgments throughout the text of the study here, but I have many, many people to thank for their help and support, both with this dissertation and more generally. First, I would like to thank Barbara Herman, for modeling and helping to shape the approach and the temperament that I have taken to reading Kant, but mainly for always striking the perfect balance between trusting me to figure things out on my own and nudging me in the right direction. I have learned from her tremendously, and I could not have asked for a better advisor. I am also grateful to Tyler Burge, for many illuminating conversations and comments, and whose teaching and thinking about Kant has exemplified for me the analytical depth, rigor, and systematicity that I have sought to achieve here. Helping to teach his class on the *Critique of Pure Reason* has been invaluable to me. I began the work that led into this dissertation as an MA student under Bill Bristow at the University of Wisconsin, Milwaukee, so I would also like to thank Bill for helping me to start out on the path that has led to this work, and especially for his continued support and many valuable suggestions. I have been honored to have had him along throughout this project.

I have also benefitted greatly from conversations and comments from John Carriero and Sheldon Smith, without whom the scope and detail of this project would have been much impoverished. I am grateful to them for their help with Kant himself, but also with the broader historical and scientific contexts for the issues discussed here. I would also like to thank Margaret Atherton, whose Berkeley seminar will always stand out to me as a model of close textual analysis, and Julius Sensat, without whom Chapter Six would not have been possible at all. I owe much of my understanding of Kant's historical context to all four of them. I am grateful

as well to Pamela Hieronymi for running UCLA's Ethics Writing Seminar, which was of great help in developing Chapter Five in particular.

Last, I do not think that I can adequately express my gratitude and appreciation for all of my colleagues at UCLA and at the University of Wisconsin, Milwaukee, and for all of my students as well. They have been an endless source of inspiration and philosophical stimulation.

This dissertation was also supported by a Dissertation Year Fellowship from the Graduate Division of UCLA, and I am exceedingly grateful for that support.

I expect that I have forgotten some people, and I am certain that I have not thanked everyone fully. For that, I apologize.

Curriculum Vitae

Education

M.A. in philosophy, University of Wisconsin, Milwaukee 2007-2009

Thesis: Incentives and Goodness in Kant's Ethics
Chair: William Bristow
Committee: Julius Sensat, Blain Neufeld

B.A. in philosophy magna cum laude, Carleton College, Northfield, Minnesota 2001-2005

Thesis: Phenomenal Holism
Chair: Roy Elveton

Conferences and Colloquia

Presentations:

| | | |
|--------------------------------|---|---------------|
| A Problem of Racism for Kant | University of Texas, Austin | November 2016 |
| Kant's Negative Equality | Kingston University, London, United Kingdom | May 2016 |
| The Structure of Kant's Maxims | University of Colorado, Boulder | April 2016 |
| The Structure of Kant's Maxims | Louisiana State University, Baton Rouge | April 2016 |
| The Value of Reason | APA Central Div. Society for Student Phil. | February 2009 |
| The Value of Reason | Cornell College, Mount Vernon, Iowa | October 2008 |

Invited Comments and Panels:

| | |
|--|---------------|
| USC ZYGO Series on Ethics in Medicine: The Rights of Refugees | February 2016 |
| USC ZYGO Series on Ethics in Medicine: Medicine in the Media | February 2015 |
| UCLA/Antwerp Law and Philosophy Graduate Conference | October 2014 |
| USC/UCLA Graduate Philosophy Conference | February 2014 |
| UC Riverside Conference in Ancient Ethics and Moral Psychology | April 2013 |

Organization:

| | |
|---|-----------------------------|
| UCLA Law and Philosophy Graduate Conference | April 2016 |
| UCLA/Antwerp Law and Philosophy Graduate Conference | October 2014 |
| USC/UCLA Graduate Philosophy Conference | February 2013 |
| UCLA Albritton Society Graduate Conference | May 2011 |
| UCLA Albritton Society | September 2010 to June 2011 |

Academic Awards and Honors

| | |
|--|------------------------------|
| UCLA Graduate Division Dissertation Year Fellowship | Winter to Fall 2016 |
| Finalist, Charlotte W. Newcombe Doctoral Dissertation Fellowship | 2015 |
| UCLA Departmental Graduate Fellowship | Spring, Fall 2015, Fall 2014 |
| UCLA Graduate Student Summer Research Mentorship | Summer 2011 |
| UCLA Foundation Doctoral Scholar Fellowship | Fall 2010 to Spring 2011 |
| UCLA Departmental Graduate Fellowship | Fall 2009 to Spring 2010 |
| UWM Distinguished Graduate Student Fellowship | Fall 2008 to Spring 2009 |

Teaching Experience

Primary Instructor:

| | | | |
|-----------------------------|---|--|-------------|
| History of Greek Philosophy | UCLA | | Summer 2015 |
| Philosophy of Mind | Johns Hopkins Center for Talented Youth | | Summer 2013 |
| Philosophy of Mind | Johns Hopkins Center for Talented Youth | | Summer 2012 |
| Ethics | Johns Hopkins Center for Talented Youth | | Summer 2010 |
| Existentialism | Johns Hopkins Center for Talented Youth | | Summer 2010 |
| Existentialism | Johns Hopkins Center for Talented Youth | | Summer 2009 |

Teaching Assistant:

| | | | |
|--|---------|------------------|-------------------|
| History of Political Philosophy – Marx | UCLA | AJ Julius | Spring 2017 |
| History of Medieval Philosophy | UCLA | Calvin Normore | Winter 2017 |
| Kant's Critique of Pure Reason | UCLA | Tyler Burge | Winter 2015 |
| Intro. to Philosophy of Mind | UCLA | Gregory Antill | Summer 2014 |
| Aristotle | UCLA | Samuel Baker | Spring 2014 |
| Philosophy of Psychology | UCLA | Tyler Burge | Winter 2014 |
| Kant's Critique of Pure Reason | UCLA | Tyler Burge | Fall 2013 |
| Descartes' Meditations | UCLA | John Carriero | Spring 2013 |
| History of Modern Philosophy | UCLA | John Carriero | Spring 2013 |
| Space and Time | UCLA | Sheldon Smith | Winter 2013 |
| Kant's Critique of Pure Reason | UCLA | Tyler Burge | Fall 2012 |
| Medical Ethics | UCLA | Diane Kierce | Summer 2012 |
| History of Modern Philosophy | UCLA | John Carriero | Spring 2012 |
| Philosophy of Mind | UCLA | Tyler Burge | Winter 2012 |
| History of Greek Philosophy | UCLA | Keith McPartland | Fall 2011 |
| Intro. to Philosophy of Science | UCLA | Sheldon Smith | Spring 2011 |
| Intro. to Ethical Theory, Writing Int. | UCLA | Pamela Hieronymi | Winter 2011 |
| Intro. to Ethical Theory | UCLA | Barbara Herman | Fall 2010 |
| Intro. to Philosophy | UWM | Edward Hinchman | Spring 2008 |
| Intro. to Philosophy | UWM | Robert Schwartz | Fall 2007 |
| Existentialism (11 Courses) | JHU CTY | David Bollert | Summers 2004-2009 |

1. Practical Principles and Laws of Freedom

My overall project in this dissertation is to use Kant's concept of a practical principle to explain the fundamental concepts, principles, and arguments of his moral theory. My goal, in other words, is to give a thoroughgoing interpretation of the core structure of Kant's moral theory, using his concept of a practical principle as my point of departure and through-line.

Specifically, I want to explain these core elements of Kant's moral theory, along with the corresponding resources that his theory has for addressing some of its more persistent objections. As such, I will mostly not have space to comment on whether Kant's concepts are correct, his principles true, or his arguments sound. To explain his theory, we will ultimately have to take all of that for granted, though I will sometimes be able to indicate places where Kant's moral theory might be resisted. More regrettably, though, I not be able to say much about whether the arguments that I offer on Kant's behalf are ultimately persuasive. In the end, though, those questions must be left to the judgment of the reader, in any case.

If I am right, though, Kant's moral theory is best understood by focusing on his concept of a practical principle. The underlying motivation for this approach is twofold. To begin with, his concept of a practical principle is clearly essential to understanding the fundamental principle of his entire moral theory: the famous Formula of Universal Law, which requires us to "*act only in accordance with that maxim through which you can at the same time will that it become a universal law.*" (G 31, 4:421)¹ Kant says that the Formula of Universal Law is itself a practical principle, for one thing (G 5, 4:392; CP 28, 5:30-31). Primarily, though, the basic

¹ I will cite Kant's texts according to the abbreviations listed on page vi; please see the bibliography for the full citations. The page references in my parenthetical citations to Kant will refer first to the relevant English translations, then to the volume and page of the *Akademie* edition of Kant's complete writings. Citations to the *Akademie* edition of the *Critique of Pure Reason* will be to the A and B editions, however, rather than to the specific volume. For Kant's German, please refer directly to the *Akademie* edition, cited in the bibliography. Citing Kant's texts is usually only the start of the argument, however, rather than its conclusion. So the full arguments for my interpretations will often only come out over extended discussion. All emphases in quotations will be in the original, unless otherwise indicated.

concept invoked in the Formula of Universal Law is the concept of a maxim, and Kant is clear that maxims are a kind of practical principle as well.²

So I will argue, in later chapters, that we can better explain what maxims are and what it is to be able to will maxims as universal laws, if we first develop a deeper account of Kant's practical principles. In doing so, I will also argue that we can offer new solutions to some of the oldest and hoariest problems for Kant's moral theory. For example, it has long been maintained, against Kant, that any maxim whatsoever can be willed as a universal law. In later chapters, however, I will argue that my approach to Kant's moral theory can propose new answers to this venerable problem, and to others, as well.

However, the concept of a practical principle is also central to Kant's views on freedom and on rational agency, and this is the second motivation for my approach. In fact, my first task here will be to explain the fundamental connection that Kant draws between practical principles and freedom, since the rest of my account will build off of that connection.

This connection itself should not be much of a surprise: as we will see more fully over the course of this chapter, Kant thinks that there would be no point in the idea of a distinctively *practical* principle, if the laws of nature told the full story of human action and agency. So without space for freedom, the concept of a practical principle could have no application whatsoever, in his mind. That already suggests an important link between the notions, for him.

However, I mean something stronger in saying that there is a *fundamental* connection in Kant's thinking between practical principles and freedom. By that, I mean to say that Kant's concept of freedom is where a systematic presentation of his views on practical principles needs to begin. I emphatically do *not* mean to say that this approach is the only one that can illuminate Kant's thinking about freedom or about practical principles. Other approaches have proven quite powerful and instructive, and I have certainly benefitted from them tremendously. However, if I am correct, we cannot fully reach the bottom of Kant's thinking about practical

² See (CR 681, A812/B840; G 31n, 4:421n; G 14n, 4:401n; CP 17, 5:19; MM 17, 6:225; MM 17-18, 6:225).

principles, unless we start with his concept of freedom. And if we need to start with freedom to fully understand his views on practical principles, then we also need to begin there to fully understand the Formula of Universal Law, and the entire core of his moral theory, too.

So I am really after a rather more substantial thesis: a competing interpretation would hold that Kant's account of practical reason is fundamental to explaining his concept of a practical principle, instead. There would be a strong basis in Kant's texts and arguments for that interpretation (G 24, 4:412-413; CP 28, 5:30-31). Or, alternately, a third interpretation would hold that freedom and practical reason are equally fundamental for explaining Kant's concept of a practical principle. There would be strong evidence for that interpretation, too, because Kant thinks that the fundamental principle of freedom and the fundamental principle of pure practical reason are necessarily one and the same (G 52-53, 4:446-447; CP 26, 5:28-29).

Still, though, I want to say that Kant thinks that freedom is really fundamental here. For it is just as much Kant's view that reason, in its proper sense, must be thought of as free (G 54, 4:448). So rational beings and rational agents must also be thought of as free, he thinks, and it will follow that we cannot fully understand his views on practical reason, either, without understanding his views on freedom as well.

So I contend that freedom turns out to be fundamental, in Kant's mind. As the Preface to the *Critique of Practical Reason* would have it, "the concept of freedom, insofar as its reality is proved by an apodictic law of practical reason, constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason" (CP 3, 5:3-4). Freedom may need "an apodictic law of practical reason" to prove its reality.³ But if we want to understand Kant's views on "the whole structure of a system of pure reason" in itself, it would still be well to begin with what he takes to be the "*keystone*" of that system.

³ In the *Critique of Practical Reason*, Kant claims that consciousness of being subject to the distinctive kind of obligation involved in moral laws "may be called a fact of reason" (CP 28, 5:31), and he argues that this fact proves that we are free, and also subject to the moral law.

This, at least, is the guiding idea of the study here, and I want to begin by explaining it in more detail. My main idea is that freedom is, fundamentally, a kind of causality, according to Kant. As such, Kant thinks that there must also be laws of free causality—laws of freedom—parallel to how he thinks that there must be laws for causation in nature (G 52-53, 4:446-447).

Now, Kant thought that Newton had essentially completed physics. But not everyone understood Newton, or Newtonian physics, to be committed to causes.⁴ But gravity is a kind of natural cause, according to Kant, and according to the version of Newtonian physics that he thought was established. As such, he concluded that there has to be a law of gravity, as well. In an analogous way, Kant thinks that free causation requires laws of freedom.

However, there is an important difference between gravity, considered as a phenomenon in nature and of natural law, and the formula of Newtonian physics “ $F_g = G(m_1m_2/r^2)$ ”. Those mathematical symbols *represent* the quantity of force involved in gravitational attraction, but they emphatically do not *constitute* that force, and more than the name ‘Kant’ constitutes the philosopher from Königsberg himself. Analogously, my central thesis here is that, for Kant, practical principles represent the laws of freedom.

However, mathematical formulae can get the laws that they represent wrong, as Newton’s formula actually turned out to do. Likewise, practical principles can get the actual laws of freedom right or wrong, according to Kant. Representation carries the possibility of misrepresentation, at least in Kant’s terminology. As we will see more fully presently.

Translated into Kant’s technical vocabulary, then, my thesis is that practical principles are, fundamentally, the cognitions that represent the actual laws of free causality. As I said, I will explain that vocabulary in more detail presently. But if my basic interpretation is correct, then this thesis is where we need to begin in order to fully explain what practical principles are, for Kant. Not just that—if I am right, then this thesis is also where a fully systematic re-presentation of Kant’s moral theory ought to begin as well.

⁴ See note 20 in Chapter Two.

So the work of this chapter will be to say more about what Kant would mean in saying that practical principles represent laws of freedom, and to explain why we should think that practical principles are fundamentally representations of that kind, for him.

However, Kant's idea of a law of freedom is really rather puzzling, to say the least, and puzzling in two main ways, which it will be useful to introduce now. For these puzzles will also stay with us throughout the study here.

First, as we will soon see more fully, Kant thinks that freedom is the property of being an absolutely self-determining causality—the property of being the ultimate or absolute cause of your actions (G 52-53, 4:446-447). So if we are free in Kant's sense, then our actions can be influenced by our desires and inclinations, for example, but never wholly caused or determined by them. As long as we are free, Kant thinks that we are always the ultimate author or cause of our actions. So even if we act according to a desire or an inclination, he thinks it is because we freely chose to do so, not because that desire or inclination caused or determined our choice. Not by itself, anyway.

However, this idea of freedom makes the idea of a “law” of freedom quite puzzling, particularly if we do not yet fully understand the analogy that Kant finds between laws of nature and laws of freedom. If I am a free agent in Kant's sense, then I am the ultimate cause of my actions. As such, I determine my actions myself, absolutely. But if there are also laws of free causation, then shouldn't those laws determine my actions, too, analogously to how Newton's law of gravity is supposed to determine the quantity of attractive force between material bodies? How, then, would my actions be *absolutely* self-determining, if they must also be determined by the laws of freedom?

The answer, Kant thinks, is that free agents must give laws to themselves—they must be autonomous, in a certain, fairly technical sense. If I give the law to myself, Kant thinks, then my actions can be determined by the law, while still counting as absolutely self-determined.

But that does not necessarily resolve the puzzle: if I am absolutely self-determining, and I give a law to myself, can't I just as well release myself from that law, too, at least by giving myself a different law?⁵ How, then, is any "law" really a law for me? Material bodies have no such discretion over whether they obey the law of gravity, by contrast.

So the idea of an absolutely self-given law sounds as if it might be incoherent, because if I give the law to myself, then it sounds as if the law does not really determine my actions. I do. But if the law really does determine my actions, then it sounds as if it must be imposed on me, and as if I do not give it to myself. How, then, can the ultimate cause of my actions be both myself and the law?

To put this first puzzle another way, Kant wants to use the idea of laws of freedom to explain how free actions and free choices do not have to be random or arbitrary. The laws of freedom are supposed to explain how free actions and free choices can still follow from something, how they can still be *grounded*, to use the technical vocabulary that we will develop here. But the risk here is that these notions are just incompatible: if a free choice must follow from something, if it is not to be arbitrary, then isn't that thing the ultimate cause of the choice, and not the agent? If I drink because I am thirsty, isn't my thirst the ultimate cause of the action, and not myself? So there might be no way to explain how an action or a choice can be both free, in the sense of absolutely self-determined, and lawful, in the sense of determined by or following from a law. For lawful or grounded freedom might turn out to be a contradiction in terms.

For the second puzzle: Kant thinks that the law of gravity determines how material bodies always behave, and that this law precludes any possible exceptions. And, in general, Kant insists that "the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited" (G 52, 4:446). So the idea of a "law" of freedom might suggest that these laws are to determine how free agents

⁵ The puzzle is a version of the famous "Paradox of Voluntarism".

always will act, or perhaps even how free agents cannot help but act. Or at least it might sound as if agents are only really free when they obey these laws, and unfree when they do not.

But isn't that a puzzle? Shouldn't free agents be free, no matter what they choose to do? Especially if freedom includes the capacity for absolute self-determination?

Here, Kant wants to distinguish perfect and imperfect freedom. A perfectly free agent will always choose to follow the laws of freedom, and actually cannot do otherwise, because of the nature of their agency. By contrast, imperfectly free agents, such as us paltry humans, are constitutionally and endemically prone to violating the laws of freedom. If we are free in Kant's sense, then we are always capable of following the laws of freedom, to be sure. For we would always retain at least the capacity for absolute self-determination. (Barring extreme circumstances, anyway, such as severe illness or intoxication.) Thus, Kant thinks that we are still free, even when we deviate from the laws of freedom. But our human nature is such that we are always susceptible to the influence of our desires and inclinations, Kant thinks, and in a way that makes us prone to deviating from the laws of freedom. Prone to misusing our freedom, as he would put it, even if we do not relinquish it.

So for Kant, the nature of a perfectly free agent guarantees that they *will necessarily* obey the laws of freedom. However, he thinks that those same laws of freedom, by their very essence or nature, instead constitute norms for how imperfectly free agents *ought always* to act, even if such agents never actually did act that way, and even if they never actually would act that way, either. So, for example, if a principle such as "thou shalt not lie" (G 2-3, 4:389) represents a genuine law of freedom, then this principle will describe how perfectly free agents will never lie, and it will likewise state that imperfectly free agents ought never to lie, on pain of misusing their freedom.⁶

⁶ As we will see shortly, Kant sometimes describes laws of freedom, categorically, as laws of what "ought" to happen, which makes it sound as if laws of freedom are essentially normative for all free agents. However, Kant always speaks this way in contexts where it is clear that he is talking about laws of freedom as laws of what ought to happen *in nature*. For one thing, in Kant's vocabulary, what "happens" is always

Kant's view on the exact nature or character of this mistake still needs to be settled, of course. But a misuse of freedom sounds a lot like a moral mistake or wrong, so we can already start to see how Kant will end up thinking that laws of freedom and moral laws turn out to be one and the same. At least in part—we will also see that not every law of freedom is a moral law, in Kant's view, nor does every law of freedom apply to every free agent.⁷

But herein lies the second puzzle: why should human beings think that they are subject to the laws of freedom, if these laws are to be norms for us? And this question really boils down to asking: why should human beings think that they are free, in Kant's sense?

So one puzzle wants to understand how Kant's concept of a law of freedom is even coherent, and the other wants to know why human beings should be subject to such laws. We will not be able to really articulate Kant's answers to these puzzles until the last chapter here. On the one hand, though, Kant's problem is to explain how following a certain kind of law will also count as being free—as being the ultimate or absolute cause of your actions. And, on the other hand, his problem is to explain why we have to think of ourselves as subject to laws of that kind.

We will have found Kant's laws of freedom when we have identified the laws that match this description. And I do not think it will spoil too much to indicate that the right kind of law is

something happening in time, and hence in nature (CR 484, A444/B471). And nature, for Kant, is only a system of appearances, not a system of things in themselves. So when Kant says that laws of freedom are laws of what "ought to happen", his claim is implicitly restricted, which leaves space for laws of freedom to be laws of what perfectly free agents will always, necessarily do. If we do not read Kant this way in these contexts, then such remarks will conflict with his repeated insistence that perfectly rational agents—who are also perfectly free—are not subject to any "oughts" (G 25, 4:414). Nevertheless, perfectly free agents might still retain a certain capacity to violate the laws of freedom, even though their perfect rationality makes it impossible for them to actually do so. Freedom, for Kant, might entail a capacity to violate its laws, even though a perfectly rational being can never utilize this capacity, due to how they will always find it unintelligible or incoherent to do so. So perfectly rational agents, as free, might be able to violate laws of freedom, even though they ultimately cannot do so, as perfectly rational, which in turn means that laws of freedom cannot be laws for how they ought to act. What a being is capable of doing, considered in one relation, they may be incapable of doing, when considered in their full nature.

⁷ We will see that a principle such as "in order to divide a line into two equal parts on a sure principle I must make two intersecting arcs from its ends" (G 28, 4:417) is to be a law of freedom, but not a moral law. Also, moral laws pertaining to sex and drink are to be specific to human nature (MM 178-181, 6:424-428), so not every law of freedom will apply to all free agents. And since freedom is to be absolute self-determination, Kant can even recognize laws of freedom that are specific to a single free agent.

to derive from the Formula of Universal Law, and from its demand to act only on maxims that we can will as universal laws.

In any case, though, given the nature of these problems, Kant's answer must lie in his concept of freedom, in his concept of laws of freedom, and in the nature of the principles that are to represent those laws. In Kant's concept of a practical principle, in short.

So, to begin making my case, we will begin by examining why we should think that Kant's practical principles are, fundamentally, representations of laws of freedom. I argue that this point is evident in Kant's treatment of laws of freedom in the Architectonic of Pure Reason from the *Critique of Pure Reason*, and also in the Preface to the *Groundwork of the Metaphysics of Morals*. So the main work of this chapter will be to explore these texts, in order to draw out how they connect practical principles to the laws of freedom. These texts, I claim, each reveal what is fundamental about practical principles, for Kant, because he uses them to give his outlines of the structure of the system human cognition. In this structure, practical principles end up representing laws of freedom, and this structure could not constitute a *system*, for Kant, unless this structure also reveals what is fundamental about practical principles. Or so I will argue in this chapter, taking these texts in order of their publication.

First, however, we should examine Kant's views on principles in general. This will require some examination of Kant's usage, when he speaks of "principles". We will also need to look at Kant's views on cognition, since my thesis contends that practical principles are a kind of cognition, in his sense. We also need to begin with Kant's views on cognition in order to raise a problem for my thesis. We can then approach the Architectonic of Pure Reason and the Preface to the *Groundwork*, and we can conclude by looking at two of Kant's more important specific explanations of practical principles.

1.1 Principles in General

To start with, Kant's use of 'principle' is rather complicated, and he does not appear to adhere very strictly to the distinction between principles and laws that I want to draw here.

Primarily, Kant often calls certain kinds of principles “laws”, and he even explicitly defines practical laws as a certain kind of principle (G 14n, 4:401n; CP 17, 5:19). My interpretation can explain how this usage is warranted, though.

If principles represent laws, as I contend, and if principles can represent laws accurately, then there will be a kind of isomorphism or analogy between laws and principles. For the content of the principle will then purport to match the nature of its law, and the nature of laws can be reflected in the content of their principles.⁸ This isomorphism is why the distinction between principles and laws can be easy to miss, though I am arguing that this distinction is really crucial. And this isomorphism is also why it makes sense to call certain kinds of principles “laws”, as when we call the formula “ $F_g = G(m_1m_2/r^2)$ ” “Newton’s Law of Gravity”. Moreover, when a principle succeeds in representing a genuine law, it makes all the more sense to call that principle a “law” as well.

However, in the interests of clarity, I will enforce this distinction between principles and laws, even if Kant does not. Next, Kant has several words that sometimes get translated with the English ‘principle’. Mainly, there is the Latinate ‘*Princip*’/‘*Prinzip*’, but also sometimes the related Latin ‘*Principium*’. Kant’s ‘*Grundsatz*’ is sometimes translated as ‘principle’ as well. See (CR 387-389, A298-302/B355-359) for examples, as well as the *Architectonic of Pure Reason*, and the First Section of the *Groundwork of the Metaphysics of Morals*. The *Metaphysical Foundations of Natural Science* also uses both ‘*Grundsatz*’ and ‘*Lehrsatz*’ in giving most of its main conclusions. Friedman translates the former as ‘principle’ and the latter as ‘proposition’.

I do not know whether there is any underlying reason for the different terminology, though ‘*Grundsatz*’ sometimes seems to connote fundamentality. For example, the *Grundsatz* in

⁸ This is not necessarily how Kant uses ‘content’ when he speaks of the “content” of a representation, at least not always. He sometimes seems to say that the content of a representation is its relation to its object (CR 340-341, A239/B298). If he does think of the content of representations that way, then his “content” more naturally resembles what we would call the “reference” of the representation. ‘Reference’ may not perfectly capture what Kant means, either, but in the absence of a full account of Kant’s views on representation, I think it will be less confusing if I speak in the terms that are more familiar now.

the *Metaphysical Foundations* appears to be structurally more important than the *Lehrsätze*. And see also (CR 388, A301/B357) and (G 8, 4:394), where Gregor explicitly—and astutely—translates ‘*ohne Grundsätze eines guten Willens*’ as ‘without the basic principles of a good will’. Even the etymology of ‘*Grundsatz*’ might suggest “basic principle” or “fundamental principle”. ‘*Grund*’ means ground; ‘*Satz*’ by itself is more commonly rendered as ‘proposition’, but as such also shows up in contexts where ‘principle’ is a plausible rendering, even in light of Kant’s technical usage, which we are about to examine. See (G 28, 4:417), for example. We are about to look at a passage where Kant treats principles as a specific kind of proposition, too.

But because of how complicated Kant’s usage is, it will be better for us to focus on what Kant tells us about principles directly, and on how he uses the term in passages and arguments that are important to the structure of his thought.

In that spirit, I have said that practical principles are a kind of representation, for Kant, which means that Kant would consider them to be a kind of cognition, in his specific technical sense of the term. So we should start with a brief discussion of Kant’s thinking on cognition (*Erkenntnis*). *Erkenntnis* is perhaps—even probably—the single most important technical concept in Kant’s entire critical philosophy, on both the theoretical and practical sides. So what I can say here must be only partial and piecemeal, though I often suspect that a fuller treatment would go a long way to resolving much of the confusion about Kant’s overall corpus.

In any case, the primary text for Kant’s views on cognition is the famous *Stufenleiter* passage from the *Critique of Pure Reason* (CR 398-399, A319-320/B375-377). There, Kant explains that cognitions are a specific kind of representation (*Vorstellung*). Roughly speaking, a representation is anything “inside” or “before” the mind. Kant’s word ‘*Vorstellung*’ itself suggests that meaning—literally, something “placed before” or “standing before”. So, in this context, something “standing before” or “placed before” the mind.⁹

⁹ It is now standard to translate ‘*Vorstellung*’ with ‘representation’, though some translations, such as Pluhar’s (1996), prefer ‘presentation’. ‘Presentation’ is a rather awkward word here, but it has some

Our picture here must remain rather vague and metaphorical, sadly, but I think the metaphor suffices to convey the general contours of Kant's theory of cognition. For cognitions are conscious representations, as against unconscious representations, and they are conscious representations of an object, in contrast to conscious representations that are subjective, such as Kant holds smells, tastes, pains, or tickles to be.¹⁰

As the *Stufenleiter* passage has it, a subjective representation "refers to the subject as a modification of its state" (CR 398, A320/B376), and Kant calls these representations "sensations". So Kant thinks that the smell of a rose or the pain of a headache only serves to indicate, reflect, or relate to some change within me or within my mind, as the subject of the sensation. Kant allows that these sensations may be associated with or caused by an object outside of me, such as the rose itself or the incessant noise outside my window. And they will presumably correspond to or correlate with a change in my body as well. But he insists that these sensations do not *represent* those objects or those changes—neither sensation "places" the rose itself or the noise itself "before" my mind, he thinks, except perhaps through the associations that I have formed through prior experience. And those sensations do not represent the changes within me to which they "refer", either, and for similar reasons.

In any case, Kant holds that the fundamental difference between a sensation and a cognition is that a cognition *does* "place" an object "before" the mind, and a cognition does that itself, not through a built-up association. As a result, a cognition can represent its object accurately or inaccurately. My perception of a rose, when I see it, is both related to the rose itself, and my perception "presents" the rose to me, Kant thinks. As a result, I can misperceive the rose, too: my perception can get the rose wrong. The rose can appear to me, contrary to how it actually is.

philosophical and interpretive merits over 'representation'. 'Representation' tends to suggest a representation of an object—Kant's "cognitions", rather than his "*Vorstellungen*".

¹⁰ Kant gives no indication of what the unconscious *Vorstellungen* are like, though nothing Kant says really prevents them from having the same classifications and divisions, at least so far as I know.

In contrast, subjective representations can be causally abnormal, such as when we feel pain in a healthy part of our body or when something tastes or smells “wrong”. But Kant thinks that this is quite different from a misperception or a miscategorization: there is no object that gets misrepresented by these sensations, only an unusual or misleading conjunction of sensation and stimulus. We can certainly be prone to further cognitive errors in such cases—we might misdiagnose an injury, or we might misidentify our food. Only the sensation itself is still not misrepresenting anything, in Kant’s view.

Moreover, it is important to note that ‘object’ must be construed as broadly as possible here: numbers, geometrical shapes, subjective representations, God, fictional characters, logical forms, normative properties, other cognitions—anything, really, can be the object of Kant’s cognitions. Not the object of just any kind of Kant’s cognitions, mind—we cannot perceive God or numbers, for example, because those objects are not sensible. And Kant thinks that we can never know whether we can represent certain objects *accurately*. (The famous *noumena*, for example.) Moreover, Kant is primarily interested in a very particular kind of accuracy: the fundamental questions of his critical philosophy have to do with how we can show that certain cognitions are *objectively valid*, meaning that their object is both a “real possibility”, and can be proved to be such (CR 115n, Bxxviii).

Now, the crucial point, as least for my thesis about practical principles: the *Stufenleiter* passage holds that cognitions come in two basic kinds: intuitions and concepts. “The former is immediately related to the object and is singular; the latter is mediate, by means of a mark, which can be common to several things.” (CR 399, A320/B377) My perception of the rose is to be an intuition, since it is to represent only a single rose, and since it is to represent that rose directly or immediately, through my senses, rather than “by means of a mark”. My representation of what roses have in common, however, is to be a concept.

So notice that the *Stufenleiter* passage’s taxonomy of kinds of cognitions does not include judgments, propositions, or, more to the current point, principles. At least not explicitly,

either as a separate category of cognition, or as a kind of intuition or concept. And Kant's published lectures on logic—called the “Jäsche” *Logic*, after its editor—gives the same account of cognition, too.

There, Kant says that “All cognitions, that is, all representations related with consciousness to an object, are either *intuitions* or *concepts*. An intuition is a *singular* representation (*repraesentatio singularis*), a concept a universal (*repraesentatio per notas communes*) or *reflected* representation (*repraesentatio discursiva*).” (JL 589, 9:91) In saying that concepts are “universal”, Kant explains that he only means to say that a concept is “a representation of what is common to several objects” (JL 589, 9:91), just as the *Stufenleiter* passage, already quoted, also holds.

What's more, Kant's argumentation sometimes relies on the idea that cognition only comes in these two kinds. For example, the First Critique's Transcendental Analytic starts by trying to arrive at a characterization of our capacity or faculty for understanding. Its reasoning begins thusly:

The understanding has been explained above only negatively, as a non-sensible faculty of cognition. Now we cannot partake of intuition independently of sensibility. The understanding is therefore not a faculty of intuition. But besides intuition there is no other kind of cognition than through concepts. Thus the cognition of every, at least human, understanding is a cognition through concepts, not intuitive but discursive. (CR 204-205, A67-68/B92-92)

Since cognitions are just representations of objects, “cognition through concepts” just means “representations of objects through concepts” or “using concepts”. In short, then, the understanding is a faculty that uses concepts to represent objects; that is supposed to follow from how the understanding is “a non-sensible faculty of cognition”, how all of our intuitions are sensible, and how intuitions and concepts are the only kinds of cognition available to us.

So Kant appears to be committed to dividing cognition into only intuitions and concepts, which consequently appears to leave out judgments, propositions, and principles. And, in fact, if Kant leaves out judgments alone, then it would appear that he must also leave out propositions

and principles, too—the Jäsche *Logic*, at least, explains that propositions are assertoric judgments (JL 605, 9:109), and that principles are “Immediately certain judgments *a priori*” (JL 606, 9:110). Which would make principles a kind of proposition, too. Moreover, we are presently going to examine some passages from the First Critique’s Transcendental Dialectic, where Kant also treats principles as a kind of proposition, though not always in a sense as narrow as the one he uses in the Jäsche *Logic*. So if *either* judgments or propositions are not cognitions, for Kant, it is hard to see how principles could be, either.

As mentioned, these omissions obviously look bad for my main thesis here, since I want to say that practical principles are a kind of cognition, for Kant. Yet it is rather odd for Kant to leave judgments in particular out of his accounts of cognition, since he does sometimes treat judgments as cognitions, too. And if judgments count as cognitions, for Kant, there will be little reason to think that propositions and principles should not be cognitions for him, also. As the passages just cited from the Jäsche *Logic* indicate.

Judgments, then, will be a good test case for the viability of my main thesis about practical principles. And for an example of Kant treating judgments as cognitions, let’s return to the paragraph from the Transcendental Analytic just cited. That paragraph, on the understanding, goes on to conclude that “Judgment is therefore the mediate cognition of an object, hence the representation of a representation of it.” (CR 205, A68/B93) The Jäsche *Logic*, too, provides a formal, explicit “*Definition of a judgment in general*”, which holds that “A judgment is the representation of the unity of the consciousness of various representations, or the representation of their relation insofar as they constitute a concept.” (JL 597, 9:101) The unity of the consciousness of various representations is a rather unfamiliar sort of object, to be sure, as is the relation between various representations insofar as these representations constitute a single concept. But if a judgment is to represent one of these objects, then it will count as a cognition nevertheless, by the lights of the prior definitions in both the *Stufenleiter* passage and earlier in the Jäsche *Logic*.

Moreover, Kant uses declarative sentences to express judgments in language: this same paragraph from the Transcendental Analytic that we have been citing gives “**All bodies are divisible**” (CR 205, A68/B93) as an example of a judgment. The judgment thus expressed certainly appears to have an object, since it is a judgment about bodies. This judgment also holds that all bodies are divisible, so it certainly appears to represent bodies as being a certain way, too. And this judgment can get its object right or wrong as well, since the judgment will be true if all bodies really are divisible, and false otherwise. So the judgment underlying this linguistic expression also appears to be a representation of an object—a cognition, in Kant’s sense of the term. We could say the same for the proposition that all bodies are divisible, and for the principle that all bodies are divisible, which we might apply as such in natural science.

Last, and perhaps most importantly, the *Critique of the Power of Judgment* clearly treats the titular power of judgment as a “faculty of cognition” (CJ 56, 5:168; 64, 5:177). There, Kant explains that “The power of judgment in general is the faculty for thinking of the particular as contained under the universal” (CJ 66, 5:179). The First Introduction to this Third Critique also says that “**the power of judgment**” is “the faculty for the **subsumption of the particular** under the general” (FI 8, 20:201). Both descriptions clearly treat judgment as a cognitive faculty, according to the definitions Kant stipulates. The *Critique of Pure Reason* treats judgment similarly, when it says that “the power of judgment is the faculty of **subsuming** under rules, i.e. of determining whether something stands under a given rule (*casus datae legis*) or not.” (CR 268, A132/B171)¹¹

So what are we to make of this puzzle? Why would Kant insist that intuitions and concepts are the only kinds of cognitions, and then go on to treat judgments as a kind of cognition, and the power of judgment as a cognitive power?

My own opinion is that Kant’s views on thought contain the answers. The paragraph from the Transcendental Analytic that we have been citing also explains that “Thinking is

¹¹ Guyer and Wood translate ‘*casus datae legis*’ as “case of the given law” (CR 268n).

cognition through concepts.” (CR 205, A69/B94). The Jäsche *Logic* agrees, since the passage cited above, where Kant distinguishes intuitions and concepts in the first place, concludes by saying that “Cognition through concepts is called *thought (cognitio discursiva)*.” (JL 589, 9:91) Judgments, propositions, and principles all fall under this heading, for Kant, as should already be clear. After all, they are all supposed to represent objects through the use of concepts.

So I contend, boldly, perhaps, that judgments, propositions, and principles are themselves concepts, at least in Kant’s mind. I do not have space to fully argue for this interpretation here, however, though it is certainly important to the current project.¹² In lieu of that argument, however, it should be noted that we would not damage Kant’s overall thinking if

¹² Briefly, though: since judgment is a faculty, Kant distinguishes between judgment as the act of that faculty (G 54, 4:448; CJ 56-57, 5:168-169) and judgment as the product of that act (described in the main text above). So here is how I understand Kant’s picture: in general, the act of judgment takes a concept and applies it to an object, through or by means of another cognition that I have of that object. That comes out of Kant’s general descriptions of the power of judgment, cited earlier in the main text. The judgment produced by this act is then a new cognition: a cognition representing that same object, but now through both of these cognitions, unified into the produced judgment. I perceive an apple and judge that it is red: this act of judgment applies my concept of redness to the apple, on the basis of how I perceive it to be, resulting in a new cognition of that same apple as red. This new cognition is the judgment produced by the act of judging, and I contend that it is itself a concept, by Kant’s definitions. For this judgment represents its object—the apple—only through the mediation of other cognitions: the perception of the apple and the concept of redness, which are combined through the act of judgment. (As the Jäsche *Logic* had it, “the unity of the consciousness of various representations, or the representation of their relation insofar as they constitute a concept.” (JL 597, 9:101)) Moreover, the judgment represents the apple through a mark as well, which could be common to multiple objects. For there must be something in my perception of the apple that allows me to apply my concept of redness to it—something linking the content of my perception and the content of my concept. This link is the mark, and multiple objects could exhibit it. So if a concept is to be a representation of an object that is “mediate, by means of a mark, which can be common to several things” (CR 399, A320/B377), as the *Stufenleiter* passage has it, for example, then a judgment will be a concept, too. There are three problems with this interpretation, though: first, it has to allow concepts to have a propositional structure. We might dislike this philosophically, but, so far as I can see, nothing in Kant’s definition of concepts prevents it. Second, Kant sometimes describes concepts “as predicates of possible judgments” (CR 205, A69/B94). So aren’t I saying that judgments can be predicates of possible judgments, for Kant? Yes, but this does not have to mean anything except that judgments, considered as products, are applied to objects, in or as the result of an act of judgment. Third, I pointed out how the Jäsche *Logic* says that concepts are universal. But Kant also says that judgments can be universal, singular, or particular in form (JL 598, 9:102). However, I also explained that the universality of concepts consists only in how concepts represent something common to several objects. And Kant also says that “It is a mere tautology to speak of universal or common concepts—a mistake that is grounded in an incorrect division of concepts into *universal*, *particular*, and *singular*. Concepts themselves cannot be so divided, but only *their use*.” (JL 589, 9:91) Judgments, though, can be universal, singular, or particular by virtue of whether they apply a concept universally, singularly, or particularly. So there is no problem here, either.

we held that concepts and cognitions through concepts—a.k.a. thoughts—must constitute different types of cognitions.

The propositional form of thought might ultimately require Kant to distinguish concepts from thoughts, in any case. But he can perfectly well accept that distinction. Even the argument quoted above, where Kant relies on dividing cognitions into only intuitions and concepts, would still go through under this three-fold distinction, given how thought is tied back to and essentially relies on concepts.

We could also conclude that Kant's remarks on cognition are simply inconsistent. In that case, my thesis that principles are cognitions would be licensed by one strand of his usage, and my project here can justify focusing on that strand by illustrating its utility in cleaning up and explaining other parts of Kant's thought.

Last, though, suppose that I am wrong about Kant, when I argue that his principles should be understood as a kind of representations. I mentioned above, though, that there is to be a certain analogy or isomorphism between representations and their objects, especially when the former are accurate to the latter. That isomorphism means that the general contours of my account can be evaluated at either of these levels—at either the level of the kind of representation, or at the level of the kind of object represented. So if Kant's principles are not really representations at all, then the spirit of my account can still be maintained, *mutatis mutandis*, by focusing on the “object level” of my account. And the isomorphism between representations and their objects will allow that maneuver to be made quite straightforwardly.

So, whichever way we go, Kant's general views about cognitions and judgments leave room for my primary thesis, at the very least, when they do not positively support my thesis. With that in mind, though, let's look at Kant's concept of a principle more directly.

The primary text for Kant's views on principles is the Introduction to the Transcendental Dialectic in the *Critique of Pure Reason*. There, Kant wants to explain the faculty of “reason in general”: as Kant puts it, his task there is “to give a definition of this supreme faculty of

cognition” (CR 387, A298-299/B355). And his definition holds that reason is “the **faculty of principles**” (*Principien*) (CR 387, A299/B356).

However, it is not immediately clear from this definition whether reason is the faculty for cognizing principles or for cognizing objects from or according to principles. (Both, I take it.) Before introducing this definition, though, Kant states that “reason itself contains the origin of certain concepts and principles, which it derives neither from the senses nor from the understanding.” (CR 387, A299/B355) Reason is a cognitive faculty, though, so these “certain” principles, which Kant thinks originate in reason, must be cognitions, too. The parity with concepts here suggests that reading, also. And then we should expect all principles to be cognitions, for Kant: it would be odd for some principles to be cognitions and some not to be, particularly when a cognitive faculty is to be the faculty of principles.

Moreover, though, in the course of clarifying what he means to say in calling reason the “faculty of principles”, Kant goes on to distinguish three progressively more exclusive senses of ‘principle’.¹³ He explains that a cognition *can be used* as a principle, can be a principle *relative to other cognitions*, or can be a principle *absolutely* or *in itself*. Reason is to be the faculty of principles in this last, absolute sense.

In any case, though, this taxonomy of principles already shows that principles are to be cognitions. According to this taxonomy, either the cognition is itself a principle, either relatively or absolutely, or the cognition is just to be used as a principle. For a cognition to be used as a principle, though, the principle must serve to represent objects. Cognitions are representations of objects, so representing objects is all that they can be used for. To use a cognition as a principle, then, that principle must itself be a cognition, according to Kant’s definitions.

Moreover, there is a clear progression in these three senses of ‘principle’, because the former senses are to apply to the latter principles. (Principles in the absolute sense are also

¹³ The subsequent discussion here follows (CR 387-389, A298-302/B355-359).

principles relative to other cognitions, and they can be used as principles, too.) But the main point for us here is that all principles must be universal, Kant thinks, albeit in different ways.

According to Kant, to use a cognition as a principle is just to use it as the major premise in a deductive syllogism. And that use only requires the principle to be of the form “All x’s are y’s”. So Kant says that even empirical generalizations can be used as principles, at least when we extend the generalization to universal application.

In contrast, Kant holds that cognitions that are principles either absolutely or relatively must be synthetic, and justifiable only *a priori*. As *a priori*, Kant thinks that both of these kinds of principles must apply to their objects universally, too, but now in the specific sense of “universal, without even the possibility of exception” (CR 137, B4). Principles in the absolute sense, however, have to be justified by concepts alone, while principles in the relative sense can be justified at least in part through either pure intuition or through the conditions that make experience possible.¹⁴ For example, Kant thinks that the axioms of mathematics are only principles relatively speaking, since his view is that mathematics relies on pure intuition, and not solely on concepts.

The further technical details of these different conceptions of principles don’t matter now—it doesn’t matter what else Kant means by “synthetic” or “*a priori*”, for example. It is worth noting that these concepts apply *in stricto sensu* only to cognitions: strictly speaking, only cognitions can be synthetic or *a priori*, according to Kant, since these are semantic and epistemic concepts, respectively (CR 141-143, A6-10/B10-14; 137-138, B3-6). But the upshot is the same, even if we ignore these technical complications: principles are a kind of cognition, and they represent universal regularities in the objects that they represent, albeit different kinds of universal regularities.

¹⁴ Hence, according to Kant there is no such thing as an absolute principle for theoretical cognition, which is why the faculty of reason gets treated in a “Transcendental Dialectic” rather than as a part of the Transcendental Analytic. Kant does not explain where either analytic principles or the principles of general logic are to fit into this schema. He also does not say whether any of these kinds of principles have to be objectively valid; having a certain cognitive form or purport might be enough to count as a principle.

Even if all principles are universal, for Kant, he may or may not think that it is appropriate to describe all principles as representing “laws”: the axioms and theorems of mathematics, while surely universal, may not qualify as “laws”, for him. The principles of general logic might, however, insofar as they represent the “laws” of thinking in general (CR 194, A52/B76). When a principle represents a universal *causal* regularity, though, as will be the case with the principles of natural science and the principles for our actions, then Kant will find no barrier to calling principles representations of *laws*. So Kant’s first general treatment of principles ends up confirming the account of practical principles that I will be giving here, at least to the extent that practical principles will represent laws, if not also laws of freedom.

1.2 The Architectonic of Pure Reason

To make a more decisive step from practical principles as representing laws to practical principles as representing laws of freedom, though, let’s now discuss the first *Critique*’s Architectonic of Pure Reason. So what does this Architectonic tell us about practical principles?

To begin with, the Architectonic of Pure Reason is, in a very fundamental way, an extension of the discussion of reason and principles from the Transcendental Dialectic. For Kant’s purpose in the Architectonic is to give an outline, at least partial, of the system of rational cognition, which he calls the system of “*cognitio ex principiis*” (CR 693, A836/B864)—cognition “from principles”. Which sense of ‘principle’ he has in mind here is not clear. It cannot be the absolute sense, however, because mathematical cognition counts as cognition “from principles” in the Architectonic (CR 693, A837/B865), as does “rational cognition from empirical principles” (CR 695, A840/B868). Neither of which would count as cognition from an absolute principle, in the terms that we just described from the Transcendental Dialectic.

In any case, though, we are only interested in what the Architectonic says about rational cognition of freedom. The crucial paragraph is very brief and relatively straightforward. As usual, though, it raises more questions than it answers definitively. Here it is, in its entirety:

Now the legislation of human reason (philosophy) has two objects, nature and freedom, and thus contains the natural law as well as the moral law, initially in two separate systems but ultimately in a single philosophical system. The philosophy of nature pertains to everything that **is**; that of morals only to that which **should be**. (CR 695, A840/B868)

For now, the full details about what Kant means by “philosophy” or “the legislation of human reason” don’t matter. It’s enough to point out that philosophy is to be a type of rational cognition, that is, a type of cognition from principles (CR 693, A835-837/B863-865).¹⁵ So the “philosophy” with freedom as its “object” will be cognition of freedom from principles.

Now, as I mentioned already in the last section, “cognition from principles” might be a little ambiguous: “from” suggests cognition originating in or deriving from principles, to be sure, in which case it would look as if those principles should be cognitions also. But for all ‘from’ means, the principles for our cognition here could be the objects that we represent, and not the representations that we form of those objects. Is “ $F_g = G(m_1 m_2 / r^2)$ ” to be the principle for our cognition of gravity, or is the principle to be the regularity that this formula purports to describe?

However, this paragraph from the Architectonic clearly treats its principles as cognitions. For Kant says that philosophy, in its role as the legislation of human reason, “has two objects”: nature and freedom. But if philosophy has objects, then it must be a kind of cognition, according to Kant’s terms. And philosophy must then cognize its objects “from principles”, if it is also a kind of rational cognition, in the Architectonic’s terms. These principles for philosophical cognition will then be used to cognize nature and freedom, and so the principles of philosophy must be cognitions themselves.

However, Kant does not call the principles for cognizing freedom and its laws “practical” principles, at least not in the Architectonic itself. But this paragraph does say that philosophy “contains the natural law as well as the moral law”, so we can infer that there is to be at least one

¹⁵ It is worth mentioning, though, that “the legislation of human reason” means “the science of the relation of all cognition to the essential ends of human reason (*teleologia rationis humanae*)” (CR 694-695, A839/B867), rather than what one might normally expect Kant to mean by lawgiving human reason.

law of freedom—the moral law—which is also contrasted to the laws of nature. In fact, this paragraph contrasts freedom in general with nature in general. So freedom will have to be a domain of agency separate from the laws of nature—at least in some sense of ‘separate’, anyway. Kant’s views on this score are notoriously complicated and obscure.

In any case, though, if the principles for cognizing freedom and its laws are to be separate from the principles concerned with cognizing laws of nature, then it would certainly make sense to call the principles for cognizing laws of freedom “practical” principles. And this is indeed Kant’s practice: in the Canon of Pure Reason, the chapter immediately preceding the Architectonic, he says that “everything that is connected with” free choice, “whether as ground or consequence, is called **practical**.” (CR 675, A802/B830)¹⁶

So it follows almost immediately from the text of the Architectonic that practical principles are to cognize or represent freedom and its laws. And since these principles are explicitly for cognizing freedom, the proper objects of these principles will be precisely these laws of freedom. The analogy with laws of nature helps to show why.

When we are doing natural philosophy in the Architectonic’s sense, we seek to cognize nature “from principle”. In practice though, Kant thinks that we seek to cognize nature and natural phenomenon through the laws of nature, as the role of these laws in physical explanation is supposed to make evident (G 63, 4:459). We seek to explain, and therefore to cognize, gravitational interactions through the laws of gravity, for example. However, we need to represent those laws in order to use them to cognize anything in nature. So in order to use the law of gravity to cognize gravitational interactions, we rely on mathematical formulae in physical theory. Those formulae are then the principles that we use to cognize nature and its laws.

¹⁶ For Kant, freedom is primarily a property of the will, and one must always be careful to distinguish will from choice (MM 13, 6:213). Yet the two end up going together, and Kant brings up free choice in the Canon in order to argue that “Practical freedom can be proved through experience.” (CR 675, A802/B830) So this distinction, while crucial, does not really affect my basic point, which is that a principle representing a law of freedom is connected with free choice, and so constitutes a “practical” principle, according to Kant’s explanation of the term in the Canon of Pure Reason.

Analogously, then, when we are doing philosophy of freedom, in the Architectonic's sense, the principles that we use to cognize freedom will take the laws of freedom as their proper objects. Therefore, on the basis of just this single paragraph from the Architectonic, with perhaps a little clarification from the Canon, we can see that Kant's practical principles represent the laws of freedom.

But are practical principles *fundamentally* representations of laws of freedom, according to Kant in the Architectonic? Kant hasn't really said anything about these laws yet, except to indicate that they contain the moral law, and that morals, at least, is concerned with what "**should be**". Nevertheless, he has still said enough to show that practical principles really are, fundamentally, representations of laws of freedom.

That follows from how Kant presents practical principles as an integral part of the structure of the system of rational cognition in general. To see why, let's briefly examine what the Architectonic says about systems of cognition.

Here is the first relevant paragraph, the second of the Architectonic, again quoted in full:

Under the government of reason our cognitions cannot at all constitute a rhapsody but must constitute a system, in which alone they can support and advance its essential ends. I understand by a system, however, the unity of the manifold cognitions under one idea. This is the rational concept of the form of a whole, insofar as through this the domain of the manifold as well as the position of the parts with respect to each other is determined *a priori*. The scientific rational concept thus contains the end and the form of the whole that is congruent with it. The unity of the end, to which all parts are related and in the idea of which they are also related to each other, allows the absence of any part to be noticed in our knowledge of the rest, and there can be no contingent addition or undetermined magnitude of perfection that does not have its boundaries determined *a priori*. The whole is therefore articulated (*articulatio*) and not heaped together (*coacervatio*); it can, to be sure, grow internally (*per intus susceptionem*) but not externally (*per appositionem*), like an animal body, whose growth does not add a limb but rather makes each limb stronger and fitter for its end without any alteration of proportion. (CR 691, A832-833/B860-861)¹⁷

To begin with, the main idea here, developed in the rest of the paragraph, is that a system is to be "the unity of the manifold cognitions under one idea". In turn, this idea is to be

¹⁷ Guyer and Wood translate Kant's Latin like so: '*coacervatio*' is "Literally, "heaped up."", '*per intus susceptionem*' is "from an internal cause", and '*per appositionem*' is "by juxtaposition" (CR 691n).

“the rational concept of the form of a whole, insofar as through this the domain of the manifold as well as the position of the parts with respect to each other is determined *a priori*.”

In short, this means that a system, in the target sense, is supposed to begin with an aggregate or a body of different cognitions—these are the “manifold cognitions”. This “manifold” is then to be unified under a single idea or concept, which is to determine which cognitions belong to the system, as well as how all of these cognitions are to be related to one another. As Kant puts it, the overarching idea for the system is to be the “concept of the form of a whole”, “insofar as” it determines “the domain of the manifold” and “the position of the parts with respect to each other”.

So, for example, consider physics. For Kant, physics is the system of cognitions of motion (MF 12, 4:476-477). I take it that something such as this is to be the idea behind physics, which is then supposed to determine which cognitions belong to physics. This idea would thereby determine which cognitions count as physical, and what ties them together, under this idea, is that they are all cognitions of motion.

Likewise, in determining how physical cognitions relate to one another, the idea is to determine the fundamental types of physical cognition. For example, Kant thinks that the idea behind physics groups one set of physical cognitions together insofar as they all consider “*motion as a pure quantum in accordance with its composition, without any quality of the movable*” (MF 12, 4:477). Never mind what that is supposed to mean: the point is that this is to be one of the fundamental types of physical cognition, according to the idea of physics.

Thus, Kant holds that a system of cognitions is an aggregate or a manifold of cognitions whose domain and structure is determined through an idea. Not just that, though—Kant also emphasizes that the idea is to determine the form and structure of its system on an *a priori* basis. For he insists that a system needs a “schema” for its complete execution.

To speak very roughly, this schema is to be an outline of the system, which is also to be determined through the idea for the system (CR 691, A833/B861). Kant adds that “a schema

that is not outlined in accordance with an idea, i.e. from the chief end of reason, but empirically, in accordance with aims occurring contingently (whose number one cannot know in advance) yields **technical** unity, but that [schema] which arises only in consequence of an idea (where reason provides the ends *a priori* and does not await them empirically) grounds **architectonic** unity.” (CR 691-692, A833/B861)¹⁸

And it is this *a priori* that is crucial for understanding what the Architectonic of Pure Reason tells us about Kant’s practical principles. Because the idea for a system is to determine the domain and structure of the system *a priori*, Kant thinks that this domain and structure must be necessary, at least relative to the underlying idea, rather than contingent on how the world turns out. (Or rather than fully contingent on the world, anyway—Kant allows room for parts of the system to be filled in through empirical investigation, as in physics.) Kant’s remarks on the schema of the system make these points very clear.

As such, when Kant’s project in the Architectonic is to outline the overall *system* of human cognition, he takes himself to be giving the necessary subdivisions and branches of human cognition. We know that much, even without knowing what the overall idea for this system is to be, and even without knowing what the idea is to be for any specific branch or subsystem.

Therefore, when the system of human cognition ends up treating practical principles as principles for cognizing laws of freedom, Kant will think that this is their necessary position and role in this system. For any other position or description would contradict the idea of the system, which can only be correctly organized in one way. Accordingly, this position will express what Kant takes to be fundamental to practical principles, as a kind of human cognition, which

¹⁸ Kant’s reference to “the chief end of reason” here is rather puzzling. Does he mean the chief end of reason in general, or the chief end of reason with respect to the particular system of cognitions that is at issue? The Architectonic does say that the idea for a system is to be an idea of an end—see the long passage already quoted above. I think Kant must mean the stronger claim, though, since he thinks that reason must ultimately be a single unity (G 5, 4:391), and because of the importance that he places on tracing the cognitive origins and boundaries of a type of cognition (CR 696, A842/B870). Kant does not give much indication of what this “chief end” of reason in general would be, but he does intimate that it must be practical (CP 102, 5:121).

is as much as to say that this position captures what he takes to be fundamental about practical principles for us. Moreover, as a result, this description of practical principles will be fundamental to explaining the domain and structure of laws of freedom, considered as a system of laws, corresponding to the system of practical principles that constitutes practical philosophy and philosophy of freedom.

1.3 Practical Principles in the *Groundwork of the Metaphysics of Morals*

So practical principles are fundamentally cognitions or representations of laws of freedom, according to Kant in the First Critique's Architectonic of Pure Reason—this follows from his views on systematicity in cognition, and from how he thinks that the system of human cognition in general uses practical principles as principles for cognizing freedom and its laws. But the *Critique of Pure Reason* is part of Kant's theoretical philosophy. How do his specifically ethical writings treat practical principles?

Kant says many things about practical principles throughout his ethical writings, and we will examine some of them in the next section, and in the next chapters. (But see note 2 above for some references.) It is hard to get a very clear picture of his overall views from those remarks, however. (As we will see.) So instead, let's begin here as we did with the First Critique. For the Preface to the *Groundwork of the Metaphysics of Morals* again provides a division for philosophy and philosophical cognition, as well as a concomitant account of practical principles.

The second paragraph of the *Groundwork's* Preface begins by dividing rational cognition into two types: “*material*” and “*formal*” (G 1, 4:387). The former, Kant says, is “concerned with some object”, while the latter is “occupied only with the form of the understanding and of reason itself and with the universal rules of thinking in general, without distinction of objects” (G 1, 4:387).¹⁹ “Formal philosophy”, he then says, “is called **logic**,” while material philosophy “has to

¹⁹ There are several puzzles about this distinction between formal and material rational cognition. First, why doesn't Kant include it in the Architectonic of Pure Reason? Is the *Groundwork's* Preface to be giving a different division, one that is equally “suitable to the nature of the subject” (G 1, 4:387), but still not the same as the one in the Architectonic? Or can this distinction be included in the Architectonic? If so,

do with determinate objects and the laws to which they are subject” (all G 1, 4:387).²⁰ And “these laws are either laws of **nature** or laws of **freedom**”, so material philosophy “is in turn divided into two”: “The science of the first is called **physics**, that of the other is **ethics**; the former is also called the doctrine of nature, the latter the doctrine of morals” (G 1, 4:387).²¹

Now, Kant does not mention practical principles at all here in the *Groundwork*’s Preface, nor does he explain what laws of freedom are to be, exactly, still less how the two are supposed to be related. We do, however, see the distinction between laws of freedom and rational cognition thereof recurring in this passage. For Kant says that material philosophy is one branch of rational cognition, while one branch of material philosophy “has to do with” freedom and its laws. This “science” of freedom is called both “**ethics**” and “the doctrine of morals”.²²

where, and why did Kant leave it out? Is it inessential to his purposes there? The second puzzle is about the distinction between formal and material cognition itself: why isn’t formal cognition—cognition “occupied only with the form of the understanding and of reason itself and with the universal rules of thinking in general, without distinction of objects”—also cognition that is “concerned with some object”, in the sense of Kant’s explanation of material cognition? Aren’t these forms and rules themselves objects, at least in the sense of objects of cognition? And not only objects, but determinate objects? Consider, for example, the representation of the *modus ponens* form of inference, as against the representation of the *modus tollens* form. Shouldn’t these be determinate objects of cognition? So what makes cognition in logic “formal” and cognition in physics “material”? Third, where is mathematics to fit into these divisions? Mathematics is not formal, according to Kant’s definition here, but he divides material cognition into only physics and ethics. I cannot try to answer these questions here, unfortunately.

²⁰ Kant does not explain here how philosophy is to be related to rational cognition, but we know from the Architectonic of Pure Reason that philosophy is to be a subtype of rational cognition. Specifically: rational cognition from concepts alone (CR 693, A837/B865). But if we restrict ourselves to just the text of the *Groundwork*, as we should wherever we can, Kant should still be read this way, since this is the weakest reading compatible with the text. Reading him as identifying rational cognition and philosophy would attribute to him a stronger claim that the text requires.

²¹ Full quotation: “material philosophy, which has to do with determinate objects and the laws to which they are subject, is in turn divided in two. For these laws are either laws of **nature** or laws of **freedom**. The science of the first is called **physics**, that of the other is **ethics**” (G 1, 4:387). I have broken it up above to try to clarify the line of thought here. Again, Kant does not explain what the relation between philosophy and science is to be, again, we ought to assume that science is to be a type of philosophy, and, again, this reading is ratified by the Architectonic, since a science there would be a system of philosophical cognition. However, neither text precludes “non-philosophical” sciences, like mathematics, which Kant thinks is a science, but not philosophy, since he thinks mathematics depends on pure intuition. Whether or not a body of cognition is a science, for Kant, is simply orthogonal to the type of cognition in that body. That follows from his description of sciences and systems in the Architectonic.

²² English grammar does not allow us to tell whether ethics is to be the science of freedom in general or only of the laws of freedom, and similarly for physics. For nothing in ‘the science of the first’ or ‘that of the other’ allows us to tell definitively what Kant wants to refer back to. However, Kant’s German is more

So one part of material philosophy will have to be concerned to discover and represent the laws of freedom, and it would be both natural and theoretically useful to mark this part out as the domain of “practical principles”. Kant gives some indication of accepting this usage, too, as when the Preface’s fifth paragraph explains that ethics is divided into two parts: a pure or rational part, and an empirical part, the latter of which “might be given the special name *practical anthropology*” (G 2, 3:388). For the idea of a “practical” anthropology as a component of ethics suggests that ethics more generally is the domain of practical philosophy.

Moreover, Kant mentions practical principles several times later in the *Groundwork’s* Preface, and in ways that support my interpretation, but never in a way that states decisively what practical principles are to be. Working chronologically through the text, Kant first says that “among practical cognitions, not only do moral laws, along with their principles, differ essentially from all the rest, in which there is something empirical, but all moral philosophy is based entirely on its pure part” (G 3, 3:389). He then goes on to explain that moral philosophy “gives to him [the human being], as a rational being, laws a priori” (G 3, 3:389).

Here, Kant counts the principles of moral laws “among practical cognitions”, which reinforces the idea that these principles should be considered both “practical” and as representations of laws. However, this is also one of the passages where Kant treats moral laws

specific. He says “*Denn diese Gesetze sind entweder Gesetze der Natur, oder der Freiheit. Die Wissenschaft von der ersten heißt Physik, die der andern ist Ethik*”. ‘*der ersten*’ in ‘*Die Wissenschaft von der ersten*’ has to be in the dative case, because the preposition ‘*von*’ requires the rest of the prepositional phrase to be declined in the dative. Hence, the whole phrase can only be a dative feminine singular construction: the dative plural here would have to be ‘*von den ersten*’, not ‘*von der ersten*’. So ‘*Die Wissenschaft von der ersten*’ must refer back to the singular feminine *Natur*, and not the plural *Gesetze der Natur*. Therefore, physics must be the science of nature in general, and not just of the laws of nature. However, ‘*der andern*’ in ‘*die [Wissenschaft] der andern*’ is in the genitive case, and therefore grammatically ambiguous between the plural and the feminine singular: ‘*der*’ and ‘*andern*’ would both be declined properly under either construal. But Kant clearly means for ethics and physics to be isomorphic at this level, so we should read him as saying that ethics is the science of freedom in general, and not just of the laws of freedom. Anyway, it is not clear whether very much in general hinges on how these sentences are understood. And nothing does for our specific purposes anyway, since ethics cannot be the science of freedom in general without also being the science of the laws of freedom, and it is this latter that is at issue for us here.

as themselves a kind of cognition. But he also describes how moral philosophy—which he has already said is a kind of rational cognition—gives laws to rational beings.

So Kant’s picture here could be one where practical principles represent moral laws, and then “give” us these laws through representing them. Practical principles would then end up representing at least some laws of freedom, too, because of how Kant equated ethics and “the doctrine of morals” at the outset of the preface.

But this text need not be read this way, either. Granted that moral laws are to form a unified system, the “principles” of moral laws could just be the moral laws that are fundamental to that system—they could be the moral law or laws that are the source of the others, and which organize all moral laws into a system. In that case, the system of moral laws would really be laws “all the way down”, and the principles of moral laws would not have to represent these laws.

The next, tenth, paragraph of the Preface goes on to say that “A metaphysics of morals is therefore indispensably necessary, not merely because of a motive to speculation—for investigating the source of the practical basic principles that lie a priori in our reason—but also because morals themselves remain subject to all sorts of corruption as long as we are without that clue and supreme norm by which to appraise them correctly.” (G 3, 3:389-390) Since reason is a faculty of cognition—as the Preface indicates from the start by outlining a system of rational cognition—“the practical basic principles that lie a priori in our reason” sound as if they should themselves be cognitions. Also, since these principles would have to provide the “supreme norm” for appraising morals, it sounds as if these principles should represent moral laws, and thereby represent laws of freedom, as before.

However, ‘practical basic principles’ (correctly) translates Kant’s ‘*praktischen Grundsätze*’, rather than ‘*praktischen Prinzipien*’. And, as I mentioned in the previous section, it is hard to interpret Kant’s notion of a *Grundsatz*. It helps my thesis that Kant says “*praktischen Grundsätze*” directly, but if we cannot say that *Prinzipien* are to be a type of *Grundsätze*, for Kant, then we are really adding problems here, not resolving them.

Last, the Preface's tenth paragraph continues: "Now the moral law in its purity and genuineness (and in the practical this is what matters most) is to be sought nowhere else than in a pure philosophy; hence this (metaphysics) must come first, and without it there can be no moral philosophy at all. That which mixes these pure principles with empirical ones does not even deserve the name of philosophy" (G 4, 3:390). Once more, Kant's idea appears to be that practical principles represent the moral law—for that is how "the moral law in its purity and genuineness" looks as if it would be found within "a pure philosophy", granted that the latter is a kind of rational cognition. Still, though, this passage does not determine Kant's usage, either.

However, besides the consideration of these specific passages, we can make one general observation about how Kant uses 'practical principle' in the Preface: in speaking of "pure" and "empirical" principles, and in connecting principles to cognitive faculties such as reason, he implicitly treats them as a kind of cognition. Kant really draws two separate, but interrelated distinctions using 'pure' and 'empirical': a "semantic" distinction, so to speak, having to do with the kind of content in a cognition, and an epistemic distinction that is concerned with how a cognition can be shown to represent its objects accurately (CR 137, B2-6). Either way, though, purity and empiricity are primarily properties of representations, for Kant, and not properties of the things represented. So when Kant says that a principle is pure or empirical, it follows that that principle ought, strictly speaking, to be a cognition. It would still make sense to describe the object of the principle as pure or empirical also, by virtue of the character of its representation. And Kant does often speak that way: we just observed him say "the moral law in its purity", for example. But this is still an extended sense of Kant's terms, derivative from their proper application to cognitions. So the fact that these concepts apply primarily to cognitions justifies us in treating the primary sense of 'practical principle' in similar terms, in which case they will be cognitions, representing laws of freedom, as per the beginning of the Preface. And this reading seems especially warranted when Kant's specific usage of 'practical principle' can be read in these terms, and is even perhaps most naturally understood that way.

So ethical and moral philosophy will include representations of laws of freedom, according to the Preface, which creates a theoretical need for a term for these representations. And Kant's usage in the Preface at least supports calling these representations "practical principles". Now, is this representing of laws of freedom supposed to be the fundamental fact about Kant's practical principles, even here in the *Groundwork*?

The Preface does not answer this question any more explicitly than the Architectonic of Pure Reason does. But we can also give the same answer on Kant's behalf here, too: we can rely on the systematicity of rational cognition once again, because Kant introduces the distinction between formal and material rational cognition in order to give the "principle" behind the ancient Greek division of philosophy into the sciences of "**physics, ethics, and logic.**" (G 1, 4:387) And reapplying our previous answer is especially apt when Kant claims that we need to add the principle for this tri-partite division of philosophy "partly so as to insure its completeness and partly so as to be able to determine correctly the necessary subdivisions" (G 1, 4:387), just as the Architectonic requires the idea for a system to determine the necessary domain and structure of thereof. By adding this principle, Kant thinks that we can go from a division that "is perfectly suitable to the nature of the subject" (G 1, 4:387) to one that is also complete, and with its "necessary" subdivisions. So, as in the Architectonic, this principle is to reveal the necessary position of practical principles in philosophy, considered as a system of cognition. By parity of reasoning, then, the fact that practical principles are to represent laws of freedom will turn out to be fundamental to them, at least in Kant's thinking.

1.4 Two Explications of Practical Principles

So, to recap: we started out by looking at Kant's general views on cognition, both in the *Critique of Pure Reason's* *Stufenleiter* passage and in the concurring explications from the Jäsche lectures on logic. These texts, I argued, should best be understood as treating principles in general as a kind of cognition. On the one hand, both texts treat them this way directly, and, on the other hand, principles satisfy the general explication of cognition that both texts share.

These texts also insist that intuitions and concepts are the only kinds of cognitions, though, which poses a problem for my reading. But I indicated how I think this problem is to be resolved. I could not fully argue for that resolution here, however, due to space constraints. But I also explained how my main proposal about practical principles should be understood, in light of the different answers we might give to this problem.

Next, we examined Kant's general treatment of principles in the First Critique's Transcendental Dialectic, where he claims that reason is the faculty of principles, and where he distinguishes three different ways in which a cognition can be a principle. So I argued that this passage reinforces the idea that principles in general are a kind of cognition, for Kant, while also adding the idea that principles must be universal, if only in how we use or apply the principle, and not universal in its own inherent validity, too. As universal, then, I argued that distinctively practical principles would represent laws, since they would be universal principles for our actions, and thus would be a kind of universal causal principle.

Even so, that argument is not enough to interpret practical principles as cognitions of laws *of freedom*, still less is it sufficient to say that this is what is fundamental about practical principles, for Kant. To reach both of these stronger conclusions, we looked at what Kant has to say about systems of cognition and about the specific location of practical principles in the system of human cognition, both in the First Critique's Architectonic of Pure Reason, and in the Preface to the *Groundwork of the Metaphysics of Morals*.

Over the course of the discussion here, I have brought up a few passages that pose problems for my interpretation of practical principles. Besides the problem about cognition already mentioned, we examined one example of a passage where Kant treats laws themselves as a kind of practical cognition. This passage was from the *Groundwork's* Preface, but there are other similar passages elsewhere, too. I explained that Kant's usage here makes sense, given the isomorphism that is supposed to obtain between the content of a cognition and the nature of the

object that it purports to represent. But I claimed that we should still separate representations from their objects, in the interests of clarity.

In any case, however, I now want to conclude with two passages where Kant almost explicitly treats practical principles themselves as cognitions. These passages are particularly important, because one starts a crucial argument in the *Groundwork*, and the other is explicitly a “Definition” of practical principles, which also supports a crucial argument. We will look at these passages in much more detail in later chapters, but they are still worth introducing now.

First, one of the main arguments in the *Groundwork* is supposed to show that the Formula of Universal Law has to be the fundamental principle of moral obligation, regardless of whether or not there is any such thing as moral obligation in the first place. I will examine this argument in much more detail in a later chapter. But Kant’s strategy for this argument is, first, to “follow and present distinctly the practical faculty of reason, from its general rules of determination to the point where the concept of duty arises from it” (G 24, 4:412). And then, from there, Kant seeks to derive the Formula of Universal Law from the concept of duty—which is the concept of a categorical imperative, Kant argues (G 30-31, 4:420-421).

Here is how this argument begins: “Everything in nature works in accordance with laws. Only a rational being has the capacity to act *in accordance with the representation of laws*, that is, in accordance with principles, or has a *will*. Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason.” (G 24, 4:412) Here, Kant identifies principles and the representations of laws, and this identification is crucial to the argument. His stated strategy is to “present distinctly the practical faculty of reason” and “its general rules of determination”. Then, in executing this strategy, he says that these general rules of determination are representations of laws, while the practical faculty of reason then turns out to be the capacity to derive actions from these representations of laws. Identifying principles with these representations of laws then allows practical reason to be the faculty for deriving actions from principles. And Kant needs that identity, if he wants to argue from here to the

conclusion that the fundamental law for this faculty of practical reason tells it to act according to principles (maxims) that it can will as a universal law.

Kant does also say that reason “is required for the derivation of actions from laws”, which makes it sound as if we can derive actions directly from laws. In context, though, “the derivation of actions from laws” looks to be a simple abbreviation of “the derivation of actions from the representation of laws”, especially because we need to represent a law in order to derive anything from it through a cognitive faculty such as reason.

So in short, Kant’s strategy in this crucial argument hinges on treating principles as representations of laws. And since we are to derive actions from these principles, they must constitute practical principles.

For the second main passage on practical principles, the body of the *Critique of Practical Reason* begins with a “Definition”: “Practical *principles*”, Kant says, “are propositions that contain a general determination of the will, having under it several practical rules.” (CP 17, 5:19) Again, this definition is crucial to the Second Critique’s argument that the Formula of Universal Law has to be the fundamental principle of morality (at least if there is any such thing). For Kant uses this definition to explain what subjective and objective principles are, and he needs these concepts in order to argue that the Formula of Universal Law is the only principle that could be the fundamental moral principle. Yet, here again, Kant defines practical principles in terms of cognition, since he says that they are “propositions”.

Now, to be sure, not everyone thinks that propositions are something like Kant’s cognitions. For example, Russell held, at least at one point, that propositions are, or at least include, structured states of affairs in the world. (See his letter to Frege, dated 12 December 1904, in Frege (1980).) So we should not assume that Kant means one thing in particular here.

I have already pointed to some other passages in Kant, such as the Transcendental Dialectic and the Jäsche *Logic*, indicating that he does think of propositions as a kind of cognition. However, Kant’s “Remark” on this definition of practical principles reinforces this

reading. In that Remark, Kant says that “In practical cognition—that is, cognition having to do only with the determining grounds of the will—the principles that one makes for oneself are not yet laws to which one is unavoidably subject, because reason, in the practical, has to do with the subject, namely with his faculty of desire, which by its special constitution can make various adjustments to the rule.” (CP 17-18, 5:20) And he also goes on to say that “A practical rule”—such as those that practical principles have “under” them, according to the definition—“is always a product of reason because it prescribes action as a means to an effect, which is its purpose.” (CP 18, 5:20)

Here, once again, Kant treats practical laws themselves as a kind of cognition, rather than as the object cognized. But, primarily, both of these passages, from the *Groundwork* and from the *Critique of Practical Reason*, are clear in treating practical principles as cognitions.

More work would be needed to show that these passages also end up treating practical principles as representations of laws of freedom. But, at minimum, this chapter has illustrated how one strand of Kant’s writing, across several of his critical texts, converges on the picture of practical principles that I am proposing. Given the issues that I have also raised, perhaps we should ultimately think that Kant’s text is inconsistent on this point, or perhaps we should ultimately even prefer a different interpretation of practical principles. Still, though, it would remain open to me to defend my interpretation on the grounds that Kant needs some concept corresponding to my interpretation of practical principles, because of the role that representations of laws of freedom need to play in his moral theory and in his theory of rational agency. The rest of the study here will be devoted to articulating this role, and my overall interpretation can also be justified by the light that it casts on the rest of Kant’s moral theory.

For we have not learned anything yet about Kant’s views on laws of freedom themselves, about what moral laws are to be, for him, or about how the two are to be related. Still less have we learned anything about the Formula of Universal Law or its component parts, such as maxims. That will be the work of the forthcoming chapters. However, if I am right about

practical principles, we now have a thread to follow: Kant's views on freedom. These views are tremendously complicated, of course, but we will begin to unpack them in the next chapter.

2. Freedom, Causation, and the Structure of Practical Principles

As the previous chapter explained, the main idea of my project here is that Kant's practical principles are, fundamentally, representations of laws of freedom. That chapter also outlined the main arguments for understanding Kant's practical principles this way, as best as could be done in the absence of a more thorough discussion of how principles in general are to fit into Kant's overall theory of representations, anyway. But if my main thesis is correct, then we will need to explore Kant's views on freedom, if we want to articulate his concept of practical principles any further. For that procedure is just what the last chapter required, in saying that representing laws of freedom is "fundamental" to practical principles.

As I said, though, none of this is to say that other approaches to Kant's practical principles cannot illuminate his thinking about them. Not by any means. But if freedom is fundamental to Kant's concept of practical principles, then freedom is still where a properly systematic interpretation of his views here needs to start.

My approach does have a significant drawback, which others lack, however: namely, Kant's thinking about freedom is quite obscure and difficult, to say the least. In the last chapter, I proposed that freedom is, for Kant, fundamentally a kind of causality or causal power. If I am correct about that, then we will need to start explaining Kant's views on freedom through his views on causation.

But Kant's views on causation are certainly quite obscure and difficult, too. Still, my goal for this chapter is to defend treating Kant's freedom as fundamentally a kind of causality, and to use his most general thinking about causation in order to articulate what he takes to be the "basic structure" of causal relations and of the principles that represent them: their basic parts, along with how these parts are to be related to one another.

This basic structure for causation is to capture the structure that all causal relations have in common, for Kant. Accordingly, the structure of any specific causal relation will turn out to be an instance of this more general causal structure. Gravitational interactions, say, must exhibit

this basic structure, and if Kant does think that freedom is fundamentally a kind of causality, then free causality must have this basic structure for him, too.

Specific causal relations might have their own specific structures, to be sure, so long as they conform to the basic structure. However, we will see that specific kinds of causal relations are to be *instances* of the basic kind of relation, so it will follow that any specific structural parts and relations will have to be subsidiary to the basic components and relations, in Kant's picture. Moreover, Kant never, so far as I know, gives examples of laws or principles with more structural parts than those in the basic structure that I will propose here. In fact, the next chapter will illustrate how the cases that are problems for my interpretation are all on the other side, since Kant gives many examples of principles that appear to lack some of the parts that my interpretation requires.

In any case, though, because this basic structure will be common to all causal relations in Kant's theory, this structure will also represent part of Kant's most general thinking about causation. Beginning at such a high level of generality has its risks, to be sure. Just for one example, this approach risks losing touch with the nature of concrete causal interactions. But this strategy also has its theoretical and interpretive advantages.

For one thing, this chapter hopes to show that Kant's most general views about freedom and causation are actually quite clear, despite the difficulty and complexity endemic to the more specific details of his views. Relatedly, remaining at this highest level of generality will also allow us to pass over those complicated technical details. So, for example, we will not have to address questions about Kant's timeless, noumenal causation, or about whether he thinks that events can be causes.¹ Mainly, though, this highest level of generality is simply the appropriate one for the question about the *basic* structure of Kant's causal relations.

¹ For example, Watkins (2005) argues that events cannot be causes, for Kant. I disagree, but that is a topic that I pursue elsewhere. See note 11 below for a little more discussion, however.

Still, despite the high level of generality at which this chapter will operate, explicating Kant's views about the basic structure of causal relations will move the current project forward. Causal principles are to represent causal relations, according to the interpretation of principles defended in the previous chapter. So Kant's views on the basic structure of causal relations will have to apply to his views on the basic structure of causal principles at the same time.

So if Kant thinks that practical principles represent laws of freedom, while laws of freedom are to be a kind of causal law, too, then practical principles will turn out to be a kind of causal principle, for Kant. In which case his practical principles will have to exhibit his basic causal structure. For their specific structure, whatever it turns out to be, will have to be an instance of this basic causal structure, in the same way that the specific structure of laws of freedom, as a kind of causal law, will have to be an instance of the more general, basic causal structure.

Then, since Kant thinks that maxims are a kind of practical principle, we will be able to use his views on the basic structure of practical principles in order to determine the basic parts of maxims, along with how those parts have to be related to one another. Thus, the work in this chapter will lay a foundation for a new account of what Kant's maxims are to consist in. I will develop that account in the subsequent chapters, use it to interpret Kant's views on willing maxims as universal laws, and apply it to some of the venerable recurring problems in understanding Kant's moral theory.

The method in this chapter, though: first, I will explain Kant's views on freedom, enough to show how they are related to his views on causation. Next, I will examine his most general views on causation, in order to determine what he takes the basic structure of causal relations to be. I will then explain, in a preliminary way, the picture of practical principles that we get from this interpretation of the structure of causal relations; further argumentation, textual analysis, and problem-resolving will wait until the next chapter. Last, I will conclude with some general discussion of how free agents relate to the structure of their own causality and to the laws and

principles thereof, according to Kant. This will involve giving a preliminary account of how Kant tries to reconcile freedom and necessity, in order to explain how a free will need not be arbitrary.

To elaborate a little, in §2.1, I will argue that the crucial point is that freedom, according to Kant, consists in the property of being an absolutely self-determining causality. For Kant, then, laws of freedom will be, fundamentally, the laws for this kind of causality.

Next, to understand Kant's most general views on causation, and to determine the basic structure of his causal relations, I will turn to the Metaphysical Deduction of the Categories from the *Critique of Pure Reason*. There, Kant seeks to derive the concept of causation from the "hypothetical" relation in general logic—from the relation between a proposition and its logical consequences, for example. Kant describes this hypothetical "if/then" relation as a relation between a "ground" and its logical "consequence". So §2.2 will use the Metaphysical Deduction to argue that causation, for Kant, is fundamentally an instance of this ground/consequence relation in logic.

Consequently, I will argue that all causal laws must relate three general parts, for Kant: a cause, as a kind of ground, an effect, as a kind of consequence of a ground, and an action whereby the cause causes the effect, as an instance of the logical ground/consequence relation itself. Moreover, Kant thinks that this underlying ground/consequence relation is logically necessary. Thus, Kant insists that causation, as an instance of a logically necessary relation, must be a necessary relation, too: given the existence of the cause, the existence of its action and its effect must necessarily follow, in the same kind of way that the truth of the logical consequences must follow from the truth of their ground. As such, I will argue that causal laws in general, for Kant, consist at least in necessary relations between the existence of a certain cause and the existence of its consequent action and effect.²

² In the hope of avoiding any possible confusion: as I understand Kant, the necessity in causal relations is not at all to imply that every possible world has the same causal relations or causal laws. Kant does not really adopt a "possible worlds" interpretation of modality at all, so far as I can see (CR 321-326, A218-226/B265-274). In fact, I think that he is rather deeply skeptical of our ability to do modal metaphysics at

§2.3 will then seek to apply Kant's basic structure for causal laws to his views on laws of freedom, so as to introduce my account of the structure of Kant's practical principles. In this context, we are talking about causal laws for agents, and particularly for free agents. So we are really talking about the causal laws that free agents choose to govern their actions.

To elaborate: if §2.1 is correct, and Kant holds that freedom is an absolutely self-determined causality, then for him the ultimate or proper cause of a free action must always be its agent and their free choices. However, Kant thinks that free actions will still require grounds, if they are not to be arbitrary. To put these two thoughts together, Kant thinks that a free cause must be a free agent, plus whatever they think makes it necessary for them to act a certain way.

Accordingly, §2.3 will argue that laws of freedom in general will need three basic parts, according to Kant: an effect, an action to bring about that effect, and a ground, which the agent thinks makes it necessary, in some sense, for them to carry out that action, to achieve that effect. In other words, a law of freedom "says" that when this ground exists, it consequently becomes necessary, in some sense, for this free agent to act a certain way to bring about a certain effect.

Kant then thinks that this "practical" necessity will come in different kinds, which we can use to distinguish the different kinds of laws of freedom, along with the different kinds of the

all. Instead, as I understand Kant, the necessity in causality consists simply in how the *existence* of an action and an effect *must follow from the existence* of their cause. This necessity can be compartmentalized to a single possible world. Moreover, as I understand Kant, the existence of an action and an effect are to be necessary only *relative to* the existence of their cause. Their existence is not, by any means, to be necessary *absolutely* or *intrinsically*, in any of the ways that God or numbers are sometimes supposed to exist necessarily, for example. (Existing in all possible worlds, essence proves existence, etc.) Accordingly, if I sometimes say that causes make effects necessary, or that causes necessitate their effects, I will mean that necessity only in the sense of a necessity *relative to* or *contingent upon* the cause, and not in the sense of any absolute or intrinsic necessity. And certainly not in the sense in which the definition or construction of a right triangle makes the Pythagorean Theorem necessary, either. Moreover, in this spirit, I will be speaking throughout of how causes "necessitate" their effects and, more generally, of how grounds "necessitate" their consequences. This is supposed to be a different, and more general, kind of necessitation than the kind that Kant discusses in the *Groundwork*, for example, at (G 24, 4:413). That necessitation is to be the kind that is connected to the concept of an imperative, and which is to be involved in how reason and its laws relate to imperfectly rational wills. According to Kant, then, perfectly rational wills are not subject to that kind of necessitation (G 25, 4:414), but they will remain subject to the kind of necessitation that I will be discussing, just because perfectly rational wills "equally stand under objective laws (of the good)" (G 25, 4:413). I take it that the specific technical notion of necessitation in the *Groundwork* is supposed to be an instance of the more general kind that we will be concerned with here. Thanks to Tyler Burge for pressing me to be far clearer on these points.

practical principles that represent these laws.³ Kant recognizes two main kinds of practical principles: maxims and practical laws. Ultimately, then, these different kinds of practical necessity will help us to determine what practical laws in general consist in, for Kant, and to distinguish them from maxims, all in later chapters.

For example, I will contend that Kant's paradigm case is maxim of the lying promise: "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen." (G 32, 4:422) This maxim both supports and illustrates my account, I argue, because it says that whenever I believe myself to be in need of money, I will act a certain way: I shall make a false promise in order to receive a loan that I know that I cannot repay. This is exactly the ground/action/effect structure that I am proposing for Kant's practical principles. Moreover, this maxim exhibits the requisite practical necessity in two ways: first, in how the false promise is supposed to follow from the belief that I need money, and, second, in how I cannot satisfy this maxim unless I act this way, whenever this ground obtains. Whether this principle is only supposed to be a maxim or whether it is also to be a practical law then depends on the specific character that the necessity here is supposed to have.

Clearly, much more will need to be said about this account of practical principles. In this chapter, though, I will conclude with some general remarks about how Kant thinks that free agents will relate to practical principles, if my interpretation of his practical principles is correct. What is it for a free agent to adopt a practical principle? What is it for a practical principle to be

³ The primary kinds of practical necessity, for Kant, are subjective and objective. But he thinks that practical necessities can also be absolute and relative, categorical and hypothetical, unconditional and conditional. For example, suppose that I act on my inclinations. Then Kant will think that these inclinations, through my free choice, will have made my actions practically necessary both subjectively and relative to those inclinations. But he also thinks that the form of certain maxims will, just by itself, make certain actions practically necessary absolutely and objectively (G 24-27, 4:412-417). These categories are not mutually exclusive, though: a single action can be practically necessary in all these ways at the same time. When we do our moral duty because it is our moral duty, then the action is supposed to be necessary objectively, absolutely, unconditionally, and categorically, just because it is our moral duty. Moreover, this action is also supposed to be necessary subjectively, because we do it, as well as relatively, conditionally, and hypothetically necessary, because we do it from our recognition of and respect for moral duty. I will elaborate the notions of subjective and objective practical necessity a little in §2.3, and more in the next chapter.

valid for a free agent, and how does the validity of a practical principle relate to whether or not a free agent has already adopted it? How is it possible for a free agent to be under the necessitation of a ground or law at the same time?

The next chapters will then continue to defend and develop this chapter's account of the structure of Kant's practical principles. In Chapter 3, I will explain how some other fundamental lines of Kant's thought converge on the interpretation of practical principles proposed here, and Chapter 4 will discuss some alternative interpretations of Kant's maxims.

I will then turn to Kant's argument that the Formula of Universal Law must be the fundamental principle of morality. Chapter Five will attempt to explain and offer a general defense of this "derivation" of the Formula of Universal Law, while Chapter Six will discuss some of the long-standing specific objections to that argument.

Chapter 7 will then conclude the study here by looking at the *Groundwork's* argument that moral obligation is real: that we really are subject to moral obligation, and, therefore, to the demands of the Formula of Universal Law.

For now, though, let's start with freedom and causation.

2.1 Free Causes⁴

So, why should we understand Kant to be saying that freedom is fundamentally a kind of causality? For the sake of simplicity, and for continuity with the forthcoming discussion of causation, I will focus here on the *Critique of Pure Reason*, and only cite supporting passages from Kant's other texts, without detailed analysis.

⁴ As will become clearer presently, our topic here is really what Kant calls "transcendental" freedom, as distinguished from "practical" freedom. In transcendental freedom, Kant says that "reason creates the idea of a spontaneity, which could start to act from itself, without needing to be preceded by any other cause that in turn determines it to action according to the law of causal connection." (CR 533, A533/B561) "**Freedom in the practical sense**", by contrast, "is the independence of the power of choice from **necessitation** by impulses of sensibility." (CR 533, A534/B562) See also (CR 675-676, A802-803/B830-831) and (G 52-53, 4:446-447). Transcendental freedom is the stronger notion, because it includes independence from all causes "outside" itself, not just sensible impulses. Kant arguably recognizes other kinds of freedom as well; see Beck (1987), for example. The relations between these notions of freedom are quite complicated, and I don't have space to discuss them here, though I say some things about them elsewhere. However, Kant does consider transcendental freedom to be freedom in the "strictest sense" (CR 116, Bxxviii), which is part of why it is the focal point of my discussion here.

To begin with, then: Kant thinks of freedom as a property of a being's will or volition.⁵ For example, in the Preface to the Second Edition of the *Critique of Pure Reason*, Kant claims that we need to be able to think of objects in two ways: both “as objects of experience” and “as things in themselves”, apart from our experience. For otherwise, he says, “I would not be able to say of one and the same thing, e.g., the human soul, that its will is free and yet that it is simultaneously subject to natural necessity, i.e., that it is not free, without falling into an obvious contradiction” (all CR 115, Bxxviii). He goes on in the same vein, a little later in this Preface:

if the critique has not erred in teaching that the object should be taken in a **twofold meaning**, namely as appearance or as thing in itself...then just the same will is thought of in the appearance (in visible actions) as necessarily subject to the law of nature and to this extent **not free**, while yet on the other hand it [just the same will] is thought of as belonging to a thing in itself as not subject to that law, and hence **free**, without any contradiction hereby occurring. (CR 116, Bxxvii-xxviii)⁶

And still later in this Preface, Kant asks us to “suppose that morality necessarily presupposes freedom (in the strictest sense) as a property of our will” (CR 116, Bxxviii).⁷

Now, granted that freedom is to be a property of a being's will, for Kant, what kind of property is it to be? The above passages already give us some indication, insofar as they explain

⁵ Kant famously uses ‘will’/‘*Wille*’ in two senses: a narrow sense, where there is at least a conceptual distinction between the faculties of will and choice (*Wille* vs. *Willkür*) (MM 13, 6:213), and a broad sense, where ‘will’ is to refer to the combination of these two faculties. We are only interested in Kant's thinking about the will in the broad sense, though, so we will not need to worry about these complications. In any case, I take it that Kant's thinking is slightly better captured by treating freedom as a property of the will rather than of volition. Freedom is to be a property both, of course, but I take it that, for Kant, freedom is primarily to be a property of the *faculty* of volition, i.e. of the will, and only secondarily of the acts of that faculty. My reasons for reading him this way should become apparent in the following discussion. Focusing on freedom as a property of the will is actually a substantive claim, however, both in Kant's own terms, and in thinking about freedom more generally. One might take free choice/*Willkür* as paradigmatic instead of free will, even in Kant's own terms, as he actually seems to do, in a certain sense (MM 18, 6:226). Or, more generally, one might prefer to think primarily in terms of free or unfree volition/*Wollen*, agency, or action for example. Perhaps even in terms of free endorsement or identification, as in a version of (Frankfurt 1971), or in terms of what kinds of attitudes it would be appropriate to take toward someone (Strawson 1962). But those issues are quite beyond the scope of our study here.

⁶ These are very important texts for the “two aspect” interpretations of Kant's distinction between the objects of our experience, a.k.a. appearances, and things in themselves. These readings contrast with the “two worlds” interpretations that I tend to favor. Nothing I say here will hinge on this question, however.

⁷ See also (G 52, 4:446) and (CP 26, 5:28-29). Freedom “in the strictest sense” is transcendental freedom, as I mentioned in note 4 just above.

freedom by relating it to the laws to which the will is subject. Here, Kant's point is primarily negative: he says that the will cannot be free if it is "subject to natural necessity" or "subject to the law of nature".⁸ Of course, "subject to the law of nature" and "subject to natural necessity" are both a little ambiguous. However, Kant means to say here that a free will is not *determined by* laws of nature, or at least not *solely determined* by the laws of nature and by the causal processes within nature that obey these laws. Nevertheless, if a free will is to act within nature, it must still be determined *according to* or *in conformity with* laws of nature, he thinks. And a free will also be *influenced by* natural causal processes, at least when the will in question is finite (CR 535-537, A538-541/B566-569; G 58, 4:453; MM 13, 6:213).

Kant confirms this point about freedom and natural laws later on in the First Critique's Third Antinomy. The Third Antinomy presents conflicting arguments about freedom: one for a Thesis, holding that there must be freedom, and one for an Antithesis, holding that freedom is impossible. The Thesis holds that "Causality in accordance with laws of nature is not the only one from which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them." (CR 484, A444/B472) In contrast, the Antithesis holds that "There is no freedom, but everything in the world happens solely in accordance with laws of nature" (CR 485, A445/B473). Hence, the crucial point about freedom is to do with causation: for Kant, freedom is not natural causation, nor is it reduced to the chain of natural causes. Instead, freedom is to be "another causality" (CR 484, A444/B472), "a special kind of causality" (CR 485, A445/B473), the status of which is at issue in the Third Antinomy.⁹

⁸ Again, also cf. (G 52-53, 4:446-447).

⁹ Yet again, (G 52-53, 4:446-447). Note that if freedom is to be a property of the will's causality, under my interpretation of Kant, and if the will is to be the faculty of practical reason as well (G 24, 4:412; CP 48, 5:55; MM 13, 6:213), then I am committing Kant to the idea that reason is itself a causal power or a kind of causality. And Kant does often speak like this, in fact: see (CR 541-542, A549-550/B577-578) and (CP 3, 5: 3) for particularly striking examples. Allison (1990, 49-50), however, argues that this language cannot be taken literally. So I will owe him an answer, which I attempt to provide elsewhere. In a sense, though, my entire project here is supposed to explain how reason is a cause, according to Kant.

Later on, when Kant is finally resolving the conflict in the Third Antinomy, he sums up what the connection between freedom and causality is to be quite nicely, saying that “In respect of what happens, one can think of causality in only two ways: either according to **nature** or from **freedom**.” (CR 532, A532/B560) As a property of a being’s will, then, freedom is supposed to describe the nature of its causality. As such, freedom is to have two dimensions: the negative dimension described above, wherein a free will is not to be determined by natural causality, at least not by natural causality alone, and a positive dimension, still to be articulated.

We are now nearing the limits of what we can say about Kant’s views on freedom without having to really sort through the complicated details. Still, it is fair to infer from the passages discussed so far, particularly from the one just mentioned, that freedom is fundamentally a kind of causation belonging to the will, according to Kant.

And we can even go a little further, in order to give a general characterization of what freedom is to be, for Kant, positively speaking. In Kant’s view, the primary difference between natural and free causation is that a natural cause is also determined by another natural cause “outside of” or “external to” it, so to speak, where this external cause is also prior in time.¹⁰ As Kant puts it, continuing immediately from the previous quotation: “In respect of what happens, one can think of causality in only two ways: either according to **nature** or from **freedom**. The first is the connection of a state with a preceding one in the world of sense upon which that state follows according to a rule.” (CR 532, A532/B560)¹¹ Kant’s subsequent discussion clearly shows

¹⁰ However, even an instant of time can be ordered temporally, according to Kant (CR 312, A202-203/B247-249). So causes and effects can still be simultaneous, he thinks, and he even claims that most causes and effects are, in fact, simultaneous. Consider gravitational attraction, for example. It may seem impossible for an instant of time to have a temporal order, but if calculus allows us to determine the slope of a geometric point, then why not? And whether time is discrete or continuous, each moment is still related to the moments before and after, which could be enough to allow an instant to have a temporal structure. More on this in the next section.

¹¹ This passage and the subsequent discussion are a problem for Watkins’ (2005) interpretation of Kant on causation, since here Kant describes natural causation as a connection between states/*Zustände* in the sensible world. Later, Kant often speaks of causation between occurrences/*Begebenheiten*, and also of causation between things that happen/*geschehen* in appearances as well. Watkins, however, maintains that causes are never states or events, according to Kant, but only the indeterminate causal powers

that he means “preceding” in a temporal sense, and he naturally goes on to point out that this preceding state in the world of sense must have its cause in turn, and so on, indefinitely.¹²

See also Kant’s description of natural causation back in the Third Antinomy’s Thesis: “Assume that there is no other causality than that in accordance with laws of nature: then everything **that happens** presupposes a previous state, upon which it follows without exception according to a rule.” (CR 484, A444/B472) Here again, Kant means “previous” temporally, and, again, this temporally previous state will also count for Kant as something that happens. So he thinks that it, too, will presuppose a still more previous state from which it follows, and so on, and so forth. Thus, once again, Kant’s claim is that all natural causes are determined by some temporally prior natural cause. In other words, for each natural cause, there is to be another natural cause, prior in time, from which the former necessarily follows.

My own view is that this description of natural causes—a natural cause is one determined by another, temporally prior natural cause outside of itself—is actually constitutive of natural causation, according to Kant. I take it that the passages quoted above are naturally read as describing just what Kant takes to be constitutive of natural causation, and the First Critique’s schematism of causation, which I will discuss in the next section, also has this implication. But nothing will hinge on whether this description is to be constitutive of natural causation or merely characteristic thereof. For my main concern is just about what inferences Kant’s notions

belonging to substances, which then produce their effects through indeterminate activities (2005 262, 296-297, e.g.). So Watkins argues that Kant is not speaking in his own voice, but within the context of the conflict in the Third Antimony, when he says that things like states can enter into causal relations. I do not think that Watkins is correct about that, but I discuss this debate in more detail elsewhere.

¹² Indefinitely, but not, I take it, *ad infinitum* (CR 518-519, A504-506/B532-534; 522-523, A512-515/B540-543). Also, the distinction between temporal and causal priority not only elucidates Kant’s views on causation in general, but it is also important for the question of whether there is a first cause. Would it be first in time, or only causally first? Either way, Kant’s view is that there is no first cause in nature (CR 538, A543-544/B571-572).

of causality allow us to draw for the basic structure of causal laws and of the principles that represent them.¹³

But be that as it may, Kant would agree that something seems paradoxical about describing natural causation as causation determined by a temporally prior, external state of the natural world. And especially so if this description is supposed to reflect what constitutes natural causation. Perhaps there is not ultimately any real problem here, but the Thesis of the Third Antinomy tries to prove that “the proposition that all causality is possible only in accordance with laws of nature, when taken in its unlimited universality, contradicts itself, and therefore this causality cannot be assumed to be the only one.” (CR 484, A446/B474)

The problem, I take it, is that if all causation is only natural, then nothing is a cause just *in itself*. If every natural cause presupposes a prior natural cause determining it, then this chain of prior causes would have to extend backwards endlessly, if there is only natural causation. Natural causation would then seem to be radically ungrounded, because each natural cause presupposes another one, and there is never a first cause to conclude or ground this sequence of causes. And if natural causes must always be determined by other natural causes, to say that natural causation is radically ungrounded is to say that there really is no natural causation after all. Therefore, the supposition that all causation is natural seems to contradict itself, as Kant indicates.¹⁴

¹³ However, when it comes to Kant’s views on natural causation, there are two elements that we should really separate: a cause that is determined by another, external cause, and a cause in appearances. For Kant, either description could plausibly constitute natural causation in general, though only the latter would require temporal ordering. There is no space for such questions here, though I attempt to sort them out elsewhere.

¹⁴ Kant agrees with the Thesis’ conclusion in one way, of course, but not in another. Treating nature as a system of appearances, as Kant’s transcendental idealism urges, is supposed to allow all causation in nature to be externally determined, while preserving “space” for natural causation to have an ultimate ground in things in themselves. In contrast, “If we would give in to the deception of transcendental realism [and hold appearances to be the things in themselves (CR 511, A491/B519)], then neither nature nor freedom would be left.” (CR 538, A543/B571) In any case, Kant’s reasoning here might not be sound. Tyler Burge suggested to me, in his comments on an earlier draft of this chapter, that the last step in the Thesis’ argument might be invalid. We may not want to grant that natural causation has to be externally determined in the first place. But even if we do grant that, and even if that external determination means

Instead, the Thesis of the Third Antinomy seeks to conclude that there must be free causation, which it describes as “an **absolute** causal **spontaneity** beginning **from itself**” (CR 484, A446/B474). To begin with, then, in the terms I have been using so far, free causation is to be causation determined by or “from” itself. So a free will would be one that determines its own causality. Later, in the Remark on the Thesis, Kant emphasizes that freedom is “a faculty of beginning a series of successive things or events **from itself**” (CR 486, A448/B476), and the resolution of the Third Antinomy treats freedom in similar terms (CR 533ff, A533ff/B561ff). Kant is also clear that free causality is the basis for imputing actions to their free causes (CR 486, A448/B476; 544-545, A554-556/B582-584). So we can conclude from all this that, for Kant, freedom is the property of a will whereby it determines its own causality itself.¹⁵

Even construed in terms of self-determined causation, a free cause can still be responsive to external factors, of course. Kant’s claim is just that a free cause cannot be *determined* by such factors *alone*. I will come back to this point in more detail later on. Besides self-determination, though, Kant says many other things about freedom and free causation, as above when he called it “absolute” and “spontaneous”.¹⁶ He also calls it “unconditioned” (CR 486, A448/B476), and

that natural causation would be radically ungrounded without something that is a cause “in itself”, it still might not follow that radically ungrounded natural causation is really impossible. Perhaps the causation we actually find in the world just is radically ungrounded. The Third Antinomy is raising a conceptual problem about freedom and causation, and Kant is asking us to accept transcendental idealism as the solution to that problem. But even if we accept the terms of Kant’s problem, perhaps his solution comes at too high a price?

¹⁵ Once again, also see (G 52-53, 4:446-447).

¹⁶ While we are mostly going to elide these complications, Kant thinks that it is essential for free causation to be *absolutely* from itself, and not only *relatively* from itself. Kant thinks that all living things begin to act from themselves in at least a relative sense: as he puts it in the *Metaphysical Foundations of Natural Science*, “Life is the faculty of a *substance* to determine itself to act from an *internal principle*” (MF 83, 4:544). (Besides essentially following Aristotle’s *Physics* (192b) and Leibniz’ “Monadology” (1989, 214), note that the *Metaphysical Foundations* provides a broader explication of life than the one in the Second Critique or in the *Metaphysics of Morals* (CP 7n, 5:9n; MM 11, 6:211).) So, for example, Kant thinks that plants and non-rational animals all begin to act from themselves through inner principles. Even Leibniz’ monads would act from themselves, in this relative sense. For monads, that inner principle would be the the “law of the series” of their appetites and perceptions (Leibniz 1989, 214-216), while the inner principle for non-rational animals could be something like “brute” instinct, reflex, or operant conditioning. Kant thinks that activity of this kind is genuinely spontaneous, because it is self-started and self-determined, but only relative to its inner principle. That internal principle, Kant thinks, is itself still subject to the

there is a whole range of properties that he thinks it must have as a kind of non-natural causation. For example, it must count as intelligible, noumenal, non-spatial, and non-temporal (CR 536, A539-540/B567-568).

So here we really are on the verge of the deepest difficulties in understanding Kant's views on freedom. Thus, I cannot attempt to explain these notions in any further detail here. I hope it is at least generally clear, though, how Kant's more specific claims might derive from and be interpreted in the light of his idea that free causation determines itself. We can also begin to see here how the features that Kant expects of moral laws and moral obligations are already prefigured in this account of free causation, along with the relation of freedom to autonomy.

For now, though, the key point is that Kant still requires there to be laws for freedom, despite how free causation determines itself. As he puts it, "every effective cause must have a **character**, i.e., a law of its causality, without which it would not be a cause at all." (CR 536, A539/B567) So here we arrive again at the notion of a law of freedom, and we can sum up the previous discussion like so: a law of freedom is supposed to be a law for a will that is free, that is, a law of freedom is to be a law for a will that determines its own causality.

Kant also thinks that the will is the faculty of volition—of willing—and that actions follow from volition. So a law of freedom will then also be a law that determines the volition, and thereby also the actions, of a will that determines its own volition. This way of describing laws of

necessitation of sensible impulses and of nature more generally, in the case of plants and non-rational animals. And Leibniz' monads are still supposed to be subject to the wisdom of God's providence and preestablished harmony (1989, 223-224). Hence, such beings are self-determined, for Kant, but not *only* self-determined, since their inner principle is also determined by external causes. In Kant's sense, a being only counts as *absolutely* self-determined, or transcendently free, when its inner principle is also totally independent of natural causation, so that its inner principle is also fully or absolutely autonomous. Though, again, a being possessed of absolute self-determination can still be influenced by and responsive to external factors. Their actions can perhaps even be determined by external causes, too, so long as their inner principle itself is still absolutely self-determining as well. (Kant appears to think that everything in nature is determined by natural laws, so whether self-determination is absolute or relative seems as if it must have to do with the status of the inner principle alone.) Allison's discussion of Kant on relative spontaneity is quite helpful here (1990, 60-64). I think this analysis here also conveys a sense of the structure of Kant's thinking about freedom: determination from itself is supposed to be basic, and to guarantee genuine spontaneity, which can then be absolute or relative, and so on. Thanks to Sheldon Smith and John Carriero for pressing me to be clearer here.

freedom may sound paradoxical, but remember that the law is to be the character of a free will's self-determined causality. A free will is supposed to determine its own causality through or according to the laws of this causality, we might say, and this is all that I will mean in saying that laws of freedom determine free volition and free action.

2.2 Causation in General

Now, granted that laws of freedom turn out to be causal laws, for Kant, what does this tell us about his views on their basic structure? What are their component parts supposed to be, and how are those parts to be related to one another?

Since Kant wants laws of freedom to be causal laws, and since laws are to be the “character” of causation, as I just noted, if we want to know what the structure of a law of freedom has to be for Kant, we must look to his account of causation in general. Of course, Kant's views on causation are tremendously complicated and difficult, as should be abundantly clear already. But, fortunately, as I indicated at the outset of the chapter, there is still a level of generality where the picture is quite clear.

For example, we can abstract away from questions about Kant's specific model of causation, like the questions about whether events can be causes, for him. So we can stay neutral with regard to more specific accounts like Watkins' (2005). Instead, we can start with the highest level of generality at which Kant discusses causation as such. Namely: at the level where causation is to be the object of a “pure concept of the understanding”.

Kant also calls these concepts “categories”, and organizes them into the famous “Table of Categories” (CR 212, A80/B106). These categories are supposed to be derived immediately from the more basic forms or functions of judgments, of which there are twelve, according to Kant's “general” logic (CR 211-212, A79-80/B104-105).¹⁷ For example, the category of substance is supposed to be an instance of the logical concept of a subject of predication.

¹⁷ Kant later calls this derivation the “**metaphysical deduction**” of the categories, in contrast to the “**transcendental deduction**” thereof, which is supposed to be an *a priori* proof of the objective validity

So there can be no higher level of generality than this one, for Kant: for him, the category of causation must be the original, fundamental concept thereof, because he thinks that the only more general concept is no longer a concept of causation at all, but of a logical relation, instead.

Specifically, Kant thinks that there are three basic logical relations between cognitions, and the category of causation is supposed to be derived immediately from one of them: from the “hypothetical”, “ground”/“consequence”, “if”/“then” logical form of judgment (CR 206, A70/B95; 208, A98/B98-99; P 63, 4:312). This derivation is an integral part of Kant’s overall transcendental idealism, of course, since it is crucial to the arguments that are supposed to show how our logical categories largely constitute and determine the form of objects in nature, such as their general causal structure. So Kant is making a hugely important and controversial step here. But we need to grant it to him, for the sake of interpreting his subsequent views.

When we do grant Kant’s derivation, though, causation turns out to be an instance of a logical relation. Therefore, causation is fundamentally a relation, for Kant—this is the fundamental fact about causation, in Kant’s account. And the overarching hypothetical form of judgment is to relate an antecedent “ground” to a consequent “consequence”, as when, for example, the truth of a proposition implies or guarantees the truth of its logical consequences, or when satisfying one concept implies satisfying another.

So if causation is to be an instance of this specific kind of logical relation, it will have to relate a cause, as a ground, to an effect, as a consequence of that ground.¹⁸ As Kant puts it later on in the First Critique, “The schema of the cause and of the causality of a thing in general is the real upon which, whenever it is posited, something else always follows.” (CR 275, A144/B183) If

of the categories (CR 261, B159). However, given Kant’s famous explanation of what a “deduction” is supposed to be, the Metaphysical Deduction should not count as a deduction, but rather a derivation (CR 219-220, A84-85/B116-117). Although Kant might count the Metaphysical Deduction as a deduction because of how it contributes to the Transcendental Deduction.

¹⁸ In other words, when Kant says that the three categories “Of Relation” are “Of Inherence and Subsistence”, “Of Causality and Dependence”, and “Of Community” (CR 212, A80/B106), he means to say that objects can relate to each other in one of three ways: as substances relate to their accidents, as causes relate to their effects, or in a reciprocal relation of activity and passivity, in the same way that he thinks cognitions can be related in categorical, hypothetical, or disjunctive forms (CR 206, A75/B95).

this is the schema of “the cause”, then the schema of causation, in contrast, will be the relation between the cause thus schematized and its effect: the “something else” that “always follows” whenever the cause “is posited”.¹⁹

So according to Kant, causation is, at its most fundamental level, a relation. He then thinks it follows that causation in general must have three, and only three, basic components: a cause, an effect, and the relation connecting the cause to the effect, which is what the category of causation properly represents, for him. For, after all, a relation requires *relata*, and a two-place relation requires exactly two basic *relata*. Then, as instances of the *relata* of a hypothetical ground/consequence relation, the cause will constitute a ground, and the effect will constitute its consequence.

But granted that causation is an instance of the logical relation between grounds and consequences, what kind of instance is it, what kind of relation? Once we depart from the domain of general logic, Kant thinks that only one answer is possible: causation must be a real instance of this hypothetical logical relation. And, in fact, we see this reflected in the schema of causation, wherein causation consists in “the real upon which, whenever it is posited, something else always follows.” (CR 275, A144/B183)

Thus, once we move from the purely logical relation of ground and consequence to a real relation between cause and effect, Kant thinks that this real relation—the causal relation itself, which connects causes and effects—will have to be the action by virtue of which the cause causes its effect. And Kant sometimes calls this action “the causality of the cause” (CR 533, A533/B561). For he thinks that the logical relation itself must be “realized”, too, in order to establish the real connection between cause and effect. And then the real instance of the logical relation will have to be the action of the cause, whereby it produces the effect. In a real relation, there must be

¹⁹ I will return to Kant’s notion of a schema/schematism of a pure concept of the understanding presently. Notice, though, that this schema already makes causation a necessary relation. If the positing of the cause means the effect always follows, there is no room left for possible exceptions to this relation.

something real connecting the *relata*, he thinks, and in causation this real connection must be the action of the cause producing the effect.

So Kant thinks that causation in general is the real instance of the hypothetical logical relation. And, as a result, the three general components of causal relations, for him, must be a cause, an effect, and the action through which the cause produces the effect.

But what are causal relations relating, exactly? In treating causation as a real instance of a logical relation, Kant's fundamental point is that causation relates existences: the existence of an action and an effect are to follow from the existence of their cause. And, in fact, Kant characteristically describes causation in general as the relation between the existence of a cause, an action, and an effect—he even pointedly refuses to say anything about how causes in general produce these consequent existences. Besides the schema of causation, see (P 62, 4:310), (G 52, 4:446), and (CP 45, 5:50-51), for example.

Crucially, at the level of causation in general, Kant does not even require the action or effect to be at subsequent moments in time, as per (CR 312, A202-203/B247-249), as well as the First Critique's Third Analogy of Experience (CR 316-319, A211-215/B256-262). Nor does Kant's general notion of causation even say anything about temporal order, even within an instant. Temporal relations are to come in with natural causation, not with causation in general.

Consider once more the schematism of the category of causation. Kant's notion of a schema here is technical and tricky, but, suffice to say, it is the most general way in which a pure concept of the understanding, such as the category of causation, relates to our faculty of sensibility, and thereby to appearances and to things in nature (CR 271-272, A137-138/B176-177). In short, the point of the schematism is “to show the possibility of applying **pure concepts of the understanding** to appearances in general” (CR 272, A138/B177), and the schematism does this by relating these concepts to our senses in general.

Roughly speaking, this is all a technical way of saying that the schematism is supposed to explain the most general way in which the categories apply to the objects of our senses, and Kant

thinks this task is to be accomplished by showing how these concepts relate to our senses in general. Then, since Kant thinks that everything we sense is in time, but not necessarily in space, he argues that the pure concepts of the understanding must be “schematized” by connecting these “pure concepts of the understanding” to time (CR 272, A138-139/B177-178).

Accordingly, the schema of causation teaches us two things about Kant’s views on causation in general. First, we learn that causation in general, as an unschematized category, can say nothing about temporal ordering, whether in a moment of time or across moments of time. Instead, only the schematized category of causation is to say anything about the temporal ordering of causes, actions, and effects. The pure concept of causation is supposed to represent a certain relation, but it is to be an open question how that representation gets reflected in time. That question can only be answered with a schema for causation, according to Kant.

Nevertheless, in seeking to relate the concept of causation in general to our faculty of sensibility in general, the schema of causation must reflect what Kant takes to be fundamental to causation in general. For all the schematism does, according to Kant, is take the concept of causation in general and apply it to time, in order to show how that concept is to apply to sensibility in general. So if we separate out the “temporal content”, so to speak, from Kant’s schematized concept of causation, we are left with his isolated concept of causation in general.

With that in mind, here again is how Kant presents the schematism of the category of causation: “The schema of the cause and of the causality of a thing in general is the real upon which, whenever it is posited, something else always follows. It therefore consists in the succession of the manifold insofar as it is subject to a rule.” (CR 275, A144/B183) So no matter what Kant means by “manifold” here, causation in general, for him, is to be nothing more than a succession of existences according to a rule. He thinks that we get a corresponding temporal succession only once the concept of causation in general becomes schematized to apply to objects in time, so it will follow that his concept of causation in general only relates existences in abstraction from temporal relations.

In fact, then, this bare relation in existences is to be Kant's basic, most general account of causation. And in order to apply this account of causation to Kant's views on free volition and laws of freedom, it will be important for causation in general to relate only the existence of an action and an effect to the existence of a cause, if only to stay totally neutral with regard to specific accounts of causation. So it will be worth elaborating a little further.

On these points, it is important that Kant considers the concept of causation to be a "**dynamical**" category (CR 215, B110), and the fundamental causal principles of the Second and Third Analogies of Experience to be "**dynamical** principles" (CR 286, A162/B201). In both cases, Kant intends this description to convey a connection to the existence of objects: first, he explains the nature of "dynamical" categories by observing that the Table of Categories

can first be split into two divisions, the first of which is concerned with objects of intuition (pure as well as empirical), the second of which, however, is directed at the existence of these objects (either in relation to each other or to the understanding). I will call the first class the **mathematical** categories, the second, the **dynamical** ones. (CR 215, B110)

And to explain "dynamical" principles, he appends a footnote speaking of a combination of objects that, "since it is not arbitrary, I call **dynamical**, since it concerns the combination of the **existence** of the manifold" of appearances (CR 286n, B201n).

Moreover, Kant intends this distinction between mathematical and dynamical representations to yield a clear conceptual distinction between causal ordering and temporal ordering: as the Remark On the Antithesis of the Third Antinomy illustrates, "**If you do not assume anything mathematically first in the world as far as time is concerned, then it is also not necessary for you to seek for something dynamically first as far as causality is concerned.**" (CR 487, A449/B477) Whether or not Kant accepts this claim is not to the point: what matters is that he distinguishes mathematical temporal ordering from dynamical causal ordering.

As I have been saying, though, Kant does believe that simultaneous events can have temporal ordering as well; see again (CR 312, A202-203/B247-249). And he thinks that causal

relations in nature will always have a temporal order and a temporal succession, even in an instant of time. For the concept of a purely dynamical succession in existences, when schematized to apply to temporal nature, will have to imply a temporal order as well, even within an instant. That is just part of how Kant thinks the schematism is to work, as we have explained it. For Kant, however, all temporal ordering depends on an underlying dynamical order: see the Second and Third Analogies of Experience, for example.

In any case, though, we come around again to the idea that *in themselves* causal relations *in general* are only supposed to involve the purely dynamical relations between the existence of a cause and the consequent existence of an action and effect, even if Kant also thinks that dynamical relations in nature must have a temporal order. In fact, Kant even sometimes expresses skepticism of our ability to form any understanding of causal mechanisms beyond the way existences are related to one another (P 62-63, 4:310-311). Regardless, though, I have been emphasizing these points in order to rely on only the most minimal account of causation in general that we can find in Kant, even if it does not turn out to exhaust his views on this topic, so as to remain as neutral as possible as to the finer details of this extremely difficult subject.

One last point is worth emphasizing, though I have indicated it already: for Kant, it is of the essence that causal relations are necessary relations. Kant's view on this point is so fundamental and obvious as to hardly need illustrating, but, again, it is very perspicuous in the schematism of causation: "The schema of the cause and of the causality of a thing in general is the real upon which, *whenever* it is posited, something else *always* follows. It therefore consists in the succession of the manifold insofar as it is *subject to a rule*." (CR 275, A144/B183, emphasis added) And see again (G 52, 4:446) and (CP 45, 5:50-51), for example, as well as (CR 265, B168) and (P 7, 4:257). In fact, though, the necessity in causal relations ought to be expected, given how Kant seeks to derive causation from the form of hypothetical judgments. For these judgments all say: if something P holds, then something Q must follow.

So here we have it, when we draw the prior discussion together, the general form of a causal law, according to Kant: the existence of a cause makes the existence of an action and an effect necessary (relative to or contingent upon the existence of the cause). And to be precise, the law, as the “**character**” (CR 536, A539/567) of the cause, is to consist in nothing more than this sort of relation, between these kinds of elements.

We can illustrate this general form quite clearly in the paradigmatic case of a causal law for Kant: Newtonian Universal Gravitation. As Kant puts it in an official “Explication” in the *Metaphysical Foundations of Natural Science*: “*Attractive force* is that moving force by which a matter can be the cause of the approach of others to it” (MF 35, 4:498). So here we have a cause, an action, and an effect: a material body has a certain mass, so it exerts a certain quantity of “moving” force on another material body, and this force is applied to that second body, which may or may not subsequently alter its motion, depending on the state of the rest of the physical system. The quantity of this action is supposed to be the output of Newton’s equation, which is also to capture the necessity in this causal relation: the amount of force that is always supposed to follow from the existence of the first body related to the existence of the second body.²⁰

²⁰ The Law of Universal Gravitation needs to be supplemented if it is to be of any use in determining how the application of gravitational force actually alters the motion of any bodies. For that, we need more information about the system in question. Are there any other bodies or forces? Are the bodies at rest or already in motion? And so on. Also, another interpretation of Newton’s Law would be that the force is the cause, the action is the application of the force, and the effect is however that force actually influences the motion of the system in question. Which way we go doesn’t matter for our purposes, and I suspect Kant doesn’t care too much, either. After all, his view is that material bodies and the forces that they exert on each other are composed out of the attractive and repulsive forces that constitute matter itself. So there is very little material difference between material bodies as causes and forces as causes (MF 48-49, 4:510-511). The point now, though, is just that both interpretations require the same three protagonists, related in the same way. Some argue that we do not need to read Newton’s Law as committed to forces as actions, or even to causes at all. The idea is that Newton’s Second Law, $F = ma$, allows us to replace any references to F ’s with references to ma ’s instead, so that physics only ends up correlating masses and motions. Newton, however, viewed gravitational force as a cause, namely, of an “endeavor to approach”, though he famously refused to “feign hypotheses” about the ultimate source of this cause (Newton 1999, Bk 1, Section 11; General Scholium). And it is well known that the idea of force as an action is necessary to resolve certain questions about the absolute motion that Newtonian Physics postulates. Suppose a car is accelerating relative to the road. Isn’t the road also accelerating relative to the car? So how can we attribute absolute motion to one rather than the other, as Newton wishes? The answer is that the car’s engine exerts a force on its wheels, changing the state of the car’s motion, as anyone can feel when they step on the accelerator. So $F = ma$ seems best read not as allowing F ’s to reduce to ma ’s, but as equating the quantity of force exerted with the quantity of acceleration it produces in a body of a certain mass. As

And, indeed, it is rather hard to see how any account of causation could avoid giving it some version of this general form. Even Hume's skepticism about real causation requires an account of how causes relate to their effects, only for him this relation boils down to a statistical/historical relation (the constant conjunction of events of certain types), a spatio-temporal relation (the contiguity of events and event types in space and time), and a psychological relation (the strong expectation that events of one type will follow events of another type). The relation is even supposed to be necessary, albeit in a merely subjective, psychological way (Hume 1978, 155-156).

Moreover, in these terms, the difference between free and natural causation amounts to this, for Kant: the action of a natural cause is to presuppose the action of a temporally prior natural cause, while a free cause is to be sufficient for its action by itself. Which helps us to see why Kant would consider free causation to be "absolute", "spontaneous", and "unconditioned".

2.3 The Structure of Freedom and Necessity

According to Kant, then, causal laws in general must consist, for their component parts, in a cause, an effect, and the action of the cause whereby it produces its effect. Moreover, Kant thinks that the relation between these terms, which ties them together into a unified whole, and which the law itself properly consists in, must be a necessary relation: given the existence of the cause, the existence of the action and the effect must follow. What does this tell us about the structure of laws of freedom, for Kant, and thereby about the structure of his practical principles?

We won't get the full argument for my interpretation of Kant's practical principles here in this chapter; I will present further arguments and deal with some problems in the next chapter. So I will leave it to my readers to decide how far these considerations about Kant's views on causation will go by themselves.

Newton himself does; again, see (Newton 1999, Bk 1, Section 11). I owe this example to Sheldon Smith. Moreover, the worry that gravity is a mysterious "action at a distance" obviously presupposes that the theory treats it as an action.

However, since free causality is supposed to be absolutely self-determined, in Kant's view, the proper or ultimate cause of every free action will be its agent: remember that free causality, as absolutely self-determined, is also to be the ultimate basis for imputing an action to its agent (CR 486, A448/B476; 544-545, A554-556/B582-584). More specifically, though, since Kant considers freedom to be a property of an agent's will, the agent will be the ultimate or proper cause of their actions, but by virtue of their freely choosing how to act. Hence, laws of freedom will be a kind of law for how free agents choose their actions. Thus, they can also be described as a kind of practical law, which would be expressed and represented in practical principles, as I have proposed.²¹

However, this analysis makes it sound as if every law of freedom should ultimately contain the same cause: if a free agent is always the proper, ultimate cause of their actions, then it looks as if every law of freedom would ultimately have to be of the form "a free agent chooses to act a certain way to achieve a certain effect". But then it looks as if there is nothing that we can do to explain or to justify why a free agent chooses to act a certain way, or why free agents should make certain choices, either. Free choices seem to have to be ultimately arbitrary and inexplicable, if laws of freedom have to have this form. But then the concept of a "law" of freedom seems to collapse in on itself. Free agency could still work this way, only it would be entirely lawless.

So the crucial point here, especially for Kant, is to explain how free choices need not be random, arbitrary, or ungrounded in anything beyond the agent and their act of choosing. Instead, Kant's view is that every free choice must still be based or grounded in something besides the bare act of choice, especially if a free choice is to count as a rational, too. And this

²¹ For Kant, all practical laws are laws of freedom, and vice versa. As he puts it, "A faculty of choice, that is, is merely **animal** (*arbitrium brutum*) which cannot be determined other than through sensible impulses, i.e., **pathologically**. However, one which can be determined independently of sensory impulses, thus through motives that can only be represented by reason, is called **free choice** (*arbitrium liberum*), and everything that is connected with this, whether as ground or consequence, is called **practical**." (CR 675, A802/B830) Hence moral laws, as laws of freedom, are also practical laws. It does seem clear, however, that there are practical laws that are not moral laws: those expressed in hypothetical imperatives, paradigmatically, but perhaps there are others also. More on this presently.

ground must actually be a ground, in Kant's technical sense: to explain the possibility of lawful freedom, Kant thinks that it is not enough to grant only a sufficient permission to act a certain way, for example. For a choice on the basis of a sufficient permission could still be ultimately arbitrary or ungrounded. A sufficient permission relates to an action only negatively, by removing barriers to performing it. In its own right, a permission, even if sufficient, does not speak at all to whether or not the action would or should be performed, positively speaking. So if the permitted action is performed, it could still be arbitrary or groundless nevertheless. So, Kant thinks, a non-arbitrary action must be necessitated by a ground, even if that ground is "just" a feeling, an inclination, or a general conception of one's happiness.

In fact, for Kant, the need for a ground to necessitate free choice is almost the entire point of speaking of "laws" of freedom in the first place, the contrast being, naturally, the *lawless* freedom to which we already alluded (G 52 4:446). Treating the will as the faculty of practical reason is help fulfill this need, also (G 24, 4:412; CP 48, 5:55; MM 13, 6:213).

These points will come up again, perhaps more fully, when I discuss Kant's "Incorporation Thesis" (R 48-49, 6:23-24; R 47n, 6:21n) in the next chapter. But the upshot is that free choice must still have a ground, according to Kant, from which a certain action and effect must follow. The ground will require an effect as its consequence, and it will require the agent to act in order to bring about that effect, even if all that the effect requires is an act of volition alone.

For example, at the outset of this chapter, I indicated that Kant's maxim of the lying promise would be the paradigmatic illustration of the structure that I am proposing: "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen." (G 32, 4:422) The ground here is the belief that you need money, since the agent is thinking that they "shall borrow money and promise to repay it" "when" they have this belief. And the intended effect of this false promise, of course, is to obtain a loan, in order to remediate the supposed need of money. Last, practical necessity comes into

the maxim in two ways: with the thought that this action, for this effect, follows from this ground, and also with the thought that the agent must act this way, to achieve this effect, when this ground obtains, or else they fail to satisfy the principle.

Now, Kant thinks that this ground, just by itself, will not be sufficient to determine whether or not its action will actually take place, even when the ground obtains. After all, the agent is to be free, and thus absolutely self-determining. So the agent, through a free act of their will, must choose to act according to this ground.²² Thus, for Kant, a free agent's choice, as free, must be absolutely self-determined, but also determined according to a necessitating ground, to be non-arbitrary. So a free agent must also freely choose the ground for their actions.

These points are tricky enough by themselves, to say the least. One immediate problem is that, under my interpretation of Kant, all free actions must still be necessitated by a ground. However, that necessity might seem to be quite hard to locate in the vast majority of our actions, to say the very least. When I decide what size coffee to order, or what to do with my Friday evening, that choice is surely free, if any is. But is that choice really necessitated as well, and necessitated in Kant's specific technical sense? Where is the ground that *requires* me to make this choice?

First, notice that there is a difference between a *general* and a *specific* necessity in action. Granted that I have decided to order coffee, and granted that I must do something on my Friday evening, I am under a general necessity to choose the size and to choose what to do. For all that, though, I might not be aware of anything that would require me to choose *this* size or to choose to do *this* with my evening. Perhaps Kant would require me to find some consideration that would lead me to a more specific necessity. For example, perhaps I want more than a small coffee, but I'm concerned that a large will make me too jittery. Combining all of these considerations into a single ground could then yield a more specific necessity to choose a medium size. (Assuming I judge that to satisfy all my coffee-related considerations—for Kant, no

²² The Incorporation Thesis is a slightly more complicated and technical version of just this point.

principle or ground is ever specific enough to determine action without an intervening judgment. This point will be crucial in Chapter 4, and hence discussed more fully there.) Or perhaps Kant would allow me to “just choose” at some point—which way he would go does not matter at the moment. What matters is that, either way, my choice can still be traced back to some form of practical necessity, and to some kind of ground.

Moreover, though, there is also a difference between unconditional and conditional necessities. As free, I can always change my mind about whether to order coffee, or about what to do on a Friday evening. Therefore, I can always release myself from even the general necessity to choose a size for my coffee, just by changing my mind about whether to have coffee in the first place. Or if I’m having trouble choosing what film to see, I can always decide to stay at home and read instead. The case is quite different, though when it comes to keeping my promises, for example.

When it comes to the ground/action/effect structure in a law of freedom, the latter distinction is captured in the kind of ground and in the kind of necessity, while the former is captured in how the effect gets specified. Is the ground and the necessity it imposes conditional or unconditional? Is the required effect acquiring a coffee of some size or is it acquiring a medium coffee?

In any case, these points will become a little clearer presently, when we take a look at Kant’s views on happiness and on hypothetical necessities. The discussion of the Incorporation Thesis in the next chapter will also help flesh out the details a little further, as well. The basic point, though, is that some practical necessities are specific and unconditional, like the necessity in saving a drowning child, while other necessities are general and conditional, and thus often quite highly attenuated or distant from the point of action, like the necessity in the current examples. (There can also be specific, conditional necessities and general, unconditional necessities—these distinctions are orthogonal.) Necessities that are specific or unconditional are often quite easy to recognize, and often easy to feel, even, while we can easily conclude that there

is no necessity at all in cases where the necessity is actually there, but general or conditional. However, as long as there is something that we are trying to satisfy in our action—our craving for coffee, our desire to relax on a Friday night—there is some ground for our choice, and some form of practical necessity.

Now, explaining how an action or a choice can be free, in the sense of absolutely self-determined, but also necessitated by a ground—that is surely no easy task. These concepts appear incompatible: as free, the agent is supposed to be the cause of the action. But if the action is not to be arbitrary, Kant thinks it has to be grounded by something besides just the agent. So wouldn't that ground have to be the cause, too? How can the cause be both the free agent and the necessitating ground?

The idea that a free agent must choose their grounds seems promising, but the apparent incompatibility of freedom and necessity also seems to recur here. For example, in the previous chapter, I mentioned how it seems impossible to choose necessities for yourself: if I choose a ground for myself, how can that ground really make it necessary for me to choose to act a certain way? If I am free, in Kant's sense, can't I change my mind at any moment, and release myself from that necessity? So, again, it seems like we must choose between genuine freedom and genuine necessity, at least if we accept Kant's theoretical goals.

Moreover, besides the apparent incompatibility of these concepts, it also seems as if there is a regress looming here: if a free agent has to choose the grounds of their actions, but their choices also have to be grounded, in order not to be arbitrary, then it seems as if they would need a ground for their choice of grounds. But that higher-order ground would have to be freely chosen, too. And then they would need a ground for choosing that ground, which would also have to be freely chosen...and so on, *ad infinitum*.

So Kant's position here is quite difficult to elaborate, and his full story will only come out over the course of the whole study here. For now, though, I can at least introduce the account of practical principles that I want to give, and then outline part of the broader answers to these

questions that my account affords. As I said before, this thread from Kant's thinking on causation is one of the main lines supporting my overall interpretation of practical principles, but I will adumbrate some others, look at the specific texts on practical principles, and deal with some possible problems in the next chapter, on the way to a fuller explanation of how Kant tries to reconcile freedom and necessity.

To begin with, then, the necessity in free choices is ultimately supposed to be an instance of the necessity in the underlying, original form of hypothetical judgments. If laws of freedom must ultimately exhibit this form, as per the Metaphysical Deduction, then each candidate law of freedom will have to propose a relation between a ground, an action, and an effect, and it will have to propose that this ground has this action and this effect as its necessary consequences. In short, a candidate law of freedom will have to propose that some free agent, or set of free agents, will have to act a certain way whenever a certain ground obtains, or else they will misuse their freedom, by violating a law of freedom. That all just follows from the account so far.

So if practical principles represent laws of freedom, these principles will say, to a free agent, that, as a free agent, they have to act a certain way, to achieve a certain effect, whenever a certain ground obtains. And a free agent, as free, must then decide whether or not this principle represents a genuine law for them.

In Kant's terminology, this means that free agents must decide whether or not this principle is *valid* for them, or whether or not it *holds for them*, if they are *subject* to it by falling *under* its requirements.²³ For Kant, then, to adopt a principle is just to judge or to come to believe that one really must act as the principle in question specifies whenever its ground obtains. To judge, for example, that you really must tell a lie in order to get a loan whenever you believe that you are in need of money. After all, the hypothetical form of judgment is to consist in a logical relation: the logical relation between grounds and consequences. So, since practical

²³ See Kant's official explications of maxims and practical laws, quoted in Chapter 3.3 below, for some examples. Strictly speaking, concepts like validity only apply to representations. But we can, and Kant does, extend its usage to the laws, necessities, grounds, and ends that are represented by a valid principle.

principles are to ultimately derive from this logical form, these principles will ultimately have to contend that a relation of this kind actually exists. And so, to adopt or accept a practical principle will just be to agree that there really is such a relation.

Kant, I take it, makes a similar point when he speaks of practical principles and practical necessities that are “problematic” (G 26, 4:414-415).²⁴ For he thinks that practical principles present a problem or pose a question for free agents: as a free agent, do I really have to act this way, to achieve this effect, whenever this ground obtains?

And Kant argues that this question can be answered affirmatively in one of two basic ways. First, I might already accept an end, which commits me to act as the principles requires. In that case, given that I accept this end, I really must act according to this principle.

Now, there is surely some sense in which rational beings already accept any principle or end that they are committed to, no matter where this commitment comes from. But that sense is not the sense at issue here: as imperfectly free and imperfectly rational agents, Kant thinks that there is always a question about whether or not we will accept or live up to our commitments. As rational, I might already accept all of my commitments. But, as imperfectly rational, I can just as much reject my commitments, too, even if I know that I am committed to them. Imperfectly rational agents are prone to contradicting themselves; that is part of their imperfect rationality.

So in speaking of an agent’s “actual” ends and principles, I mean to refer to the ones that they actually accept in this stronger sense, where commitment is not enough, by itself. In this sense, an agent must act to adopt an end or a principle, even if they are already committed to it. For the commitment at issue is just a commitment to adopt an end or principle, and these can always come apart, in imperfectly rational beings. I take it that I am following Kant here.

²⁴ Strictly speaking, the *Groundwork* says that problematic principles are a type of hypothetical imperative, so I am broadening his concept a little: by the above lights, a principle that purports to represent a categorical law could be “problematic”, if it is unclear whether the principle in question is valid. But it makes no difference whether this extension is granted to me; I am happy to use the notion simply to illustrate the point about practical principles.

In any case, suppose that I accept Kant's maxim of the lying promise, or even just suppose that I am tempted by this principle. Do I really have to act this way, though? If I have my own happiness as an end, as I surely do, and if I conceive of my happiness in such a way that it requires me to tell lies to get loans when I think I need money, then an end that I already accept would require me to accept the maxim of the lying promise, too. I might be wrong about my ends: my conception of my own happiness could be wrong, or it might not actually commit me to the maxim of the lying promise. And there might be deeper, overriding moral considerations preventing me from having to act this way, all things considered. Moreover, I could still fail to fulfill this commitment, perhaps through *akrasia* or another form of irrationality. But relative to the end that I've already adopted, I find myself committed to this principle, too.

Second, I might have to follow a principle because the nature of the principle itself requires me to, regardless of whether or not I actually do accept any of it. Kant thinks that this kind of requirement is possible in three different ways: first, the principle could state that an action is necessary as a means to some end, which an agent might or might not have. If such principles are correct, then they state how anyone and everyone must act, in order to achieve that end.²⁵ Second, we could be committed to a principle through a "natural necessity", such as,

²⁵ Kant says that such principles yield "hypothetical imperatives" (G 25, 4:414), which I will discuss more fully in Chapter 5. Attentive readers might have noticed that I am leading up to Kant's distinction between subjective and objective principles, and that I am treating the principles that yield hypothetical imperatives as objective. Contrary to initial appearances, perhaps, Kant is clear that all imperatives are objective, even hypothetical ones (G 24, 4:413). Kant thinks that hypothetical imperatives say that something must be done, in order to achieve a certain end. However, that conditional requirement is to be objectively binding, even though agents do not have to perform that specific action unless they accept the specified end. For Kant, that makes hypothetical imperatives objective principles, but with an embedded subjective condition. To borrow an example from Kant (G 28, 4:417): "To bisect a line, make two intersecting arcs from its ends" is based on a fact about Euclidean geometry. So this principle is to be valid for everyone, and thus valid objectively. But "Make two intersecting arcs from the ends of this line" will only be valid for those agents who are committed to bisecting this line. If that commitment comes only from their choice of ends, then they will be under an objective necessity, but only relative to their subjective choice of ends. For Kant, this objectivity is evident in the kind of mistake that agents make when they fail to fulfill the hypothetical imperative here: they make a mistake not just relative to their choice of ends, but also relative to Euclidean geometry. So the necessity can be ultimately subjective, but it is still to be objective, whenever it is in force. (Some readers of Kant think that he does not allow

paradigmatically, through how the principle is connected to our own happiness. This idea is puzzling, to be sure, but I will say more about it momentarily. Third, a principle could be valid for me because rejecting it would commit me to a contradiction of some kind, independently of what I actually will and why I actually will it. “Reason itself” requires me to accept the principle, we might say. I won’t say much more about this now, since explaining Kant’s thinking here will be the work of the rest of the study here.

Now, when I am committed to a practical principle because of an end that I already accept, Kant says that the principle is “assertoric”, and he says that principles are “problematic” when they represent an action as necessary for an end that we might or might not have (G 26, 4:414-415).²⁶ By contrast, if reason itself commits me to a principle, he calls that principle “apodictic”, since it is “of itself objectively necessary” (G 26, 4:414-415).²⁷ The difference between apodictic and non-apodictic principles is also supposed to be the source of the difference between hypothetical and categorical imperatives (G 26, 4:414-415).²⁸

Anyway, to recap, so as to elaborate a little: I started out this section just trying to make sense of Kant’s idea of a law for a will whose causes “are not so determining” (CR 534, A534/B562) as to exclude free choice, but where those causes can also be something besides just the agent and their free choices. The first step was to point out how, for Kant, a free agent would then have to choose which laws to accept, which, according to the interpretation of Kant that I have been developing here, is just to say that they must choose which grounds and which practical necessities to act under. So if some of these grounds are valid for me, and I freely choose to follow them, then won’t Kant be able to say that my choice is both free and grounded?

principles to be both subjective and objective. I disagree; this point will come up a fair bit in the next chapter. For now, just notice that there is no contradiction in saying that an agent is committed to a principle both through their choice of ends and through the nature of the principle itself.)

²⁶ See note 24 above.

²⁷ See the two explanations of categorical imperatives on (G 25, 4:414). ‘Apodictic’ connotes proof by demonstration.

²⁸ See note 25 above.

But then a puzzle arises: what are the standards supposed to be for a free agent's choice of grounds? If practical principles cannot be right or wrong about the practical necessities that they represent, then there would not be any genuine practical necessities, and our free choices would be arbitrary. Freedom would be lawless. So what would make a principle valid for a free agent, such that they would make a mistake—misuse their freedom—if they did not adopt that principle?

But practical grounds, by their very concept, are supposed to determine how a free agent has to act. So it is hard to see how these grounds could be freely chosen, in such a way that this choice could also be right or wrong. The paradox of voluntarism looms here, and an infinite regress does, too. First, a freely chosen ground seems as if it cannot really necessitate, because another free choice could give up or reject the ground at any time. And a free agent also needs a ground for their choice of grounds, or else their choice would still be arbitrary and lawless, by parity of the original reasoning. But their ground for choosing grounds would have to be freely chosen also, or else their choices stop being free. And then this higher-order ground for choosing grounds would have to be chosen according to a ground in turn, and so on.

So I started to thread these needles on Kant's behalf by noting how he thinks that we can refer the agent's free choice of grounds, necessities, and laws back to an end that they already accept, which would then require them to accept certain principles in turn. They could always relinquish or reconceive their ends, to be sure, although Kant does think that

There is, however, *one* end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely *could* have but that we can safely presuppose they all actually *do have* by a natural necessity, and that purpose is *happiness*. (G 26, 4:415)²⁹

²⁹ How a "natural necessity" can give ends to a free being is a rather large puzzle, to say the least, which I cannot take up here. It is not clear that Kant means "natural" in the sense of appearances, however, since he says this end belongs to our "essence", where he may even mean our essence not as humans, but as "dependent beings".

So Kant thinks that happiness, under some interpretation of it, can always be the basis for principles that are valid for free agents, at least of the dependent variety, and at least if they are under no overriding moral requirements or imperatives.

Still, Kant thinks that principles that are only valid for us through happiness, that we are committed to only through happiness, are only assertoric, and not apodictic, even if happiness is one of our ends through “a natural necessity”. For a natural necessity is not necessarily a rational necessity, for Kant, and he thinks that reason itself is indifferent as to whether or not we have happiness as an end. That is evident from how Kant thinks that morality, and hence reason itself, never requires us to pursue our own happiness for its own sake, but only, occasionally, through how unhappiness is a temptation to vice (G 12, 4:399).

So if we already have happiness as an end, then certain things must certainly follow by reason alone. But if we did not already have happiness as an end, we would not necessarily contradict ourselves either by adopting it or by refusing it. Only principles that are rationally necessary in this specific sense are to be apodictic, though, which is another way Kant allows us to explain how a free agent can be committed to a practical principle, despite having to choose the grounds of their actions.

So when it comes to the validity of a practical principle and of the practical necessity that it represents, Kant thinks that there must either be some sense, natural or rational, in which the agent *must necessarily* accept this relation as a basis for their free choice, or there must be some sense in which *they already do* accept this relation. (And recall that imperfectly free and imperfectly rational agents are prone to rejecting their commitments, both intentionally and by accident.)

In the former case, the principle of freedom is to be objectively valid, in the second case, subjectively valid. In turn, these subjectively valid principles can come in two kinds: either the agent already accepts the relation immediately, or they accept some other relation, which

requires them to accept this one as well, though, rationally speaking, they could just as well reject the relation in either case.

So a subjectively valid principle is to represent an action as necessary for a purpose an agent actually has, while an objectively valid principle is to tell an agent what they must do. In these terms, some objective laws of freedom are also subjective, when agents actually accept them, too. Moreover, Kant thinks that some objective principles are apodictic and some are still assertoric—because he thinks that some principles of freedom are objectively necessary, but not “of themselves”. Happiness, for example, because it is an end that we all must have, but because of a natural necessity, not an intrinsic rational necessity. Hence, happiness is an end that we both must have and which we always actually do have, so that the basic principle of pursuing happiness is to be both objectively and subjectively valid for us, though not valid apodictically.³⁰

How any principle of happiness can be objectively valid for us, according to Kant, is both a philosophical and a textual puzzle. Our imperfect rationality, however, is supposed to mean that we are persistently capable, even intentionally capable, of acting against our true happiness, and even of acting against our own conception of happiness (G 12, 4:399). We can get happiness wrong in a number of ways, and we can fail to pursue it as we ought to do, too.

As such, Kant himself recognizes that the prudent pursuit of happiness could yield “objective **laws of freedom**” (CR 675, A802/B830), at least in principle, even if we can never actually possess such objective laws in practice (G 28-29, 4:417-419). If human happiness were as simple a matter as bisecting lines in Euclidean geometry (G 28, 4:417), then principles of happiness would surely be objective, Kant thinks. As he puts it, “If only it were as easy to give a determinate concept of happiness, imperatives of happiness would agree entirely with those of skill and would be just as analytic” (G 28, 4:417), in the sense where sufficient analysis of your end tells you what you must, objectively, do to attain it. If “it is supposed that the means to

³⁰ Earlier, I mentioned how morality and reason do sometimes require us to pursue our own happiness, albeit never directly, never for its own sake (G 12, 4:399). So there can be apodictic principles about happiness, but not grounded in it.

happiness can be assigned with certainty” (G 29, 4:418), then principles of happiness would be objective. But, alas, Kant thinks this supposition must always be false.

For, unfortunately, Kant argues that we neither have, nor are we capable of having any such determinate concept of happiness (G 28-29, 4:417-419). So in fact, Kant thinks, we are under no objective principles of happiness, except the one that tells us to pursue our happiness in general. For, in his view, without a sufficiently determinate concept of happiness, happiness can never require anything specific of us. Still, happiness remains an end that we must have, “by a natural necessity”. So “pursue happiness” is to be objectively valid for us, although not apodictically, and always only when it does not conflict with the demands of morality.³¹

For the necessity here is still supposed to come from us, from our natures, and not from reason itself or from the intrinsic nature of prudent action. So principles of happiness would be objective, but still only assertoric, according to the *Groundwork*. For Kant thinks that principles of happiness remain based in an “*actual purpose*” (G 26, 4:415), albeit that our nature requires, rather than because of the intrinsic nature of the actions themselves, or because of reason itself.

Still, we are always committed to our own happiness, even if we actually reject it, either through temptation, through *akrasia*, or through certain kinds of Stoic ethical theories (CP 106, 5:126-127). This commitment to happiness is like the way in which we are always committed to reason, even though, sadly or not, we are still all too capable of rejecting reason in fact. The commitment to both reason and happiness is “built into us”, so to speak. The difference is that we are committed to reason “in itself”, because of what reason itself is, rather than because of what our specific natures are like.

³¹ At any rate, this represents my best efforts to make the cited passages from the *Critique of Pure Reason* and the *Groundwork* cohere on this subject, when the former explicitly says that there can be objective laws of freedom based in “considerations about that which in regard to our whole condition is desirable, i.e. good and useful”, considerations that “depend on reason”, even, while the *Groundwork* insists, among other seemingly conflicting things, that “imperatives of prudence cannot, to speak precisely, command at all, that is, present actions objectively as practically *necessary*” (G 29, 4:418).

I will have more to say about practical necessity in the next chapter. For now, though, I cannot say anything more about how we come to accept grounds and principles, for Kant, or about what he thinks can or cannot be the ground of a valid law of freedom. All I can say for the moment is that, according to Kant, each candidate law of freedom presents, as a “problem” for free choice, a certain relation between a ground, an action, and an effect. Moreover, the law is to suggest that this relation is practically necessary, and Kant thinks that these grounds and relations can really be valid for free agents, regardless of whether or not anyone chooses to act upon them. And, last, Kant thinks that this practical necessity comes in two main kinds: subjective, when it derives from an agent’s actual choices, and objective, when it derives from the nature of the principle itself and from the intrinsic nature of what it represents.

So the crucial point is that Kant thinks that we can still always speak of the ground of free actions, despite the fact that the proper cause of a free action is always ultimately to be its agent, in his view. Moreover, this ground is also to count as an ultimate cause of the action, but only insofar as the agent has chosen to act upon this ground: the ground will count as a cause of the action only relative to the agent’s free choice, in other words. For Kant thinks that an action and an effect still necessarily follow from the ground, even though the agent must also freely choose to act according to this necessary relation.

Kant’s view here might still be rather paradoxical and mysterious, but at least it is an outline of how to necessitate a free choice, since it offers an explanation for how a necessitating ground can be valid for a free agent. It has not, however, even tried to explain how an action or a choice pursuant to such a ground would be free. Even if we agree that we misuse our freedom by failing to act according to a valid ground, mightn’t we misuse our freedom if we follow a valid ground, too? Couldn’t that also be incompatible with freedom? For wouldn’t happiness, or even reason itself, still be a constraint on choice? Even if I “freely” choose to follow such a necessity, wouldn’t I still experience it as a constraint on my choice? Then shouldn’t I still conclude that

happiness or reason is the only ultimate cause of my action? And wouldn't that mean that my choice is not really *absolutely* self-determined after all, as Kant requires?

Kant's answer will be that free choices all share a common intrinsic nature or form, so that certain choices of actions and grounds for action count as free just by dint of the kind of choice that they are. As a result, Kant thinks that there will be a fundamental principle that describes the form that constitutes free choice. Our choices would then count as both grounded and free whenever they conform with this principle and exhibit this form. (Or whenever they proceed from a will that is always capable of respecting this form just as such; a free agent still freely chooses to misuse their freedom, in Kant's view.) And this will be the form of autonomous lawgiving, for Kant: of freely choosing our own grounds, or of freely giving genuine necessities to ourselves.

So what is this form to be, by virtue of which our choices are to count as both free and grounded? In short, the form that allows us to will our principles as universal laws. Kant thinks that by obeying the Formula of Universal Law, our choices can be necessitated by grounds, but still absolutely self-determined. The ultimate ground and cause of our actions can then be both ourselves, our free choices, and a necessitating ground.

Or so Kant argues, and so I am going to try to explain in the subsequent chapters. For now, the important point is that Kant thinks that laws of freedom, as causal laws, must still contain a ground from which an action and an effect necessarily follow, despite the fact that the agent is always the ultimate and proper cause of their actions, too. In short, then, the three general components of a law of freedom are to be a ground, an effect, and an action to bring about that effect. Moreover, Kant thinks that the latter must necessarily follow from the former, at least problematically, waiting upon a free agent's free choice.

Accordingly, we can see now that Kant's practical principles, as representations of laws of freedom, should have the same three general components, related in the same general way:

they should represent some kind of necessary relation between a cause or ground, an effect, and an action connecting them, by means of which an agent is to produce the effect.

To be sure, there are still problems and possible objections to overcome, both philosophical and interpretive. But I will begin to address them in the next chapter, where I will also give some further positive arguments for understanding the basic structure of Kant's practical principles this way.

2.4 Conclusion

Let's return now to the beginning and retrace our steps. We began with Kant's views on general logic, from which we get the hypothetical form of judgment, which represents the relation between a ground and its logically necessary consequence. The Metaphysical Deduction then seeks to take this form of judgment and apply it to reality. Taking a logically necessary relation between grounds and consequences and applying it to real objects is supposed to yield a really necessary relation between a cause and its effect. Moreover, Kant thinks that the underlying logical relation must be realized in turn by an action of the cause to produce the effect. And, to be more specific, as a real relation of the highest generality, causation is to relate only the existence of a cause, the existence of an effect, and the existence of an action. Accordingly, causal relations in general have the following form, for Kant: the existence of an action and the existence of an effect necessarily follow from the existence of a cause.

So far, causation is not supposed to be anything but a purely dynamical relation between existences. So Kant next seeks to apply this concept of causation to our faculty of sensibility in general, by applying it to time. That "schematism" of causation is to yield up the "schematized" concept of causation. This schema of causation is only supposed to add the idea that the dynamical structure of the relation between a real cause, a real action, and a real effect must be reflected in a corresponding temporal structure. Since a cause is to be dynamically prior to its consequent effect and action, the schema of causation then adds the idea that causes are prior in time to their actions and effects, even if all three components exist simultaneously.

In any case, though, the next step in Kant's analysis is to apply the general concept of causation to a free will, that is, to a will that is free, but not lawless or arbitrary, and so still in need of grounds for its actions. Applying the notion of causation in general to a free will, then, gives us the general form of a law of freedom: the existence of a ground makes it necessary for an agent to act a certain way to bring about a certain effect. By which we mean that if the agent accepts, or is committed to accept, the relation between this ground and these consequences as a law for their actions, then they must act this way whenever this ground obtains. And a law of freedom can then be valid either *objectively*, if the agent must necessarily accept this relation, or *subjectively*, if they do already accept it, either directly or indirectly. Otherwise, a law of freedom remains only problematically valid.

Thus, since practical principles represent laws of freedom, according to the interpretation of Kant argued above, these principles must have this structure also. Let's turn now to considering some further reasons for accepting my interpretation of Kant's practical principles, and to addressing some specific problems for my account.

3. The Structure of Practical Principles

So far, I have used Kant's views on cognition and freedom to argue for two main claims: that practical principles, for Kant, are fundamentally representations of laws of freedom, and that laws of freedom are fundamentally laws for the causality of an absolutely self-determining will. Accordingly, I concluded that Kant's practical principles are fundamentally a kind of causal principle, since they represent the laws for the causality of an absolutely self-determining will.

Next, I used Kant's most general account of causal relations, as expressed in the *Metaphysical Deduction of the Categories*, to argue that the basic structure of causal relations, for him, consists in a cause, an effect, and an action of the cause whereby it produces this effect. Moreover, I argued that causal relations have to be necessary, for Kant: the existence of the cause means that the existence of its action and its effect must follow, necessarily.

To speak strictly, then, Kant thinks that causal principles, as representations of causal relations, must take this kind of necessary relation as their proper objects. So I ultimately concluded that Kant's practical principles, as principles for a free causality, must contain a ground for action, an action, and an effect to be achieved by that action. And I argued that practical principles represent the relation whereby this ground makes it necessary, in some sense, either objective or subjective, for a free agent to perform this action for this effect.

With all that in mind, now I want to look at some other lines of Kant's thought, which are both crucial for him, and which also show his practical principles to have this same basic structure. Specifically, we will now examine the Incorporation Thesis from *Religion within the Boundaries of Mere Reason* and the Typic of Pure Practical Judgment from the *Critique of Practical Reason*. Those passages will constitute the work of §3.1 and §3.2. Then we will finally be in a position to study Kant's actual explications (§3.3) and examples of practical principles (§3.4), as they occur throughout his critical writings.

So we will continue to approach Kant's thought and writing by moving from the general to the particular. Approaching Kant's concept of practical principles in this indirect, oblique way

is probably less than ideal; it would likely be preferable to start with what he says about practical principles directly, and build our interpretation off of those remarks. But, as will be apparent soon enough, Kant's direct remarks about practical principles are rather enigmatic, and do not yield a clear or clearly consistent picture of their basic structure. So it will make more sense to continue to treat the clearer general points first, even if we will not need to use them in order to interpret what Kant says about practical principles directly.

Reinforcing my interpretation of Kant's practical principles in these ways will then help position us to address some of the broader problems for Kant's moral theory. For example, one of the oldest and most fundamental problems for Kant has been the persistent unclarity about what can be included in practical principles. The Formula of Universal Law asks us to act on maxims that we can will as universal laws, so we obviously need to know what can go into our maxims, if we are to have any hope of applying this principle to guide our actions in anything but a crude, rudimentary, and intuitive way. But if I am right about Kant's views on practical principles, then we start to get traction on this problem by asking: What, specifically, can the ground, action, and effect in a practical principles consist in? When we act, what ground do we act on, what does it require us to do, and what effect are we aiming to achieve?

We are still a ways off from approaching those questions, however. Even after finishing the direct case for my interpretation in this chapter, we will still need to address alternative views on practical principles in the next chapter, and then we will need to discuss why Kant thinks that the Formula of Universal Law should be the fundamental principle of moral obligation in the first place. Only then will we finally be in a position to begin resolving specific problems with the Formula of Universal Law.

3.1 The Incorporation Thesis

At the end of the last chapter, I explained how Kant wants to use the notions of autonomy and laws of freedom to explain how freedom and necessity can be compatible with one another: how our wills can be free, but without thereby becoming arbitrary, random,

ungrounded, or, in a word, *lawless*. To be sure, Kant gives himself a particularly stringent version of this problem. To be free, Kant thinks that our actions need to be absolutely self-determined, and to be non-arbitrary, he thinks that the very same actions must also be necessitated by a ground that isn't just the agent's free choice. As a result, Kant has to explain how a ground can require us to act a certain way—can put us under a practical necessity—without preventing the will from being one of the ultimate grounds or causes of the action, too. The puzzle here is how to make the ultimate cause of the action free choice, and also a ground necessitating free choice. The free choice of laws for ourselves is supposed to unify this kind of freedom with this kind of necessity: the solution lies in autonomy, in the sense of being a law to yourself, or of giving genuine practical necessities to yourself.

Now, solving this problem about freedom and necessity is one of the deepest goals of Kant's entire critical project, arguably even his single most important philosophical concern. And his famous "Incorporation Thesis" from *Religion Within the Boundaries of Mere Reason* is a crucial component of his solution to this problem. Kant needs to be able to explain how a ground and an agent's free choice can be brought together, unified with one another, into the ultimate cause of an action. The Incorporation Thesis is a necessary, but not a sufficient, part of that explanation.

The Incorporation Thesis, called thus by Henry Allison (1990, 39-40), holds that

freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself, according to which he wills to conduct himself); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom). (R 49, 6:23-24)

Here, the Incorporation Thesis itself is just the idea that an incentive cannot determine a free choice, unless it is "incorporated" into a maxim.

As we will see over the course of the discussion here, Kant's main idea in the Incorporation Thesis is that a free agent's ground needs to be a part of the practical principle

that they choose for themselves—their ground needs to be “incorporated” into their principle, as Kant puts it. So it is easy to see why this incorporation would be necessary to explain how a ground can show up in a law that we give to ourselves. “Incorporating” a ground into a principle will then be necessary to explain how a ground is to be ultimately unified with a free choice.

In short, then, Kant’s notion of autonomy hinges on the Incorporation Thesis. However, this incorporation is not sufficient to explain how a choice of grounds and principles is free. Incorporating certain grounds into your principle can still be a misuse of freedom, or even an abrogation of it. So we will still need to understand the form of free choice. But that is a question for further on down the line.

Anyway, this passage is actually the second time that Kant states the Incorporation Thesis in the *Religion*: the first instance occurs in a footnote only a couple of pages earlier (R 47n, 6:21n), which we will examine presently. That note does not use ‘incorporation’, however, and this second passage occurs in the body of the *Religion*’s text, so it is usually treated as the main statement of the “Incorporation” Thesis.

In any event, though, I now want to argue that this Incorporation Thesis means—says directly, in fact—that grounds must be part of the basic structure of maxims. I will then argue that grounds must therefore be a part of the basic structure of all practical principles in general, too. But grounds require consequences: that follows just from the kind of thing that grounds are, even just in the *Religion*’s own terms. So under the Incorporation Thesis, it will turn out that all practical principles will have to have the basic structure adumbrated in the previous chapter.

To start with, though, this passage does not explicitly say what ‘incentive’ is to mean. However, Kant does clearly imply that incentives are to determine free choice. After all, his basic point here is that an incentive can “determine” a free power of choice “to action” only if it is incorporated into a maxim. So it is clear from the textual context that incentives are to determine the power of choice.

Strictly speaking, though, this passage does not explicitly rule out the possibility that free choice could be determined by something that is not an incentive. However, Kant's point here is not really about incentives *per se*, but rather about the kind of determination that is compatible with free choice. Why must incentives be incorporated into maxims, if incentives are to determine free choice? Because free choice can only be determined through maxims, so anything determining a free choice must be incorporated into a maxim.

This argument is, at best, implicit in the text. But Kant does not explicitly argue for the Incorporation Thesis at all, so reading him this way allows us to see why he should have held it. Moreover, the other Incorporation Thesis passage will confirm this reading, as we will see shortly.

Anyway, though, the upshot is that Kant's reasoning here is perfectly general: by parity of whatever reasoning, even if free choice were determined by something that is not an incentive, that thing would still have to be incorporated into a maxim, or else the choice thus determined could not be free after all. So it is fair to conclude that 'incentive' covers anything that can determine free choice, and Kant's treatment of incentives in his other writings confirms that reading, as we will also see shortly. He even insists that the laws of freedom themselves can, and even sometimes ought to be, the incentives of our wills. For example, morally speaking, "the incentive of the human will (and of the will of every created rational being) can never be anything other than the moral law" (CP 62, 5:72).¹

In short, then, the Incorporation Thesis really says that free choices can only be determined by incentives that are incorporated into maxims.

¹ Kant surely does not mean for this remark to say that we must always be doing something morally obligatory. After all, the Formula of Universal Law carves out a class of morally permissible actions, which do not have to be motivated by the moral law itself. Instead, the moral law can be the ground or incentive of our actions in two ways: as the fundamental boundaries that we respect in our actions, or as the positive ground for a specific action. Moreover, this passage from the *Critique of Practical Reason* also finds Kant insisting that "no incentives at all can be attributed to the divine will" (CP 62, 5:72), parallel to how perfectly rational wills are not subject to any imperatives or "oughts". This technicality does not affect the overall line of my argument here, however, because the important point here is ultimately about the role of grounds in relation to free choice and about the kind of determination compatible with free choice. So my core points can be expressed without relying on Kant's terms for specific kinds of grounds.

Next, notice that Kant's language of determination is enough, even all by itself, to understand an incentive to be a kind of ground: "determining" a free choice and "determining" an action means that once a certain incentive is given, a certain choice and a certain action must follow. Kant confirms this interpretation with his parenthetical clause explaining how incorporating an incentive into a maxim involves making that incentive "into a universal rule" for how we "will to conduct ourselves". For that rule will say, universally, that when the incentive obtains, we must "conduct ourselves" according to the rule.

And, in fact, Kant's whole framing of the issue as a problem about making incentives "coexist with the absolute spontaneity of the power of choice (of freedom)" implies that incentives are to make certain choices necessary. To appear incompatible with that absolute spontaneity, an incentive would have to require an action, independently of that spontaneity. In which case, though, incentives turn out to be grounds of choice, in Kant's technical sense.

So now the Incorporation Thesis says that free choices can only be determined by grounds that are incorporated into maxims. So we can see right away that this Incorporation Thesis is supposed to be a necessary condition for Kant's reconciliation of freedom and necessity. That is apparent just from how Kant's question in this passage is about how "freedom of the power of choice" can be "determined to action through any incentive", and from how this determination is impossible "*except so far as the human being has incorporated [the incentive] into his maxim*", so that "only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom)." As determining free choice, incentives count as grounds thereof. This passage therefore finds Kant trying to explain how a free power of choice can still be grounded, granted that grounds must have necessary consequences, such as actions and choices.

And, once again, autonomy is to be the answer: we have to make incentives "into universal rules" for how we "will to conduct" ourselves. For, Kant says, "only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of

freedom).” And the only way to turn an incentive into a universal rule like this is to incorporate it into a maxim, he says.

Thus, Kant’s point here is ultimately about the grounds of the free power of choice, and this is what incentives are to be. Kant’s explications of incentives in other texts confirms this reading: according to the *Groundwork*, an incentive is “the subjective ground of desire” (G 36, 4:427), and the *Critique of Practical Reason* avers that an incentive is “the subjective determining ground of the will of a being whose reason does not by its nature necessarily conform to the objective law” (CP 62, 5:72).²

In fact, the *Religion* itself also confirms this reading of ‘incentive’ in its other statement of the Incorporation Thesis (which, as I mentioned before, actually occurs first in the text). It goes:

That the first subjective ground of the adoption of moral maxims is inscrutable can be seen provisionally from this: Since the adoption is free, its ground (e.g. why I have adopted an evil maxim and not a good one instead) must not be sought in any incentive of nature, but always again in a maxim; and, since any such maxim must have its ground as well, yet apart from a maxim no *determining ground* of the free power of choice ought to, or can, be adduced, we are endlessly referred back in the series of subjective determining grounds, without ever being able to come to the first ground. (R 47n, 6:22n)

We need not worry about the conclusion that Kant is driving at here, or about whether his argument for it is at all persuasive. Instead, simply notice that Kant’s guiding idea is, once again, that the grounds of free choices must always occur in a maxim: “apart from a maxim no *determining ground* of the free power of choice ought to, or can, be adduced”. That is the *Religion*’s second statement of the Incorporation Thesis, which implies that if free choice is to be grounded at all, it must be through a maxim containing a ground. Kant says that “any such maxim must have its ground”, and “no *determining ground* of the free power of choice ought to, or can, be adduced” “apart from a maxim”. So any ground for free choice must occur in a maxim.

² See note 1 above.

Moreover, this passage also treats incentives as a kind of ground: “Since the adoption [of moral maxims] is free, its ground...must not be sought in any incentive of nature”, at least apart from a maxim. So this passage, too, supports understanding incentives as a kind of ground, and it also holds that these grounds need to be incorporated into maxims, if free choice is to be grounded, and rendered compatible with the necessity in a ground.

Now, Kant’s language of “incorporation” is a little enigmatic. His German here is ‘*aufgenommen*’, which is fairly translated with ‘incorporated’, since the German more literally connotes taking on, taking in, or taking up. So if the Incorporation Thesis holds that grounds, in the form of incentives, must be incorporated, taken up, or taken in to maxims, it would be natural to understand this Thesis as saying that incentives need to be part of the basic structure of maxims.

However, this language of incorporation could also suggest that an agent already has their principles complete and fully formed, before incorporating an incentive into that preexisting, preestablished principle. In that case, it might seem as though grounds would not have to be a part of the *basic* structure of a maxim, since then our maxims would just be there, fully formed and ready at hand, waiting on us to incorporate incentives into them.

However, I take it that Kant really has the exact opposite picture instead: incentives come first. (Logically speaking, not temporally speaking.) I am tempted to act a certain way, let’s suppose, perhaps by a feeling or an inclination. If I am free in Kant’s sense, however, then that feeling or inclination alone cannot be enough to cause me to act. Instead, I must also freely choose to act on this feeling. The Incorporation Thesis, however, says that free choices can only be determined by grounds that are incorporated into maxims. So this feeling can only be the basis for a free choice if it becomes a ground for action, and if it is incorporated into a maxim as such a ground. And if that choice is to be free, then it must be me who makes this feeling into a ground, and incorporates it as such into my maxim.

Now, the former picture holds that I would already have a fully formed maxim somewhere, just waiting for this feeling to be incorporated into it as its ground. But it is wholly unnecessary to posit such maxims. Instead, I can “build up” a maxim around this feeling, which I can do simply by considering the feeling as a possible ground for me, rather than by incorporating it into a maxim that I already have.

I consider this feeling as a possible ground for me just by representing that feeling to myself as requiring me to act a certain way. In Kant’s terms, that is already to represent it as a ground for that action. Ultimately, then, to represent a feeling as a ground is just to understand the feeling in a certain way: to conceptualize what the feeling is prompting me to do, or to make a judgment about what it would take for me to satisfy that feeling. Then I just have to decide whether I really ought to act at that feeling’s prompting.

Thus, simply by representing my feeling as a possible ground for my actions, I already frame for myself a principle capturing what this feeling requires me to do. This is the way in which grounded freedom requires its grounds to occur within a principle, Kant thinks, and also the way in which the principles for such freedom must then have grounds as part of their basic structure. For here, the identification of a ground for action is essential to the (logical) generation of the maxim, and drives the process of its (logical) construction. In fact, then, representing something as requiring you to act a certain way is already, *eo ipso*, to generate a practical principle, wherein the former is incorporated as a ground, requiring the latter.

However, even if Kant intends the first model of incorporation, he would still hold that maxims cannot actually determine our actions without an incorporated ground. Free choice, Kant says, can be “determined to action” only through an incentive—a ground—that is “incorporated” into a maxim. So it will still follow that any maxim that we can actually act on must have a ground as part of its structure. Likewise, since actual free choices cannot be grounded except through an incentive that occurs within a maxim, that fact should resolve any

doubts that there might be as to whether incentives are a part of the *basic* structure of a maxim for grounded free choice, rather than a part of some non-basic structure.

So the Incorporation Thesis implies that grounds, in the form of incentives, must be included in the basic structure of our maxims. This is already enough, though, to establish that maxims must have the basic structure that I have been proposing. We do not even need to reach back to the Metaphysical Deduction in order to see this. For the Incorporation Thesis passages are both clear that incentives must have choices and actions as their necessary consequences. After all, this was how we concluded that incentives had to be grounds in the first place. So either way, an action or a choice is to be the effect of the ground, which the agent is to bring about by their choice or volition itself. Ground, end, action, necessary relation between them.

In general, moreover, what is important about the ground—what makes it a ground at all, in fact, even just in the *Religion's* own terms—is the role or function that it plays in determining choice and action. So a maxim cannot include a ground as such, as a ground, without also including its consequences, as well as the necessary relation between them. The Incorporation Thesis implies, then, that maxims must include the entire necessary relation between a ground, an effect, and an action; that follows from the Incorporation Thesis, just because Kant holds in it that incentives must be incorporated into maxims.

Now, I have been arguing that Kant's reasoning in the two Incorporation Thesis passages in the *Religion* holds for all grounds of free choice whatsoever: a ground of free choice must be incorporated into a maxim, as part of its basic structure. However, the passages discussed so far are clear that incentives are not equivalent to grounds in general: Kant repeatedly describes them as “subjective” grounds. So even if the Incorporation Thesis tells us that maxims must represent a necessary relation between an incentive, an action, and an effect, it might still be hard to determine what this Thesis would imply about the relation between grounds in general and practical principles in general. So far, the Incorporation Thesis only seems to cover subjective necessities, subjective grounds, and subjective principles. What about objective ones?

As I have just said, both of the Incorporation Thesis passages are clearly concerned with the grounds and principles that individual agents actually act on: the first passage asks about the grounds that can determine a free power of choice “to action”, while the second passage starts off with a question about how and why moral maxims are adopted, instead of immoral maxims. Moreover, it is clear in both passages that the Incorporation Thesis itself is intended to apply to the actual choices and actions that free agents undertake. After all, we have seen that the overall point of the Thesis itself is to help explain how there could actually be free choices that are actually not arbitrary or ungrounded.

So the Incorporation Thesis passages are about grounds and principles that are subjective, in the sense of being the grounds and principles that free agents actually act upon. Subjective grounds and principles contrast, of course, with objective grounds and objective principles, which free agents ought to act on, even though they might not in fact do so. This contrast is particularly evident in Kant’s explications of practical principles, which we will look at presently. But Kant does also sometimes contrast incentives, as subjective grounds, with objective grounds. For example, the *Groundwork* says that “The subjective ground of desire is an *incentive*; the objective ground of volition is a *motive*; hence the distinction between subjective ends, which rest on incentives, and objective ends, which depend on motives, which hold for every rational being.” (G 36, 4:427)³

In any case, though, since incentives and maxims refer to how agents actually do act, anything that we learn about their basic structure will automatically and immediately extend to objective grounds and principles as well. To say that we ought to act according to a ground or a principle is to say, at least, that our actual grounds and principles ought to conform to the objective grounds and principle. (It might also say that the objective ground and principle actually ought to be our subjective ground and principle. I tend to read Kant that way, but there

³ Notice the ‘hence’ here: Kant is clearly thinking that subjective and objective grounds have subjective and objective ends as their consequences, as my general account of his views on grounds would predict.

is disagreement about whether that reading is correct. I will have some things to say about that in §3.3 below.) But if our subjective grounds and principles have to conform to objective ones, then their basic structure must correspond exactly to the basic structure of their objective correlates. Otherwise, the ground or principle that we actually act on could not fully capture the structure of the objective requirements that we are under. Accordingly, we would not be doing as we ought, even if we followed an objective ground or principle to the best of our ability. Moreover, if “ought implies can”, as Kant held (CP 26-28, 5:29-30), then the basic structure of our maxims—the basic structure of the principles that we actually act on—must constrain and determine the basic structure of all practical principles. We could never exceed or transgress this basic structure, so it will follow that we ought not to do so, either.

So subjective and objective principles must have the same basic structure, according to Kant. However, Kant also means for the distinction between the subjective and the objective to be exhaustive as well. For one thing, as we will see more fully soon, Kant presents this distinction as exhaustive, when he uses it to classify principles and grounds. For another, this distinction is itself conceptually exhaustive: the entire domain of principles and grounds is covered by the distinction between principles, say, that must hold for an entire class of agents and principles that could hold for only a single agent, or for no one at all.

So there is no problem in extending what the Incorporation Thesis teaches us about incentives and maxims to all grounds and all practical principles. Therefore, the Incorporation Thesis implies that all practical principles must have the basic structure that I have proposed.

3.2 The Typic of Pure Practical Judgment⁴

Now, the next major line of Kant’s thought, which converges on this same structure for practical principles, is the “typic” of pure practical judgment. Pure practical judgment is just judgment according to pure practical principles, such as the fundamental principles of morality.

⁴ The discussion here follows and is intended to summarize the main ideas in the Second Critique’s “Of the Typic of Pure Practical Judgment” (CP 58-62, 5:67-71).

I'll say more about what these principles are momentarily; for now, just note that Kant raises a problem about this kind of judgment, which their "type" is supposed to solve. And here I want to argue that this solution requires practical principles to be representations of a necessary relation between a ground, an action, and an effect, as my interpretation contends.

The difficulty about pure practical principles, such as the Formula of Universal Law, is that these principles are supposed to lack any empirical or sensible content; that is just what their purity is supposed to consist in (CR 137, B3; CP 58-59, 5:67-68). For example, Kant holds that none of the concepts invoked in "*act only in accordance with that maxim through which you can at the same time will that it become a universal law*" (G 31, 4:421) include a representation of a color or a smell, for example, or anything else that would be essentially sensible or empirical.

Nevertheless, despite this purity of content, we human agents are still supposed to use the Formula of Universal Law to guide our actions. Accordingly, we must use this Formula in order to make judgments about how to act.

Naturally enough, Kant calls such judgments "practical" judgments.⁵ And agents like us must often act in sensible nature, so that our actions must often be sensible as well. (Omissions and acts of volition are a separate question.) So Kant runs into a question about how we can use a pure practical principle, such as the Formula of Universal Law, to make judgments about how to act in the sensible world. How can we use a principle that has no sensible or empirical content in order to make judgments about sensible actions in the empirical world? That would involve applying representations with no empirical content to empirical objects. How can that be possible? How can an empirical object satisfy a representation with no empirical content?⁶

⁵ Strictly speaking, judgments about how to act are only one type of practical judgment. In general, "practical judgment" comprehends any judgment about how the world ought to be. That follows from how laws of freedom are practical laws of "what ought to be", as we saw back in Chapter One.

⁶ This is parallel to the problem to be solved by the schematism of the categories in the First Critique, discussed in Chapter 2.2 above. Kant explicitly draws this parallel in the Typic of the *Critique of Practical*

Kant's idea in the Typic is that we can rely on the concept of a law of nature in order to properly apply pure practical principles to empirical objects and actions. Accordingly, by virtue of fulfilling this function, he calls laws of nature themselves the "type" of pure practical principles (CP 60, 5:69). But how does the concept of a law of nature allow us to apply pure practical concepts and principles to objects and actions in the sensible world?

On the one hand, laws of nature have the same form as the laws that pure practical principles represent, since they both have the form of laws in general. On the other hand, though, laws of nature have the same matter or content as actions in nature, since laws of nature relate sensible objects to one another according to causal regularities in nature. In this way, laws of nature are supposed to be able to show us how "the *form of lawfulness* in general" (CP 61, 5:70) is exhibited in empirical objects. We can then use the concept of a law of nature in order to link the sensible content in particular actions and their maxims back up with the form required by pure practical principles.

So the concept of laws of nature acts as a kind of "middle term" in our judgments under pure practical principles, and that is why laws of nature are themselves the type of pure practical principles and pure practical judgment both.

Kant explicitly introduces the concept of a type for pure practical principles only in the Second Critique. In retrospect, however, we can see this idea already prefigured in the *Groundwork's* Formula of the Law of Nature: "*act as if the maxim of your action were to become by your will a **universal law of nature.***" (G 31, 4:421) For the Typic in the *Critique of Practical Reason* explains that

The rule of judgment under laws of pure practical reason is this: ask yourself whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will. (CP 60, 5:69)

Reason, but he also denies that the type of pure practical judgment is a schema in the same way (CP 59-60, 5:69). Presumably this is why he chooses to call it a "type" instead of a "schema", despite the broad functional similarity.

In this way, laws of nature turn out to be “a *type* for the appraisal of maxims in accordance with moral principles” (CP 60, 5:69), for we need to use the concept of a law of nature in order to determine whether our maxims conform to pure practical principles, such as the basic moral principles.

So what does this mean about the basic structure of practical principles in general? The crucial point here is that the fundamental moral principles are supposed to be pure, while they are also supposed to constrain both the actions and the maxims involved in our agency in empirical nature. That is all evident in the Formula of Universal Law. So the judgments that we have to make in order to apply this Formula to the empirical world are actually the paradigm case of the problem that the type of pure practical judgment is supposed to solve.

Then, since laws of nature are the type of pure practical judgment, Kant’s idea is that we can use the concept of a law of nature to mediate between the fundamental principles of morality and our judgments about which particular maxims to adopt for our agency in nature. And to carry out this function, laws of nature must exhibit the same form as both the pure moral laws and the maxims that we adopt pursuant to them.

For Kant says that pure practical principles need to ask “whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will”. For the moment, we needn’t concern ourselves with why Kant thinks that this is our basic practical question, or with the mechanics of how we are to answer this question for ourselves. I will have some things to say about these issues in later chapters. For now, though, the point is only that this question is not intelligible, except under the assumption that the maxims for our actions, the laws of nature, and the pure practical principles can all share a basic form in common. Judgment according to pure practical principles requires us to consider our maxims as though they were laws of nature, so it follows that these all must be capable of the same basic form: the form of lawfulness in general.

To be clear, though, Kant intends for this form of law to be a test for whether or not our maxims are morally permissible. As he puts it, “If the maxim of the action is not so constituted that it can stand the test as to the form of a law of nature in general, then it is morally impermissible.” (CP 60, 5:69-70) And this is how the form of law, as exhibited in laws of nature, must work: if this form is to serve as a guide to moral judgment, then it must be capable of distinguishing moral maxims from immoral maxims, and, in general, of separating what conforms to pure practical principles from what does not. So not every maxim will be able to exhibit the form of law, even though every law does exhibit that form. Still, though, maxims *in general* must be capable of the form of law, even if particular maxims cannot have it. And then maxims must have the same basic structure as laws of nature and pure practical principles, for otherwise there could be no question at all of their being able to have the form of lawfulness in general.

In short, then, the Typic of Pure Practical Judgment means that moral judgment—and hence moral agency in general—would be totally impossible for Kant if maxims, pure practical principles, and laws of nature did not have the same basic structure. And if we cannot act on moral laws or moral obligations, then morality in general becomes impossible.

Thus, according to Kant, the very possibility of morality itself depends on whether or not maxims, laws of nature, and pure practical principles all share the same basic structure. For Kant is now arguing that we could never use the fundamental principles of morality to guide our actions at all unless we can use laws of nature as the type of both maxims and pure practical principles. But we cannot use laws of nature as the type of these principles unless they all have, or are at least capable of having, the same form: “the *form of lawfulness* in general”. In exhibiting this form, the Typic explains that “to this extent laws as such are the same, no matter from what they derive their determining grounds.” (CP 61, 5:70) So pure practical principles and laws of nature must have the same basic structure. And our maxims must be able to have this form, too, along with its corresponding basic structure, if we are to be able to determine whether

or not our maxims can conform with the moral law. For we cannot even begin to consider what would transpire if our maxims were to become universal laws of nature, unless our maxims have the same basic parts as these laws, arranged in the same basic way.

So what is this “*form of lawfulness* in general”, which maxims, laws of nature, and pure practical principle must be able to share, if pure practical judgment, and even morality itself, is to be possible, according to Kant? And what basic structure goes along with it?

Laws of nature are supposed to be our guide here, since their form is supposed to be the key to linking maxims and pure practical principles. We could look back to the *Critique of Pure Reason* here, and specifically to the Metaphysical Deduction again. In fact, the Tycic encourages precisely this, by referring to the Transcendental Analytic and by comparing types and schemas (CP 59-60, 5:68-70).

However, we do not actually need to look back to the First Critique, because the Tycic answers these questions, too, in its own, independent terms. In setting up the problem that the type of pure practical judgment is supposed to solve, Kant explains that

a practical rule of pure reason *first*, as *practical*, concerns the existence of an object, and *second*, as a *practical rule* of pure reason, brings with it necessity with respect to the existence of an action and is thus a practical law, not a natural law through empirical grounds of determination but a law of freedom in accordance with which the will is to be determinable independently of anything empirical (merely through the representation of a law in general and its form) (CP 59, 5:67-68)

To unpack this passage: the pure practical principles at issue in the Tycic will represent these “practical rules of pure reason”, because these rules are also “practical laws”. And Kant says that these rules will “concern the existence of an object”, and also “bring with them necessity with respect to the existence of an action”. Hence, we should conclude that the existence of this object is to be the result of the action that the rule requires, since the rule “concerns” the existence of the object, while requiring the action. (Although nothing prevents the action from being the object to be achieved as well; remember that “object” is to be read as broadly as possible.) Moreover, the underlying practical law is to be “a law of freedom in accordance with which the

will is to be determinable” “merely through the representation of a law in general and its form”. Since the will is to be determinable through this representation, it also follows that this representation “of a law in general and its form” will be a determining ground of the will, at least in potentiality. And since the law requires the will to be determinable merely through this representation, it follows that this ground is to be contained in the law, also.

So here, in the Typic, we can see Kant attributing the necessary ground/action/effect structure to pure practical principles. This is an important piece of evidence for my overall interpretation of practical principles in general. But the contrast that Kant draws between natural laws and practical laws is more important. For, in making this contrast, he says that a practical rule of pure reason is “a practical law, not a natural law through empirical grounds of determination but a law of freedom”. So his point is that laws of nature are to work through “empirical grounds of determination”. By now, I hope the next step in the analysis will be fairly familiar: “grounds of determination”, by their concept, have necessary consequences, which means that the Typic is saying that laws of nature have the basic structure that I have been proposing.

So Kant’s Typic presents laws of nature as having this same basic structure. Pure practical principles must then have this structure also, in part because laws of nature are to be their type, and in part because, when it comes to the form of law in general, “to this extent laws as such are the same, no matter from what they derive their determining grounds” (CP 61, 5:70), as Kant puts it, a little later on in the Typic. But then maxims must have the same basic structure, too, because they are to conform to pure practical principles, and because the concept of a law of nature is to mediate this conformity.

Now, unlike the distinction between subjective and objective principles, the distinction between maxims and pure practical principles is *not* exhaustive. So we have not shown yet that *all* practical principles must have the same basic structure, in Kant’s view. However, pure practical principles, of morality, anyway, are to be the standard for judging the moral

permissibility of any practical principle whatsoever. So, by parity of the Typic's reasoning, all practical principles must have this same basic structure.

Kant's solution to the problem of pure practical judgment, then, hinges on the idea that all practical principles have the basic structure that I have proposed. As does the possibility of morality itself, because morality depends, Kant thinks, on the possibility of pure practical judgment. If this analysis is correct, then, it will turn out that Kant actually must hold that all practical principles have this basic, practically necessary ground/action/effect structure, or else his entire moral theory would have to fall apart.

3.3 Kant's Explications of Practical Principles

To this point, we have examined three major strands of Kant's thought, each of which independently leads to the idea that practical principles represent a necessary relation between a ground and its consequent action and effect. This structure for practical principles comes out of Kant's views on causation, on freedom, and on morality itself, as expressed, respectively, in the Metaphysical Deduction of the Categories, the Incorporation Thesis, and the Typic of Pure Practical Judgment. In fact, we have seen that this structure is essential to Kant's views in all three of these areas, where it is inextricably bound up with three of the deepest parts of his philosophical project: defending our ability to know the objective necessity in causal relations, reconciling freedom and necessity through the concept of autonomy, and vindicating the purity of moral obligations. And the logical relation of ground and consequence turns out to be fundamental in all three of these domains, and to lie at the heart of the relations that practical principles are to represent.

Now, though, we finally need to turn to Kant's direct remarks on practical principles, where he explains what practical principles are to be, and where he gives examples of them. For both his explications and his examples of practical principles raise problems for my interpretation of their structure, which I now want to try to resolve.

So let's begin with his official explications.⁷ As I mentioned before, Kant divides practical principles into two basic kinds: maxims and practical laws. Accordingly, he characteristically explicates practical principles by contrasting these two kinds. For example, in the First Critique's Canon of Pure Reason, Kant says that

Practical laws, insofar as they are at the same time subjective grounds of actions, i.e. subjective principles, are called **maxims**. The **judgment** of morality concerning its purity and consequences takes place in accordance with **ideas**, the **observance** of its laws, in accordance with **maxims**. (CR 681, A812/B840)⁸

Moreover, a crucial footnote from the *Groundwork of the Metaphysics of Morals* holds that

A *maxim* is the subjective principle of acting, and must be distinguished from the *objective* principle, namely the practical law. The former contains the practical rule determined by reason conformably with the conditions of the subject (often his ignorance or also his inclinations), and is therefore the principle in accordance with which the subject *acts*; but the law is the objective principle valid for every rational being, and the principle in accordance with which he *ought to act*, i.e., an imperative. (G 31n, 4:421n)

And Section I of the *Groundwork* also says that

A *maxim* is the subjective principle of volition; the objective principle (i.e., that which would also serve subjectively as the practical principle for all rational beings if reason had complete control over the faculty of desire) is the practical *law*. (G 14n, 4:401n)⁹

⁷ I prefer to call these remarks "explications" rather than "definitions", so as to avoid prejudging how much they tell us about what practical principles are to be. Is Kant giving constitutive descriptions? Are these descriptions supposed to capture what is fundamental to practical principles? Kant does not usually give us very much explicit guidance when he is "defining" concepts, and usually only seems to need such descriptions to apply to all of their "*definienda*".

⁸ I mentioned above that there is controversy as to whether practical laws can literally be maxims, in Kant's view. For example, Allison (1990, 87) and Aune (1979, 24-25) both think, for different reasons, that practical laws cannot be maxims. This passage is obviously a fairly large problem for their reading. I will give some running commentary on this issue in the next few footnotes.

⁹ Besides the essential agreement with all of the passages quoted above, note that Kant says here, almost outright, that practical laws can be maxims. I think this is also the subtext of how maxims are the principles upon which we act, while laws are the principles on which we ought to act. From this, it appears to follow right away that we cannot actually act as we ought unless we make the law our maxim. To avoid reading Kant this way here, Allison argues that this explication involves a weaker notion of subjectivity than the former one does (Allison 1990, 87). Here, he says, maxims are only subjective in the sense of being the principle on which an agent acts, while before maxims had to be subjective in the sense of conforming to "conditions of the subject" as well. This stronger subjectivity, Allison contends, is incompatible with the idea that laws can be maxims. This is a good observation about the literal text, and Kant is characteristically brief about what 'subjective' and 'objective' are to mean. (Although Kant is not necessarily unclear by dint of said brevity.) In any event, though, the principle that an agent actually acts on must pretty obviously contain "the practical rule determined by reason conformably with the

Compare these explications to the “Definition” that Kant gives at the very outset of the main text of the *Critique of Practical Reason*:

Practical *principles* are propositions that contain a general determination of the will, having under it several practical rules. They are subjective, or *maxims*, when the condition is regarded by the subject as holding only for his will; but they are objective, or practical *laws*, when the condition is cognized as objective, that is, as holding for the will of every rational being. (CP 17, 5:19)¹⁰

And to these two nearly adjacent passages from the *Metaphysics of Morals*:

conditions of the subject” as well. How could we possibly act on a rule that did not conform to the conditions of our subjects? But then it looks as though we cannot act on, or even in conformity with, objective principles unless they, too, “conform with the conditions of the subject”. But then objective principles would count as subjective as well, even in Allison’s more robust sense. So, *contra* Allison, it is difficult to see why even this apparently more robust notion of the subjectivity in maxims should preclude laws from being maxims, given how Kant speaks here. Especially when at least some subjective conditions are also objective, e.g. the condition of being bound by the fundamental laws of freedom. Allison also makes a separate point about imperatives, which I will address in the next chapter, along with his own view on maxims.

¹⁰ Some commentators take this “definition” of maxims to hold that agents must view their maxims as valid for themselves and nobody else. Likewise, these commentators think that this definition says that practical laws must be viewed as valid for all rational agents. Thus, they conclude that maxims cannot be practical laws, for Kant. Finding this suggestion unwelcome, some interpreters want to draw a distinction on Kant’s behalf between maxims that can also be laws and “mere” maxims, which cannot be (Beck 1960, 81-82). While I do not object to this distinction, I do find it to be rather odd. Because Kant’s moral theory is addressed to imperfectly rational, imperfectly free agents, there is always space between recognizing something as valid for you and allowing that validity to govern your actions. So there is always space between believing that something is a law for you and adopting that law as a law for your actions. In Kant’s vocabulary, this situation is most naturally described by asking about whether or not we make the law our maxim, and I do not think that Kant’s “definition” here forces us to speak another way, either. Kant’s definition actually does not say very much: it says only that if a condition “is regarded by the subject as holding only for his will”, then the corresponding principle is a maxim, while if a condition “is cognized as objective”, then the corresponding principle is a practical law. So this definition does *not* say that a principle is a maxim *only* if “the condition is regarded by the subject as holding only for his will”. Furthermore, as the main discussion here will emphasize shortly, this definition distinguishes maxims from practical laws based only on how the condition is represented: based on how it “is regarded by the subject”, in the case of maxims, and on how it is “cognized” in the case of practical laws. But I can represent a condition or a principle as valid for me without adopting it for myself. In the previous chapter, I explained that to adopt a principle is to come to believe that the necessity represented in the principle is valid for you. But representing something as valid is not the same as believing that it is valid: I can suppose that a principle is valid, or entertain the idea that it is valid, while remaining scrupulously neutral about whether or not it is actually valid. In fact, the Formula of Universal Law requires precisely that, when it asks us to consider our maxims as though they were universal laws. These exercises require us to represent a principle as valid in a certain way, but they decidedly do not commit us to believing in that validity. As a result, Kant’s definition here simply does not say anything about the status of a principle that I both cognize as objectively valid, and also adopt for myself. So it remains open for us to say that a principle that I adopt for myself is my maxim, even if I regard that principle as objectively valid, too. As usual, I think we should take the weaker reading, especially when it squares better with Kant’s other remarks and avoids this potential puzzle. Thank to Tyler Burge, whose comments here made me realize that my first treatment of this issue was quite inadequate.

A principle that makes certain actions duties is a practical law. A rule that the agent himself makes his principle on subjective grounds is called his *maxim*; hence different agents can have very different maxims with regard to the same law. (MM 17, 6:225)¹¹

A *maxim* is a *subjective* principle of action, a principle which the subject himself makes his rule (how he wills to act). A principle of duty, on the other hand, is a principle that reason prescribes to him absolutely and so objectively (how he *ought* to act). (MM 17-18, 6:225)

So what do we learn about practical principles from these different explications? First, maxims and laws are both principles for action, at least when we construe action broadly enough to include volition. (As I think we pretty clearly should, for Kant, since we have already seen that the will is itself a faculty of causality.¹²) So maxims and laws are both “practical” principles, and Kant repeatedly treats them as such directly, too, not just by making them principles of action.

But maxims and practical laws are principles of action in different ways. Maxims are, variously,

- “subjective grounds of action” and principles for “the **observance** of” moral laws (CR)¹³
- “the subjective principle of acting”, “the principle in accordance with which the subject *acts*”, and the “the subjective principle of volition” (G)
- principles that “contain a general determination of the will”, but “are subjective” (CP)
- “a rule that the agent himself makes his principle on subjective grounds”, “a *subjective* principle of action”, and a “rule” for how an agent “wills to act” (MM)

¹¹ I take it that Kant means that we can perform an action required by a moral law for very different reasons, or in accordance with very different grounds, to use the terminology introduced in the previous chapter. So Kant’s point is that different principles can “overlap” as to their action. This will be especially true for the moral laws that make up the Doctrine of Right, where what matters is only whether or not our actions conform with the law, and not why they conform. Instead, our reasons for obeying the moral law are relevant only to whether we are virtuous, and not to whether we are just (MM 24-25, 6:230-231). Passages like these are good evidence that maxims must contain grounds, as I have been arguing.

¹² In fact, I take it that Kant recognizes two kinds of action: volition, which is action “internal” to the will, and the “external” actions subsequent to volition. Not every volition leads to “external” action, though, and there may even be space for Kant to say that some of our external actions are not willed, even though I understand his view to be that external actions are individuated by the underlying principles of our volition. In any case, there is no real problem in my saying that practical principles contain an action, since we know perfectly well what lying is, for example, even if the full nature of specific lies depends on the principles from which they are ultimately derived. Thanks to Tyler Burge for pressing me to clarify.

¹³ There can clearly be maxims that are not just for the observance of moral laws, and this passage surely does not deny that.

While laws are

- “the *objective* principle”, “the principle in accordance with which [an agent] *ought* to act” (G)
- principles that “contain a general determination of the will”, but “are objective” (CP)
- “A principle that makes certain actions duties” and “a principle that reason prescribes...objectively” for how an agent “*ought* to act” (MM)

And earlier in the First Critique’s Canon, Kant uses the notion of “practical freedom” to derive “objective **laws of freedom**” that “say **what ought to happen**” (CR 675, A802/B830).

So these explications are remarkably consistent across Kant’s entire critical corpus. And when we put them together, we find that maxims are principles according to which agents actually act, while laws are principles according to which agents ought to act.

Moreover, Kant clearly intends for both “principles according to which agents act” and “principles according to which agents ought to act” to include elements of necessity. For maxims, Kant’s idea is that to adopt a maxim is to undertake a commitment to act a certain way: it is to adopt a principle according to which you undertake to act. As a result, our actions become necessary, relative to the maxims that we choose for ourselves. We act “according” to them, Kant sometimes says, and we are supposed to satisfy them in our actions. They also constitute “rules” for ourselves, and Kant also connects them with grounds, and with the determination of the will.

And Kant intends for practical laws to be principles for how we “ought to act” in the sense of “ought necessarily”, rather than the sense of a mere recommendation. This element of necessity in “ought to act” is evident in the connections that Kant draws between practical laws and laws (CR 681, A812/B840), imperatives (G 31n, 4:421n), rational necessity (G 14n, 4:401n), the determination of the will (CP 17, 5:19), and duty (MM 17, 6:225; MM 17-18, 6:225).

So it is clear that both maxims and laws are to be principles necessitating action, just of different kinds: subjective versus objective, for our actual actions versus for how we ought to

act.¹⁴ So we should expect laws and maxims to have a basic structure in common as principles of action—as practical principles. And as principles of action, broadly construed, practical principles will turn out to be a kind of causal principle also. This is particularly evident in Kant’s descriptions of maxims as principles of observance, action, volition, the will, or agency. And the point must extend to principles for how these things ought to be conducted, too.

So these explications ratify the approach that the previous chapter used to determine the basic structure of Kant’s practical principles. But on the whole, at least, it is also true that these explications themselves give little indication as to what that basic structure should be.

To be sure, some of these explications do provide some evidence that practical principles have the basic structure that I have proposed. I have been trying to show how the crucial point is about whether or not practical principles have to include grounds for action. Treating maxims and practical laws as including an element of necessity certainly suggests that they ought to

¹⁴ It is not clear how these two distinctions are supposed to relate to one another, if at all. I tend to think that the subjective/objective distinction is fundamental, since Kant consistently emphasizes it, especially in the Second Critique’s Definition. But Rüdiger Bittner argues that we cannot explain maxims through this contrast at all, because such an explanation would be circular (Bittner 1974, 487n). His argument, though, depends on two misreadings of Kant’s procedure in the *Groundwork*’s Second Section. He thinks that Kant tries to derive the content of the moral law from the concept of a will that acts according to maxims, so that it would be circular to define maxims by their contrast to the objective moral law. I am not sure how this is supposed to follow, and Bittner does not elaborate. But suppose that Kant uses the distinction between objective and subjective principles to define the concept of a will that acts according to maxims, and then uses that concept to determine the content of the fundamental objective principle: I do not see the circle in that procedure, since it argues from the kind of principle to the content of the principle. In any case, though, Bittner can only object to defining maxims through their contrast with moral laws. To speak strictly, though, Kant does *not* contrast the subjectivity of maxims with the objectivity of *moral* laws, but with the objectivity of *practical* laws. The latter category is more general than the former, since moral laws are only a type of practical law: hypothetical imperatives also express practical laws (G 24-25, 4:413-414). So Bittner seems to misunderstand Kant here, and in doing so he rests his argument on a false premise. How far that matters, though, is hard to say without a clearer idea of what the circle is supposed to be. Regardless, though, Bittner is wrong, and pretty clearly wrong, I think, to hold that Kant’s fundamental concept here is that of a will acting according to maxims. Bittner does give an argument for reading Kant that way, but I do not have space to address it here. In the relevant passage, however, Kant says that the will is “the capacity to act *in accordance with the representation* of laws [*Gesetze*]” (G 24, 4:412), and he calls these representations principles (*Prinzipien*), not maxims (*Maximen*), even though he has already introduced the concept of a maxim by this point in the argument (G 14n, 4:401n). Moreover, Kant then proceeds by distinguishing perfectly and imperfectly rational wills—those that always act as reason says they ought from those that need not do so. Hence, it seems that ‘principles’ means practical laws, not maxims, though Kant could also mean both. Neither interpretation, though, supports treating maxims as the fundamental concept, and especially not under Bittner’s own account of maxims, where they are our general rules or policies of life, our “*Lebensregeln*”. So I do not see why the subjectivity of maxims cannot be their fundamental feature, which distinguishes them as a type of practical principle.

include grounds for this necessity. Moreover, the explications quoted above do sometimes treat maxims and practical laws as containing grounds or conditions. The explication in the Canon of Pure Reason comes close to saying so directly when it declares that “Practical laws, insofar as they are at the same time subjective grounds of actions, i.e. subjective principles, are called **maxims**.” (CR 681, A812/B840) Kant also refers to grounds or conditions in the *Groundwork*, saying that “A *maxim* is the subjective principle of acting, and must be distinguished from the *objective* principle, namely the practical law. The former contains the practical rule determined by reason conformably with the conditions of the subject (often his ignorance or also his inclinations)” (G 31n, 4:421n). Likewise, the *Metaphysics of Morals* has it that “A rule that the agent himself makes his principle on subjective grounds is called his *maxim*” (MM 17, 6:225).

Now, these passages certainly do not say outright that maxims and practical laws must contain grounds. However, the “Definition” in the *Critique of Practical Reason* is much more decisive:

Practical *principles* are propositions that contain a general determination of the will, having under it several practical rules. They are subjective, or *maxims*, when the condition is regarded by the subject as holding only for his will; but they are objective, or practical *laws*, when the condition is cognized as objective, that is, as holding for the will of every rational being. (CP 17, 5:19)

Here, Kant explicitly contrasts maxims and laws based on the nature of “the condition” and on the manner in which that condition is represented—in Kant’s words, by how that condition is “regarded by the subject” or “cognized”. Specifically, the difference has to do with whether the condition is cognized as valid either subjectively or objectively: “as holding only for his will”, or “as holding for the will of every rational being.”

Accordingly, the best reading of this “Definition” would hold that “the condition” is a part of what the principle represents. How else could Kant use this difference in how the condition is cognized to distinguish maxims and laws? Moreover, Kant’s Definition ends up implying that these conditions are grounds, in Kant’s technical sense. The important point about these conditions, the one that serves as the basis for distinguishing maxims and practical laws,

even, has to do with their validity: with whether the condition “holds” only for an individual will, or for “the will of every rational being”.¹⁵ So Kant is saying that the difference between maxims and laws has to do with how the condition is represented, and, specifically, with the domain of agents for whom the condition is supposed to hold. Thus, the implication is that those agents are represented as having to act as the principle specifies whenever the relevant condition “holds”. These conditions then turn out to be grounds, in Kant’s technical sense. The idea that practical principles “contain a general determination of the will” suggests the same thing as well, since a determination of the will presupposes a ground for that determination, for the reasons that I explained above.

As such, Kant’s explications of practical principles provide at least some direct, independent textual evidence that all practical principles have the same basic structure that I have been describing. And when one explication directly treats practical principles as containing grounds, while the others are suggestive, but less decisive, it would be fair to treat the former as reflecting Kant’s full views, especially when this explication is explicitly given pride of place as a “Definition” in a critique of reason.

Still, Kant’s main purpose in these explications is to contrast maxims and practical laws, which is why he does not say very much about their structure. So even if these passages do not support a positive case for my interpretation, as in the analysis just given, the least that we can say about these explications is that they do not contradict my reading, either. For Kant appears to presuppose that maxims and practical law share a basic structure, as practical principles, and the distinctions that he draws between them do not require them to have different structures from each other, or to deviate from the structure that I propose.

To start with, Kant says that maxims are the principles that agents act on, while practical laws are the principles that agents ought to follow. However, whether an agent does follow a principle or ought to follow a principle is a modal distinction: it is the distinction between a

¹⁵ See note 10 above for an important clarification, however.

principle that is *actually* followed and one that it is objectively *necessary* to follow. Accordingly, this difference does not require an underlying difference in the structure of the principles, at least when all that we mean by their structure is their general components and the general relation between them. To be more specific, despite this difference between maxims and practical laws, both can still represent only a ground, an action, an effect, and a way in which the latter two are necessary relative to the former. Nothing in that structure says anything about whether the principle will be followed or whether it is objectively necessary to follow it, nor does anything about this modal difference force either maxims or practical laws to have a different structure from each other, or a structure other than the one that I propose.¹⁶

The same sort of analysis also applies to whether a principle is objective or subjective. In the terms of Kant's explications, the difference here is one of validity: a subjective principle is valid for the agent who acts upon it, while an objective principle is valid for all rational beings whatsoever, at least of a certain kind. (Remember that principles of happiness can be objectively valid, at least in principle.) As such, this distinction likewise does nothing to force a difference in the basic structure of maxims and practical laws. Instead, it has merely to do with the specific nature of the ground represented in the principle and the corresponding kind of practical necessity in the action and the effect. If the ground is valid for a subject, at least as reflected in their action, then the principle is their maxim, while if it is valid for all rational subjects, whether they act on it or not, then it is a practical law.

To put the same point a little differently: in the terms of Kant's categories, the difference between subjective and objective principles is a difference in the *quantity* of the principle. In the

¹⁶ Of course, we can consider a modal difference in types of principle to be a difference in their structure if we please, but that would not affect the crucial point, given how we are thinking about the structure of a principle. In other words, a modal difference might require a difference in their specific structures, but not in their basic structure. The same goes for Allison's distinction between first-order and second-order practical principles (1990, 87). The difference between acting straight-away and adopting a maxim is a difference in the action that a principle would have you undertake, or perhaps in the effect to be accomplished, but not a difference in the basic structures of the principles. Likewise, *contra* Aune (1979, 24-25), neither kind of principle need specify the agents it is to apply to, at least not in their basic structures. More on this presently.

Second Critique's "TABLE of the categories of freedom with respect to the concepts of good and evil", Kant explains that the categories "Of quantity" are "Subjective, in accordance with maxims (*intentions of the will* of the individual)", "Objective, in accordance with principles (*precepts*)", and "A priori objective as well as subjective principles of freedom (*laws*)" (CP 57, 5:66).¹⁷ And, helpfully, despite how the Table "is intelligible enough in itself", Kant still stops to explain the categories of quantity, saying

Thus, for example, one knows at once from the above table and its first number where one has to set out from in practical considerations: from the maxims that each bases on his inclinations, from the precepts that hold for a species of rational beings insofar as they agree in certain inclinations, and finally from the law that holds for all without regard for their inclinations, and so forth. (CP 58, 5:67)

Kant's elaboration here clarifies the sense in which these are to be categories of "quantity": the distinction has to do, as indicated, with the set of rational beings for whom the principle is valid, be it "from the maxims that each bases on his inclinations" or "from the law that holds for all without regard for their inclinations". Therefore, as with the modal distinction just discussed, there is no need to treat the distinction between subjective and objective principles as necessitating a difference in the general components of those laws, or in the general relation between those components. Instead, the difference in quantity can be a difference in their specific structure.¹⁸ So, at minimum, Kant's explications of practical principles do not create any trouble for my interpretation of their basic structure.

Except for one possible issue: I have been arguing that all practical principles represent laws of freedom, and that laws of freedom are a kind of causal law. However, Kant consistently explains that maxims are to be subjective principles, which might suggest that they cannot

¹⁷ Here again, observe how laws are both objective and subjective, which creates a strong impression that practical laws can be maxims, in Kant's mind. This is worth keeping in mind for the quotation just below, as well.

¹⁸ Likewise, when the difference between maxims and practical laws is understood as a modal difference and as a quantitative difference, it become quite difficult to see why practical laws cannot be maxims. For there is no problem in saying that a single principle is both actual and necessary, nor is there any incompatibility in saying that one and the same principle is valid both for an individual and for all rational beings, of a certain kind.

represent *laws* of any kind. As I argued in the last chapter, the concept of a law, for Kant, includes a certain kind of strict universality: a law is to hold universally not just in fact, still less as the result of a generalization, but without even the possibility of an exception. So how can a subjective principle represent a law, in that sense? Can't at least some subjective principles hold only for a single agent, according to Kant's explications?

The answer, I take it, hinges on Kant's concept of freedom as an absolutely self-determining causality. Because a free will is absolutely self-determining, each free will chooses which principles to adopt—chooses its maxims—independently of all the others. To be sure, some principles may certainly be valid for me objectively, or even subjectively, independently of my actual choices. But the principles that I adopt for myself, as a free agent, are compartmentalized from the principles that you adopt, as another free agent. Moreover, in many cases I can even choose principles for myself without having to judge that those principles are valid for similarly situated free agents.

Accordingly, Kant can accept laws of freedom and causal laws that hold only for a single free agent, just because freedom involves the capacity for absolute self-determination. Those laws can even have the requisite kind of "strict" universality, because I can adopt maxims that are to hold for all of my actions universally, and without the possibility of exception. In fact, to adopt a maxim, rather than just to form something akin to a rule of thumb, is precisely to take on that kind of universal commitment. Even though free agents can alter their maxims at any moment, they still maintain a strictly universal commitment to their maxims, so long as they adhere to those maxims.

So I must admit that Kant does not usually speak as though there is any such thing as a subjective law of freedom or a subjective causal law. But his conceptual framework does not preclude them, either, and his concept of freedom positively requires space for them.

3.4 Kant's Examples of Practical Principles

Now, as for Kant's examples of practical principles, I explained in the last chapter that he does sometimes give examples that exhibit this ground/action/effect structure quite neatly. The maxim of the lying promise in the *Groundwork*, for example. Others, however, do not: to mention just two, Kant describes a "maxim of neglecting his natural gifts" (G 32, 4:423), and he uses "thou shalt not lie" as a paradigm of a moral/practical law (G 2-3, 4:389). These principles appear to contain only an action, so aren't they counterexamples to my interpretation here?

For now, let's put aside whether or not anyone is actually required to follow these principles, and focus just on what they say. Even so, these principles are still supposed to impose a requirement, either objective or subjective, on our actions: they require us not to lie, or to neglect our natural gifts. We have to do that in order to satisfy the content of the principle.

Thus, the practical necessity that my account requires is still there, at least to the extent that these principles purport to necessitate our actions. And what are these principles really requiring? At minimum, they require us to satisfy their content, which means that they set an end for us, which is to be the effect of our actions: the end or goal of never lying, or of neglecting our natural gifts. (Which is not the same as the goal of preventing lies, for example. The principle may say that no one should lie, but that says very little about how we should treat the lies of others.) We need not explicitly or self-consciously adopt this goal, to be sure. But we must still accomplish that end in order to satisfy these principles.

Moreover, these principles also require us to perform certain acts of volition. We must act in order to satisfy them, in order to fulfill the ends that they set for us. And in these cases, those acts will be acts of volition: we must will never to lie, or to neglect our natural gifts. There are multiple ways of executing this volition, of course: as with the ends themselves, we need not explicitly or self-consciously follow these principles. It would be enough if we obeyed other principles or performed other acts of volition that just happened to result in honesty or in allowing ourselves to deteriorate. It would even be enough if we chose to just do nothing at all. All the same, our actions and volition need to result in fulfilling these ends. So these principles

still require actions of us, even if they are just acts of volition, and even if no specific act of volition is strictly required by the content of these principles.

So we find that practical principles must require both an end/effect and an action as a means to that end, even if no specific “external” action or act of volition is directly required. What about a ground, then, since Kant sometimes includes a ground in the statement of a principle, but sometimes seems not to?

All practical principles represent an action as practically necessary, as I just illustrated. So if they do not explicitly include a ground for this necessity, then they end up representing the action as unconditionally, absolutely, or intrinsically necessary. Categorically necessary, in short, as the form of “thou shalt not lie” or “neglect your natural gifts” reflects, at least when we consider only their surface form. But then the principle in question would just say that it is intrinsically necessary not to lie, that lying is practically forbidden because of something about the nature of lying, perhaps, or because of something about the nature of this principle itself, such as its form. Either way, though, the principle is implicitly its own ground: something about its own nature or form explains why it is practically necessary to follow it. And then nothing prevents us from including this ground in the “deep” structure of the principle.

In fact, a principle must include a ground, or else it cannot capture how it is supposed to necessitate our actions in the first place. So here we have a procedure for finding the full structure of practical principles, which also shows them to have the structure that my view proposes: I pointed out earlier that practical principles are principles for action, either for how we do act, or for how we ought to act. As such, they must require an end or an effect that we are to accomplish. But then we can always recover an action as a means to that end, and both the end and the means will always presuppose a ground from which they are to follow necessarily, at least if the principle is supposed to represent or express a practical requirement.

Still, if I am right about the structure of Kant’s practical principles, it would be fair to wonder about why Kant never explains this structure, and also about why he does not always

give examples exhibiting it. His relative silence here does not contradict my interpretation *per se*. But if this structure is as central to his overall views as I have been making it out to be, it is rather odd that he never explains it more directly, and that he does not hold himself to adhering to it, either. Looseness in his examples would not be so strange if he stated this structure directly, and neglecting to explain the structure would not be so unusual if his examples all clearly had the same structure. But doing neither of these things is, admittedly, rather odd, particularly given Kant's systematicity and characteristic methods.

Since the question here is about why Kant did *not* do something, it is hard to do much more than speculate. So the best answer that I can give is that explaining the full structure of practical principles was just never a sufficiently integral part of Kant's immediate purposes at any specific point in his critical philosophy. That explanation would be most relevant in his critical ethics. There, however, his main goals are to elucidate three things: the fundamental principle of morality, the source of its validity for us, and its connection to autonomy. The structure of practical principles is certainly important to a full understanding of these arguments. But his main ideas and arguments can be conveyed primarily through his analysis of the different kinds of objective validity and through the concept of an imperative, without needing a fine-grained account of the structure of practical principles, of the sort that I am trying to provide.

So Kant's main philosophical goals mostly allow him to leave the full structure of practical principles roughly sketched and largely undetermined, and this appears to be his actual practice. The fact that he neglects to explain this structure is at least some evidence that he viewed it as insufficiently necessary for his main purposes, even if, from our point of view, this might seem to be a miscalculation on his part, and even if a more fine-grained account is necessary for a more detailed presentation of his views and arguments.

A more detailed account is far more useful when it comes to actually applying the fundamental principle of morality. But Kant does not believe that it is necessary for him to

explain that application in any great detail, nor does he view it as the proper role of the moral theorist. He says, of the Formula of Universal Law, that “Common human reason also agrees completely with this in its practical appraisals and always has this principle before its eyes.” (G 15, 4:402) So in general, if we take this kind of remark seriously, he could have plausibly, albeit perhaps naively, thought that he did not need to stop to explain to us the structure of our own principles, rather in the way that he does not deem it necessary to give more than highly schematic outlines of how to apply the Formula of Universal Law.

And as for Kant’s specific examples: he often introduces examples of practical principles so as to illustrate how to evaluate them, morally speaking. But moral assessment and reasoning need not always depend upon the specific nature of every particular component of a principle. It is often enough, for example, to know that the proposed action involves lying, or that the proposed effect involves murdering a human being. In the former case, neither the ground nor the effect need matter to the moral analysis; in the latter, the ground and the action can be morally superfluous. For example, if certain kinds of actions are categorically forbidden just as such, then their ground and their effect are superfluous to their moral evaluation just by definition. This does not mean that the ground or effect can be excluded from the agent’s principle, only that we do not need to rely on their specific natures in order to evaluate the morality of their actions. So the fact that Kant does not always specify a ground, an action, and an effect in contexts of moral evaluation need not imply anything about whether a fully articulated maxim or practical law requires them.

Instead, the fact that Kant does sometimes include a ground in his specification of a principle is strong evidence that grounds must be included in a fully articulated principle. This time, consider the *Groundwork*’s first two examples: suicide and a lying promise. I have already analyzed the latter, but, for the former, Kant says that the “maxim, however, is: from self-love I make it my principle to shorten my life when its longer duration threatens more trouble than it promises agreeableness.” (G 32, 4:422) Strictly speaking, Kant here describes self-love as the

ground for adopting a principle, not as the ground contained in or adopted by the principle. (Though nothing prevents it from being both grounds, of course.) When Kant returns to these examples in the *Critique of Practical Reason*, however, he asks about what would happen “if *everyone* permitted himself to deceive when he believed it to be to his advantage, or considered himself authorized to shorten his life as soon as he was thoroughly weary of it” (CP 60, 5:69).

Because Kant is asking here about what would happen if everyone acted a certain way, he clearly means to be articulating maxims. Moreover, both maxims clearly contain grounds for the action that is to be evaluated, in the form of beliefs about advantage and weariness of life. So if grounds were not to be included in the basic structure of practical principles, it would also be quite odd for him to occasionally include them in his examples. And especially so in contexts where the moral analysis of a principle *does* hinge on the nature of its ground, as it does with the maxim of the suicide, among others. After all, his reasoning in the case of suicide is that a universal law must contradict itself, if it is based in self-love, while also requiring the destruction of that very self (G 31-32, 4:421-422). Kant’s reasoning here has often been found unconvincing, but no matter: it is his reasoning, and it depends on the ground of the maxim. And if moral analysis sometimes does depend on the ground in a maxim, then, absent any substantive evidence to the contrary, we ought to expect all principles to include grounds.

So whether or not Kant explicitly says so, his arguments are strong evidence that grounds need to be included in the basic structure of practical principles. And then my whole interpretation follows, as I have illustrated. Moreover, when Kant’s procedure is odd whichever way we read him—either as leaving out a crucial explanation or as including grounds in practical principles when they are not to be part of their basic structure—then we have to revert to the sort of systematic considerations that I began with, if we are to understand him properly.

3.5 Conclusion

Still, the same residue of strangeness remains. If the analyses in the current and previous chapters are correct, then Kant’s views in three crucial domains—on freedom, on causation, and

even on morality itself—all lead to the idea that practical principles are representations of a necessity in the relation between a ground, an action, and an effect. And some of Kant's explications and examples exhibit this structure, too. But some do not, and he never explicitly identifies this structure for us, either.

Moreover, in the next chapters I am going to go on to try to explain how giving practical principles this structure will help us to understand and answer objections to some of the other fundamental elements of Kant's moral theory. So if my overall account of Kant's moral theory is correct, then there is an unmistakable lacuna right at the center of his theory, where an explicit account of the basic structure of practical principles ought to be. That remains quite odd, whatever the explanation. So the best that we can do is to look at the lines leading into that gap, and then at the lines leading out of it, and try to infer what must be going on inside to connect the two. And that has been, and will continue to be, my project here.

To conclude, I want to briefly compare my interpretation of the structure of Kant's practical principles with a representative sampling of the other views in the literature, and then quickly tie my account together. The next chapters can then discuss alternate views of maxims in more detail, and apply my interpretation to some problems for Kant's moral theory.

One of the odd features of the literature on Kant is that there are a lot of different views about the structure of practical principles, but relatively little extended discussion of which view is correct and why. For a representative sample of the views on the structure of maxims: Henry Allison writes that maxims have “the form: When in S-type situations, perform A-type actions.” (Allison 1990, 89-90)¹⁹ In contrast, Christine Korsgaard claims that maxims “usually have the form “I will do Action-A in order to achieve Purpose-P.”” (Korsgaard 1996, 57-58) And Rawls

¹⁹ Allison is actually a little unclear about what should be included in the structure of maxims. He sometimes suggests that an agent's interests can, or even should, be included in their maxims (Allison 1990, 90). I will discuss this point in a little more detail in Chapter 4.4, when we will look at Allison's views in more detail.

describes maxims as stipulating that “I am to do *X* in circumstances *C* in order to bring about *Y* unless *Z*.” (Rawls 2000, 168)²⁰

Despite such differences, however, most commentators focus on more general issues about what maxims are, rather than on which of these specific accounts is correct. And it should be clear that I disagree with these alternative accounts more on grounds of completeness and precision than on grounds of positive error. Instead, I think that my own account reflects how these other accounts are all partially, or even largely, correct. My interpretation corrects theirs mainly by trying to supplement, complement, and clarify them.

For it should be clear how my interpretation overlaps with Allison’s, Korsgaard’s, and Rawls’. The primary difference is that I include grounds, where Allison has situations, and where Rawls has circumstances. This difference might look subtle, but, in my view, it is essential. Putting situations or circumstances in the structure of practical principles is close to correct, but how are we supposed to specify the circumstances? In the normal, unanalyzed meaning of “circumstances”, some circumstances and some aspects of circumstances are relevant to the action and some aren’t, morally speaking, instrumentally speaking, and in terms of the agent’s own understanding of what they propose to do and why. So a bare reference to circumstances or situations is pretty clearly too vague and too general.

My account rectifies that imprecision by stipulating that the principle must instead include a ground, in Kant’s technical sense. These other views are certainly worth deeper discussion, however, if only to see how deep these differences go. But even if my account is not in any deep disagreement with Rawls or Allison, it is still marginally clearer, especially in terms of the general statement of the basic structure of practical principles.

²⁰ I am unclear as to why Rawls thought the “unless *Z*” clause is necessary. I might have thought that the specification of the circumstances would render such a clause otiose. Oddly, Rawls might agree: the “unless” clause seems to disappear from the analysis, once we move from a universalized maxim to considering this maxim as a law of nature.

In any case, centering Kant's notion of a ground is the distinctive feature of my whole approach here. The concept of a ground for action is really what all of the foregoing considerations about freedom, causation, incorporation, pure practical judgment, and about Kant's explanations and examples all converge on. And, accordingly, practical principles turn out to be representations of a certain kind of ground/consequence relation.

Therefore, when Kant distinguishes maxims and laws on the basis of their different kinds of validity or practical necessity, he is also distinguishing them on the basis of the kind of ground that they involve. The ground is what is to make the action and the effect practically necessary. So if the resulting principle is to be subjective, it must be because the ground is subjectively valid, and because that ground makes its effect and action practically necessary, but only subjectively. And similarly for objective principles.

And we saw earlier that Kant calls subjective grounds "incentives", and that he sometimes calls objective grounds "motives". So we can also say that maxims consist in an incentive, an action, and an effect, while laws consist in a motive, an action, and an effect. And these principles represent the relation between these terms, which binds them together into a single, unified whole.

Moreover, at the end of the last chapter, we saw that subjective validity derives from an end that an agent actually accepts. That end will then turn out to be the ground of their action, and Kant thinks that much, if not most, of our agency has this structure, where we act in order to achieve an end, because we have adopted that end. He even says that the "*faculty of desire is a being's faculty to be by means of its representations the cause of the reality of the objects of these representations.*" (CP 8n, 5:9n; MM 11, 6:211) So we represent something to ourselves, adopt that thing as the end to be accomplished by our actions, and then act in order to make that thing real. That action is then subjectively necessary, relative to the end that we have adopted, which then serves both as the end and as the ground of the action.

Objective validity, by contrast, is a matter of whether an agent must act a certain way, either through a natural necessity, an instrumental necessity, or because of “reason itself”—because acting another way would force them to contradict themselves. The sources of these necessities then count as objective grounds—as motives.

Thus, as Kant insists, maxims turn out to be principles according to which an agent actually acts, while laws are principles according to which an agent ought to act. However, deviance from a principle remains possible in both cases. We do not always do as we ought, to be sure, and *akrasia* and subjective irrationality can prevent us from living up to our maxims, as well. To adopt a principle is no guarantee that we will always follow it, and violating a principle does not imply that we have relinquished it entirely, either.

Now, in Kant’s technical vocabulary, an effect that is to be achieved by an agent can also be called an “end” (G 36, 4:427; MM 15, 6:222; MM 146, 6:381). Strictly speaking, an end is a type of effect. An end is an effect that is related to the will, and, more specifically, represented as something that can be brought about through volition and action. To adopt an end is to represent it as something that you ought to bring about, or that it would be good to bring about. So it is actually slightly more precise to speak of ends here, rather than just effects.

In any case, though we can summarize the interpretation defended here in this way:

- (1) In general, practical principles are representation of a necessary relation between a ground and its consequent action and end (effect).
- (2) Maxims are representations of the (subjectively) necessary relation between a subjectively valid ground (an incentive) and its consequent action and end (effect).
- (3) Practical laws are representations of the (objectively) necessary relation between an objectively valid ground (a motive) and its consequent action and end (effect).

To be sure, I have not said anything yet about what kinds of things can be grounds, actions, or ends, for Kant. But the important thing to understand here is that practical principles are primarily principles for the will. So grounds, actions, and ends are all relational or functional

roles, relative to volition. An action is something that the will does, to achieve a certain end, and a ground is to make it necessary for the will to perform that action, to achieve that end.

So in general, an end is whatever is to be brought about through action, an action is whatever is to bring an end about, and a ground is whatever makes the former necessary. The question is really about what kinds of things can fulfill these roles, and so far there is no reason to be particularly restrictive about any of them. Ends and actions, for example, could be states of affairs in the “external” world, the world beyond volition itself, or they could be purely “internal” to volition itself. And as for grounds: this question is really asking about the kinds of things that can generate practical necessities. From the subjective point of view, this is only a matter of the kinds of ends that we can adopt, and of what these ends can commit us to. There is also no real puzzle about happiness as a ground, or about means/ends necessities. “Reason itself” as a ground, though, is rather puzzling. So let us now turn to Kant’s views on the subject, which finally leads us to the Formula of Universal Law itself.

First, though, we should consider some alternative views on Kant’s maxims.

4. Alternative Views of Practical Principles

Now, having completed the case for my interpretation of Kant's practical principles, we should consider some of the other views in the literature on Kant. Then we will be able to discuss Kant's main arguments about the Formula of Universal Law, beginning in the next chapter.

As I mentioned at the end of the last chapter, there is not, by any means, very much agreement amongst Kant's commentators about what practical principles consist in for him. Still less is there anything close to a consensus that the account that I have given is the correct one. In fact, as far as I know, my account is entirely novel, though worked up from preexisting parts, to be sure. For, as I also illustrated at the end of the last chapter, the accounts of practical principles in the literature on Kant's practical philosophy all agree with my account in some ways and to some degrees, while disagreeing with it in others.

It would behoove us, then, to canvass a few of the more prominent lines of interpretation, so as to see, from the perspective of Kant's texts themselves, what they get right and wrong about Kant, independently of their disagreements with my account. Seeing how these other interpretations fail to give the best reading of the texts will provide some further indirect support for my account, as will the ways in which my account allows us to offer new solutions to some important problems for Kant's moral theory, which we will explore over the next chapters.

The bulk of the discussion in the literature focuses on maxims, rather than on practical principles in general. So that is how we will proceed here, also. Since practical laws are structurally analogous to maxims, contrasting some different accounts of maxims will at least convey the main contours of the broader disagreements about practical principles in general.

Specifically, let's look at three common ideas about maxims. The first holds that maxims are a kind of intention. The second, inspired by Rüdiger Bittner and Otfried Höffe, treats maxims as a kind of policy or "rule of life". Third, Allison argues that maxims are "first-order" principles for our choice of actions, rather than for our choice of principles. After discussing these views, we can then conclude by looking more closely at some formalized treatments of the structure of

maxims, like those of Allison, Korsgaard, and Rawls, which we examined briefly at the end of the previous chapter.

4.1 Maxims and Intentions

As far as I know, the idea that maxims are a kind of intention does not form the basis of any sustained interpretation of Kant's notion of a maxim. Onora O'Neill comes close: in both *Acting on Principle* and *Constructions of Reason*, O'Neill glosses maxims as intentions throughout her discussions of how the Formula of Universal Law can guide our actions. However, in both texts she is also clear that references to maxims as intentions can be eliminated in favor of references to certain kinds of practical principles instead (O'Neill 2013, 16; 1989, 84n).¹

But even if Kant's maxims are not ultimately understood to be a kind of intention, this idea is still worth discussing, because of how common it is as a rough gloss or a first approximation of Kant's thinking. To begin with, then, treating maxims as a kind of intention is helpful, at least to a certain extent. Kant's view is that the agent's volition is to be the locus of moral analysis and evaluation. (That should be abundantly clear in the Formula of Universal Law itself.) And what we intend to do is at least a close neighbor of what we will to do. So treating maxims as a sort of intention puts those loci in the right neighborhood, at least.

Moreover, intentions are integrated into our practical reasoning, too, at least to the extent that our intentions help us to structure our instrumental, means/ends reasoning. Intending to do something helps us to work out what we need to do achieve that thing, after all. As principles for our actions, maxims are also to play a role in our practical reasoning: our reasoning about how to satisfy them, for just one primary example (G 24, 4:412). So treating maxims as a type of intention can also indicate the role that maxims are to play in practical reasoning, at least in part. (More on this presently.)

¹ O'Neill also discusses specific intentions ancillary to maxims. Those can remain, of course, even when we give up on the idea that maxims themselves are a kind of intention.

Last, there is also some precedent in Kant's texts for understanding maxims to be a kind of intention. In the last chapter, for example, we discussed the Table of the Categories of Freedom from the *Critique of Practical Reason*. There, we noticed that the first practical category of quantity is the concept of subjective quantity: "Subjective, in accordance with maxims (*intentions of the will* of the individual)" (CP 57, 5:66). Here, Kant appears to explicitly and directly treat maxims as a kind of intention. But it is crucial that Kant is speaking of intentions "*of the will*" here, rather than just intentions *simpliciter*. That suggests that the concept of volition should still have primacy in interpreting maxims, rather than the concept of intentions.

Accordingly, as a rough gloss or a first approximation, we could certainly do worse than to compare maxims to intentions. And as long as that comparison is restricted to a first approximation, or used only as an entry point into a deeper understanding of Kant's maxims, there is not much to object to in it. As the basis of a deeper explication of Kant's thinking about maxims, however, treating maxims as a kind of intention quickly becomes more or less totally inadequate.

Onora O'Neill herself does a very thorough job of outlining some of the immediate points of tension between the two notions (2013, 13-16; 1989, 84), so I will not belabor the point much further here. We should add, though, that the main problem with this line of interpretation would be that 'maxim' is a Kantian term of art, which is also obviously technical, and which, as such, stands in obvious need of philosophical and textual analysis. 'Intention', however, needs just as much philosophical analysis, if not more, while also carrying the risk of the illusion of understanding, due to the role that 'intention' plays in common parlance and folk psychology. As a result, I often suspect that it would be better to just explain maxims with "subjective principle of volition" and leave it at that, rather than gloss 'maxim' with 'intention', even if the comparison can be helpful as a rough approximation of what Kant has in mind.

In that spirit, I will not venture anything like a full account of intentions here, either in themselves or for Kant. But we usually seem to speak of intentions in four contexts: when we want to describe what we intend to accomplish, what action we intend to perform, how we intend to perform that action, or whether an effect, an action, or a way of carrying out an action is intentional.² However, when we talk about our principles, we do not usually speak of intentions or of the intentional, except in describing what we intend to do as a consequence of our principles. So intentions appear more to do with how we execute or apply our principles, rather than with our principles themselves, or with how we come to adopt our principles.

In Kant's vocabulary, then, intentions seem to be better thought of as practical representations in these four domains. Accordingly, intentions would be a certain kind of representation of ends, of actions, or of ways of carrying out actions. Moreover, actions, etc., would seem to be intentional or unintentional only relative to those kinds of representations. In no case, then, would an intention itself count as a principle of action. Understood in these terms, intentions could be components of practical principles, and they would certainly be crucial elements in acting on principles, in any case. But an isolated representation of an end, of an action, or of a way of carrying out an action would not be enough to constitute a practical principle, either for Kant, or in whatever common sense of the term there might be.

In fact, even a single, unified representation of an end, an action, and a way of performing an action would still not be enough to constitute a practical principle. For such a representation does not even seem to be enough to *determine* a practical principle, let alone constitute one. Instead, it seems pretty clear that many, if not most, intentions are compatible

² Anscombe famously starts off *Intention* by mentioning three cases where we use 'intention': expressing an intention to do something in the future, describing an action as intentional, and asking about the intention with which something was done (1963, §1). And she proceeds to focus her analysis on intentional action, which she wants to understand in terms of a particular sense in which we can ask about why an action is performed (1963, §5). Actions are intentional, she claims, under the descriptions answering this specific "why" question. I think the contexts mentioned above cover the contexts that Anscombe adumbrates. And answering her "why" question is close to asking about a maxim, but not quite the same. If you ask me why I am pumping water, and I say "To poison the people in the house", this answers your question in the right way for Anscombe, but I have still told you very little about what my maxim might or might not be. More on this in §4.4 below.

with any number of different maxims, no matter how you wish to understand what maxims consist in.

For example: suppose that my intention is to get a loan from you, or to help others when in need, as far as I can, or to always treat others with respect. And fill those intentions out with the missing ends, actions, and ways of performing those actions as necessary and however you like. What would those intentions tell us about my principles?

Very little, it seems, since any of those intentions could be derived from any number of underlying principles. It is true that some of these intentions might seem to be principles in their own rights. In particular, the intentions to help others and to treat others with respect appear to have the proper sort of form for a categorical principle. But there are still many different grounds for adopting precisely these intentions. For example, besides the perception of moral duty, I could intend to help others out of genuine fellow-feeling, or out of considerations of prudence. Each of these grounds will lead to different patterns of action under this intention, and to different patterns of success and failure relative to the intention. As such, it would be natural to describe these same intentions as following from different principles in these cases, and we should expect Kant's practical principles to agree. His example of the honest shopkeeper is a helpful illustration of precisely these points (G 11, 4:397).

However, if we are to interpret maxims as a kind of intention, then we will need to be able to recover a single maxim from any intention of the proper kind that we might stipulate or propose. So the fact that intentions of all of these kinds are compatible with many different maxims implies that we cannot identify the two in any very straightforward way.

Granted, Kant certainly believes that we are not always in a position to properly identify either our principles or our intentions (G 19-20, 4:407). In practice, then, we will often be unable to recover our maxims from our intentions, and *vice versa*. Still, though, once an intention of the right kind is determined, or even just stipulated, we will need to be able to determine the corresponding maxim, under the assumption that maxims just are that kind of

intention. Practical difficulties in identifying our exact maxims and intentions to any arbitrary degree of certainty do not alter that underlying philosophical point.

Despite this, it will certainly be the case that there is an intention corresponding to every maxim that we adopt. If my principle is to tell a lie in order to get a loan whenever I need money, then I necessarily intend to tell a lie in order to get a loan whenever I need money, in at least some normal sense of ‘intend’. But the root point here is that intentions, as we just described them, have to do with how actions are executed, and so occupy a separate space from the principles that we adopt in order to govern our actions. The action, end, or way of acting is intended or intentional: that all pertains to the execution of the action. As such, it should not be much of a surprise to find that different principles can be executed in the same fashion. But then those principles cannot be identified with a kind of intention, at least not in any very straightforward way.

None of this is to say that principles are the function of *Wille* (in the narrow sense), while intentions are functions of *Willkür*. For Kant would deny that (MM 18, 6:226). But we are still working with a distinction between legislating action and executing it, which implies that maxims cannot be identified with a kind of intentions. Not easily, in any event.

4.2 Maxims and Policies

Another common gloss, which actually is the basis of a sustained interpretation of Kant’s concept of a maxim, holds that maxims are a kind of policy that we adopt in order to govern our actions. For example, as I mentioned at the start of this chapter, Rüdiger Bittner and Otfried Höffe are two of the main proponents of this line of interpretation, and they hold that maxims are “rules of life” or “*Lebensregeln*”.

The intuitive connection between principles and policies should make the appeal of this line of interpretation clear. In fact, this line of interpretation seems to have all of the advantages of the “maxims as intentions” gloss, while at the same time giving some intuitive insight into the structure of a maxim. For thinking of maxims as policies seems to preserve the connection to

volition and practical reasoning, while a policy also sounds as though it has to be a policy to act a certain way, in whatever conditions are specified within the policy. Moreover, it would be natural, though by no means required, to suppose that the purpose of the policy is itself a part of the policy, at least as far as reasoning about and from the policy is concerned.

In fact, I think that comparing maxims to policies is perfectly acceptable, even if it is a bit too indeterminate about the structure of maxims. But this comparison is acceptable only if we insist that maxims can be policies at any level of generality. To be more specific, a “maxims as policies” interpretation is acceptable, but only so long as it allows us to have maxims/policies for anything ranging from a single action, to the entire course of our lives, to even more general ideas about how to adopt maxims/policies for the entire course of our lives. For nothing in Kant’s official explications of maxims gives even the slightest indication that maxims must stay at a certain level of generality. For example, remember that the *Critique of Practical Reason* says that practical principles must contain “a general determination of the will” (CP 17, 5:19). But nowhere does Kant say that these determinations must be at any specific level of generality. Moreover, the idea that rational agents are free and autonomous would seem to imply that they are capable adopting both maxims and policies at any level of generality.

However, the main attempts to give sustained “maxims as policies” interpretations of Kant usually want to insist on keeping maxims at a specific level of generality. Actually, in certain senses, this is the entire *raison d’être* for such interpretations. For the main “maxims as policies” interpretations want maxims to be principles for the entire course of our lives, or for our lives considered as wholes—maxims are the aforementioned “rules of life” or “*Lebensregeln*”, as these accounts would have it.

To be sure, these *Lebensregeln* views of maxims do not require us to hold our maxims for our entire lives, since we will always remain free to relinquish or alter our maxims. Nor must our actions cleave to our maxims strictly and universally throughout our lives, either. We ought to adhere to our maxims, at least subjectively speaking, but failure to live up to our principles is

not always the same as relinquishing them. Instead, what is essential to maxims, under the *Lebensregeln* views, is just that maxims are to represent our views or commitments about how the entire course of our lives ought to go—about the entire kind of life that we commit ourselves to leading through our maxims. So whether we change our minds about that, or whether we sometimes fail to live up to our commitments—that can often be immaterial to determining our actual maxims.

With all of that in mind, the main sources for these *Lebensregeln* views are Rüdiger Bittner's "*Maximen*" and Otfried Höffe's "*Kants kategorischer Imperativ als Kriterium des Sittlichen*". I will focus on Bittner here, partly due to space constraints, and partly because much of the discussion here will apply to the *Lebensregeln* approach in general, and not just to Bittner's specific version of the view.³

For *Lebensregeln* views, crucially and fundamentally, want to distinguish maxims from precepts (*Vorsätze*), and this distinction is to be a distinction in their level of generality.⁴ For simplicity, I will construe maxims and precepts as different kinds of practical principles, though I think nothing will hinge on this convenience. And as Bittner puts it, the difference between maxims and precepts is supposed to be that "maxims are more general than precepts; the question is, more general in which sense." (Bittner 1974, 487)⁵

³ As far as I know, there is no published translation of "*Maximen*", so all translations are my own. I have prepared a full, but quite rough translation of the entire article.

⁴ Allison invokes the maxim/precept distinction also, and I will have some things to say about his version of that distinction in the next section. But this section will leave little surprise as to general stance I will take towards it there, too.

⁵ Throughout this section, parenthetical citations will be to this article of Bittner's, but I will cite only page numbers, for simplicity. Interestingly, though, Bittner also claims that principles can be too general to be maxims, if they are too general to determine a single coherent course of life. As he puts it, "'I will become happy' or 'I will to live in accordance with nature' are no maxims, since they contain under themselves entirely different lifestyles. Whether wealth or popularity makes one happier, whether moderation or every greatest pleasure is the natural thing, still stands in question. We cannot say that someone has a representation of what their life should look like, that they know what they will, when they only know that they will to live happily or in conformity with nature. The ascent to still more general propositions lets the substance of maxims, namely, to be determinate principles of a life, slip through one's fingers at the same time." (490) There is no need to dwell on this complication, however. If the contrast that Bittner finds

Bittner admits that this distinction between maxims and precepts will not always be sharp in practice (488). But he does think that this distinction can be limned with examples, so that its general contours, and perhaps even its underlying criterion, will be clear (487). So, as examples of precepts, he cites “I will eat with friends every Monday evening” and “to get up every morning at seven o’clock” (487). And, for examples of genuine maxims, he suggests “continuous work for the enlargement of my wealth” (488) and “suffer no unavenged insult” (486), the latter of which Kant himself cites, too (CP 17, 5:19). On the basis of such examples, we are supposed to notice that maxims are more general than precepts (487), and the idea that maxims are *Lebensregeln* is then supposed to explain this difference.

For, to start with, Bittner insists that the greater generality that he finds in maxims cannot consist in covering more cases, or in holding for a longer period of time (487). For example, he points out that you can surely live more days than you suffer insults in need of avenging. So a precept about getting up in the morning can, and usually will, come into effect much more frequently than a maxim about avenging insults. Moreover, I can hold a precept about getting up every morning at 7:00 for exactly the same period of time that I hold a maxim about enlarging my wealth, and quite possibly even longer.

Instead, Bittner thinks that my maxim is more general than a precept because my maxim “has as its object the type and direction of my life as a whole” (488). As he puts it, summarizing his *Lebensregeln* view,

Maxims are rules of life: they pronounce what I will be as a person—one who no one can insult with impunity, or one who is not worried by the needs of others, a life of greed, or a life of enjoyment. They contain the meaning of my life, if “meaning” is taken not as transcendental fulfillment, but rather simply as the fashion in which I think of my life as a whole: “meaning” not as goal, but rather understood as the sense of direction. (489)

between maxims and precepts cannot be made out, then there will be little more hope for this contrast, either. Moreover, many of the same points will apply to either distinction.

Bittner also says that maxims are “the design of an entire life in a determinate purpose” (490), that they are “formed as a philosophy of life” (489), and that they express “the subjective representation of a good life” (489).

So, according to Bittner, maxims are more general than precepts, because the former have to do with the entire kind of life that we choose to lead, while precepts are for conduct with less searching import, or perhaps for conduct in more specific domains of our lives. (Bittner does not really explain the precept side of the contrast.) Thus, we are supposed to think that Kant’s maxims are rules of life—*Lebensregeln*.

This kind of interpretation has a lot of natural appeal, and even a certain romance to it. Thinking of maxims as *Lebensregeln* nicely captures their weight and existential import, and it also resonates well with much of Kant’s thinking about autonomy. Relatedly, the idea of a *Lebensregel* helps give us some purchase on the questions about how we should go about choosing maxims for ourselves and identifying them in others. Moreover, Bittner does not need to disagree with me about the basic structure of maxims: for all he has said, *Lebensregeln* might have the basic structure that I have been proposing here.

The difference between our views is just that my account has less rigid strictures about what can count as a maxim. As I hope the subsequent discussion will make clear, Bittner and I could have the same view, if he could convince me that only certain kinds of considerations could be the ground (or perhaps the end) of a maxim, while other kinds of considerations support precepts instead.

But why should we think that Kant would distinguish maxims and precepts in the first place? Importantly, Bittner does not explain where his crucial distinction between maxims and precepts is to be found in anything that Kant explicitly says anywhere in his actual texts. Kant does occasionally use terms that can be translated as ‘precept’, though never, so far as I know, with a specifically technical meaning at all, still less with a meaning of the kind that Bittner is

interested in. As we saw in the last chapter, Kant frequently distinguishes maxims and practical laws. But he never says anything like that to distinguish maxims and precepts.

Instead, the closest that Bittner comes to locating this distinction in Kant's text is by suggesting that we can find it by comparing examples of "precepts" with the examples of maxims that Kant actually gives (487). As I just illustrated, that comparison is supposed to allow us to see that maxims are to be more general than precepts, and then the "maxims as *Lebensregeln*" interpretation is supposed to step in to explain that difference.

However, when Kant never explicitly distinguishes maxims and precepts, calling some of these examples "precepts" risks begging the question about whether he would recognize this distinction in the first place. Granted that some of these examples are more general than others, why should that make us think that Kant would want some of them to be precepts and some of them to be maxims? Why don't these examples just teach us that maxims can come at different levels of generality?

In fact, a more comprehensive survey of Kant's examples will confirm that that is actually his view: maxims are just more and less general; they are not distinguished from Bittner's precepts. So Bittner's analysis only favors a *Lebensregeln* interpretation because it selectively ignores several of Kant's crucial examples.

Primarily, consider the maxims of the lying promise, and of embezzling deposits. As I have been emphasizing, the former says "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen." (G 32, 4:422) And, for the latter, Kant asks "whether I could through my maxim at the same time give such a law as this: that everyone may deny a deposit which no one can prove has been made." (CP 25, 5:27)

Now, these principles have to be operating at the same level of generality. They are, after all, a kind of inverse of one another: one is about lying to get a loan, while the other is about lying to avoid returning a deposit. And Kant explicitly identifies both of these principles as

maxims. Bittner, though, is committed to making them precepts, since he claims that Kant made a mistake in treating the embezzling deposits maxim as a maxim (487n).

That allegation is quite hard to accept, however, when Kant never explicitly contrasts maxims and precepts in the first place. In any case, though, it is absolutely impossible to think that Kant could be making that kind of mistake with the maxim of the lying promise, since that maxim is to be one of his four fundamental cases for illustrating how to apply the Formula of Universal Law in the first place. So given that these principles are equally general, it is just impossible to deny that both principles should get the same treatment as maxims, rather than as precepts.

Bittner's account, though, really does need these principles to be precepts, not maxims, as he implicitly recognizes by wanting Kant's treatment of the maxim of embezzling deposits to be mistaken. For, to say the least, it is quite difficult to understand such maxims as invoking anything quite so grand as the "meaning" of my life or the "direction of my life as a whole". Under these maxims, I suppose that "what I will be as a person" will be someone who tells lies to get loans, or someone who embezzles deposits when they can get away with it? That hardly seems like what Bittner has in mind for a *Lebensregel*. Or, if it does, then it becomes quite hard to see why a maxim about getting up in the morning cannot be a *Lebensregel*, too. If the maxim of the lying promise says something about "what I will be as a person", then shouldn't the "precept" about waking up every morning at 7:00 say something on the same order? "What I will be as a person": someone who gets up daily at 7:00 a.m.? Don't these principles give a "direction for my life as a whole" in the same kind of way, or on the same kind of level?

At minimum, the maxims of the lying promise and of embezzling deposits can surely be traced back to *Lebensregeln*, like those about maximizing our wealth or happiness. Kant himself puts the maxim of embezzling deposits in precisely such a context. That connection might allow Bittner to count these maxims as *Lebensregeln* after all, but only at the cost of the entire distinction between *Lebensregeln* and precepts. For if a principle can be a *Lebensregel* through

its origin in or through its derivation from a *Lebensregel*, then nothing stops Bittner's precepts from being *Lebensregeln*, too, by parity of reasoning. For we can surely find such an origin or derivation for "precepts" as well.

So if the maxims of the lying promise and of embezzling deposits are really maxims, as Kant says that they are, then either some maxims are not *Lebensregeln*, or the idea that all maxims are *Lebensregeln* yields no tenable distinction between maxims and precepts. (Or both, as I think is actually the case.) Whichever way, though, Bittner's interpretation is just not borne out in Kant's examples.

Thus, the full range of Kant's examples of maxims does not allow us to find a sense in which they are characteristically more general than Bittner's precepts. And poking at cases like these makes it start to seem as if the precept/*Lebensregel* distinction cannot really be made out clearly even in its own terms, apart from whatever basis it might or might not have in Kant's writing. What is the "direction of my life as a whole" supposed to mean anyway, and why doesn't every principle whatsoever provide such a direction?

Bittner, though, does give two more arguments for understanding Kant's maxims as *Lebensregeln*. First, he claims that a "maxims as *Lebensregeln*" interpretation can help us to solve a major problem in how the Formula of Universal Law is to be applied. As an example, he asks us to consider "a rule like: I will eat with friends every Monday evening" (487). If we understand this rule to say that my friends will host me for dinner every Monday evening, then it looks as if this rule cannot work as a universal law. Someone has to do the hosting, after all, so not everyone can visit friends for dinner on Mondays. (And the same goes if my rule is to host my friends, rather than to visit them.) So the Formula of Universal Law would appear to reject this rule. Still, it does not seem as if there is anything morally problematic with following this rule.

So Bittner proposes that we can solve this problem by treating this rule as a precept, not as a maxim (487).⁶ The Formula of Universal Law tells us to act on *maxims* that we can will as universal laws, after all, so it does not appear to say anything at all about following *precepts*. Hence, if we can use the idea that maxims are *Lebensregeln* in order to reduce this rule about my Monday evening dinners to a precept, then it looks as though the Formula of Universal Law won't require us to ask after whether or not we can will for everyone to eat at their friends' on Monday evenings in the first place. And then it looks as though we can get around the problem here, too. The utility of the distinction between precepts and maxims as *Lebensregeln* in solving this problem would then afford us with another argument for attributing this distinction to Kant.

But surely this cure is worse than the original disease. Strictly speaking, Bittner only says that precepts are “not subject to *direct* moral examination from possible universality.” (487, emphasis added) However, Bittner does not explain what it would be for a precept to be subject to the Formula of Universal Law only indirectly. (And never mind how, by Bittner's own lights (487n), Kant applies the Formula of Universal Law directly to the “precept” of embezzling deposits (CP 25, 5:27). That would appear to be another mistake on Kant's part.) Since the Formula of Universal Law only tells us which maxims we are to act on, it seems safe to presume that indirect evaluation under the Formula of Universal Law must go by way of a connection to maxims: to evaluate a precept “indirectly”, in other words, that precept has to be tied back to a maxim somehow.

But the difference between maxims and precepts is supposed to consist solely in the specific way in which maxims are more general than precepts. So the only way to connect a precept to a maxim would be to derive the former from the latter. Precepts would then have to

⁶ Strictly speaking, Bittner seems to want to evolve this solution to this problem out of a prior distinction between maxims and precepts, based on their characteristic kinds of generality. But if that distinction promises to solve a problem like this for Kant's theory, then we are free to understand that solution as yielding an independent argument for that distinction.

be less general principles derived from more general principles, in which case precepts would have to be instances or specifications of maxims. Otherwise there could be no way to evaluate a precept indirectly through a maxim.

Now, though, there is no longer any reason at all to think that precepts are not just a type of maxim. For one thing, the gradations in generality will not admit of any sharp distinction between the two. (As Bittner admits.)⁷ When we were trying to use examples to produce the distinction between precepts and maxims, we could perhaps tolerate vague boundaries between them. But positing a vague distinction in order to solve a problem, as we are now considering—that starts to seem rather *ad hoc*. Especially when deriving precepts from maxims in this fashion annihilates the purported specific difference between them all the more by Bittner’s own lights.

To wit: to solve the problem at issue, we are supposed to posit that maxims are *Lebensregeln*, so that the problematic kinds of rules can be precepts, and not maxims. But an instance of a *Lebensregel* ought to be a *Lebensregel* itself, because an instance or a specification of a *Lebensregel* would have to be an integral part of how we think about our lives as a whole. So, because of how precepts would be integrated into our *Lebensregeln*, precepts would have to count as *Lebensregeln* themselves. And then we would lose the sense in which there is supposed to be any specific difference between maxims and precepts, so long as we grant that the Formula of Universal Law is to evaluate precepts through maxims. So then there would really be little reason to think that the Formula of Universal Law cannot evaluate “precepts” directly.

For what is the point of this “indirect” analysis under the Formula of Universal Law in the first place? What is it supposed to show? At minimum, it would have to show that it is permissible to act on a precept, because its associated maxim can be willed as a universal law. But direct analysis of this precept is supposed to show, *ex hypothesi* of the problem and of the

⁷ Allison, however, seems to think that there is such a thing as a most specific principle, which would imply an action directly, without leaving any scope for practical judgment. And Bittner gives some indication that he would agree with that. If that is correct, then we might be able to draw a more or less sharp distinction between maxims and precepts after all, even on the basis of their levels of generality. I will address this idea in the next section.

strategy of this solution, that this precept cannot be willed as a universal law. So this whole strategy of response hinges on declaring this direct analysis to be inapt or improper. Why should it be, though? If precepts are instances or specifications of maxims, and the only difference between them is their level of generality, then why should it be that maxims are apt for direct analysis under the Formula of Universal Law, while precepts are not? Why does the lesser generality of a precept make them unsuitable for consideration as universal laws? Why should it be inapt for us to consider whether we can will them as such?

If Bittner cannot give a compelling answer here, then his proposal may not even be coherent, since it may have to accept both the direct analysis of the precept and a conflicting indirect analysis. Why, then, should we suppose that the indirect moral analysis is better or more penetrating than the direct one? If the precept is derived from the maxim as its instance or specification, why shouldn't *modus tollens* from the direct analysis show that the indirect analysis must be mistaken? For doesn't the direct analysis of the precept show that the indirect analysis reaches a false conclusion?

So I think we must conclude that whether precepts are subject to the Formula of Universal Law directly or only indirectly can be of no consequence at all. Instead, what matters is only whether or not precepts are *ultimately* "subject to moral examination from possible universality". If they are, then the problem remains, because then they are still ultimately subsumed in the same kind of moral analysis, either directly or through an associated maxim. So if Bittner's proposed solution is to work, he really needs to say that precepts are intrinsically exempt from analysis by the Formula of Universal Law.

But no matter what you want to call them—rules, precepts, what have you—surely the outlook for Kant's moral theory cannot be improved by carving out a class of principles that is intrinsically exempt from analysis by the Formula of Universal Law. By Bittner's own lights, the difference between a precept and a maxim is to be one of generality. As such, the nature of a precept provides no guarantee whatsoever that precepts will be inherently moral. Compare "I

will get up every morning at 7:00 a.m.” with “I will steal from my local convenience store every morning at 7:00 a.m.” or “I will beat my child every morning at 7:00 a.m.”. Kant’s moral theory would fail, catastrophically, if it had to admit those principles as moral, just because they count as precepts and not as maxims. In contrast, Kant’s moral theory would fail in its own terms if it needed to import another standard, independent of the Formula of Universal Law, in order to evaluate precepts. The Formula of Universal Law is, after all, to be the fundamental and supreme principle of moral obligation (G 5, 4:392). Which is just to say that it must be the ultimate standard for all of our action and volition, whether it gets expressed in a maxim or in a “precept”.

And, really, arguing that Kant should admit a class of precepts on the basis that precepts circumvent the Formula of Universal Law ought to have seemed hopeless from the start. Moreover, we probably ought to have expected that the problem about principles like “I will eat with friends every Monday evening” would turn out to be too robust to solve via a technicality about the wording of the Formula of Universal Law anyway. (I will offer another solution in a subsequent chapter.) But solving this problem is almost the entire motivation for distinguishing maxims and precepts. So, in the final analysis, if both maxims and precepts ultimately have to be subject to the Formula of Universal Law, then that by itself should be enough to show that Bittner’s distinction between maxims and precepts cannot really withstand much scrutiny.

Bittner does have a third argument, though, which purports to find a difference in the kinds of considerations that could rationally lead us to change our maxims, as against those that could make it rational to alter our precepts. In essence, Bittner’s argument here is that particular facts can make it rational to change a precept, while only some new “insight” (488) into “what type of life I lead” or into “what I will become or already am becoming as a person” (488) can make it rational to change my maxim. As examples, Bittner says that it can be rational to change my precept about having dinner with my friends on Mondays if I discover something particularly enticing on television those nights (488). By contrast, though, suppose that my

maxim of maximizing my wealth has caused one specific person to dislike me. Bittner thinks that noticing that fact is not sufficient to allow me to change my maxim (488). A particular fact might lead me to violate my maxim, he thinks, but it cannot be enough to make it rational to change my maxim, unless that particular fact reveals something to me about my life as a whole.

And Bittner thinks that we can make parallel points about adopting new precepts and maxims. “New precepts could be supported by experience in the sense of recognition of external facts.” Maxims, though, are “based on cognition that concerns the whole of a life”, on a “life-experience”, or “a comprehensive view on what life in this world is” (489).

This argument makes an interesting, and perhaps powerful, suggestion. If this difference in underlying considerations really does rationalize different kinds of principles, then it does sound as though that should motivate a distinction in kinds of principles. But there are still a couple of problems with this argument.

One problem is that Bittner does not really explain why it should be irrational to change a maxim on the basis of a single, particular fact *qua* single, particular fact. Again, Bittner admits that the distinction between a cognition of a particular fact and a “life-experience” will not be sharp, at least not in practice (488). And he seems to think that experience of particular facts *qua* particular facts can only justify precepts covering a set of particular cases, and not maxims covering the whole of a life (489). That appears to be why he thinks that a particular fact has to reveal something about the way of the world, if it is to justify adopting or changing a maxim.

But it is hard to see why that should follow. Just as a matter of boring old *modus tollens* logic, if I conclude that a principle has a single consequence that I do not or cannot accept, then that should be enough to make it rational to reject the principle. Not necessary, necessarily, just rational. Just as a matter of pure logic, that one case need not reveal anything about any other case at all, still less about the whole way of the world. One false consequence is logically enough by itself to reject its antecedent. So I struggle to see why it should have to be irrational to give up

on a maxim of greed, just on the basis that it forces me to harm or offend one person *qua* that one person. Actually, that seems to be what I properly ought to do.

At any rate, though, I doubt that this distinction in rationalizing cognitions really yields a tenable distinction between maxims as *Lebensregeln* and precepts. For example, let's return to our paradigm precept about getting up every morning at 7:00. Why can't I adopt that precept on the basis of a "life-experience"? For example, can't I have a life-experience corresponding to the adage that "Early to bed, early to rise, makes a person healthy, wealthy, and wise"? Or why can't I adopt a precept about weekly dinners with my friends on the basis of a life-experience about the role of friendship in human life? (Perhaps I just read Books VIII and IX of the *Nicomachean Ethics*?) But now these "precepts" look as though they should really be *Lebensregeln* after all.

In short, it looks as though no principle is intrinsically a precept or a maxim, if the basis of the distinction is supposed to lie in the kind of cognition that can make it rational to adopt or change a principle. Perhaps Bittner would respond to this line of reasoning by suggesting that my life-experience has not really led me to adopt this precept itself, but rather a more general *Lebensregel*, as in the "Early to bed..." adage. For all that, though, my experience could still lead me to adopt the precept about getting up at 7:00, too, and then we would need to consider that precept to be an instance or a specification of my *Lebensregel*. As Bittner himself admits, *Lebensregeln* are not self-interpreting (490-491), so they stand in need of specification. What counts as being "early to rise"? Getting up at 7:00 looks to be a good candidate. But if I adopt this precept as a specification or an instance of a *Lebensregel*, then, as I have been urging, we simply lose the distinction between precepts and *Lebensregeln*. And with it, Bittner's distinction between maxims and precepts is lost again, too.

Bittner and Allison do use the idea that maxims are not self-interpreting in order to offer one more pass at distinguishing maxims and precepts on the basis of their levels of generality. But that approach will be better treated in the next section. For now, though, I conclude that it is untenable to treat maxims as policies at a specific level of generality. I do not think that that

should be such a surprise: it is hard to see how levels of generality could support distinctions in types of principles. There is just too much gradation there, as Kant himself recognizes in the *Architectonic of Pure Reason*, when he objects to using levels of generality to determine the boundaries of metaphysical cognition (CR 696-697, A842-844/B870-873).

4.3 Maxims and First-Order Practical Principles

In Chapter 3.3, I mentioned Henry Allison's view that

Imperatives, whether hypothetical or categorical, are of the wrong logical type to be maxims. They are second-order principles, which dictate the appropriate first-order principles (maxims), rather than themselves being maxims. Consequently, according to this formulation, maxims could conform (or fail to conform) to imperatives or practical laws, but they could never themselves become such laws. (Allison 1990, 87)⁸

As I also indicated there, Allison's view is, strictly speaking, entirely compatible with the account of Kant's maxims that I have been developing, at least as far as their basic structure is concerned. In my terms, this difference between first-order and second-order principles is a difference in the action that the principle requires, not in the structure of the principle. For a principle can tell you to adopt certain maxims, or it can tell you to perform certain actions directly, and neither case requires a change to the ground/action/effect structure of the principle. Nor does Allison's view have to conflict with my account of validity in practical principles. For I can be right about the structure of practical principles in general and about what their validity consists in, and Allison can still be right about what all actual maxims and practical laws are like.

⁸ Aune (1979, 24-25) also holds that the logic of maxims and laws are different, and different in a way that prevents maxims from being laws. He holds that maxims refer only to their specific agent and not to all agents of a relevant kind. In Aune's view of Kant, maxims can be universally valid, of course, because each agent can act on a condition that is also valid for all agents of a certain kind. But maxims still lack a universal quantifier ranging over some class of agents, so they lack the logical structure required to actually be a practical law, under his interpretation. Under my account, though, neither maxims nor laws quantify over agents directly, but only through the kind of necessity that they are supposed to involve. From my point of view, Aune is confusing the logic of the expression of the principle with the logic of the underlying causal law. The logic of the expression makes the quantifier look like the dominant logical feature, when it is actually submerged into the relation posited in the law.

However, Allison's view still faces some immediate difficulties with several of Kant's remarks on maxims and laws. (Which I also noted in Chapter 3.3) And because Allison's account implies that maxims can never be laws and laws can never be maxims, rejecting it will create more space for my account. For if maxims *can* be laws, and *vice versa*, then the two must obviously have the same general structure, as I have been contending. Rejecting Allison's interpretation does not prove that maxims and laws can be one and the same, of course, but it does help allow the natural readings of Kant's remarks to this effect to stand. Rejecting Allison's view will then indirectly support my account.

As far as I can tell, then, Allison does not say very much about why we should interpret Kant his way. He asserts on (1990, 131) that there is an argument in (1990, Ch. 5), but he does not identify it there, and it is rather difficult to locate on his behalf. However, his remarks suggest three possible lines of argument, so let's consider them in turn. The first line derives from the passage quoted above, where the claim seems to be that imperatives are second-order principles, so maxims must be first-order principles.

Now, the Formula of Universal Law is clearly a second-order practical principle, since it makes explicit reference to how we ought to choose our maxims. However, Kant gives perfectly clear examples of imperatives and of moral laws that do not explicitly mention maxims: consider the aforementioned "thou shalt not lie" (G 2-3, 4:439), as a prime example. And that is to say nothing of the Formula of Humanity itself, which reads only "*So act that you use humanity, whether in your own person or in the person of another, always at the same time as an end, never merely as a means*" (G 38, 4:429). So there appears to be no consistent textual basis for thinking that Kant would insist that all imperatives, along with the laws that they express, must be second-order. At least not in the sense of explicitly stating how we are to choose our maxims.

However, Allison could try to get around these examples by arguing that the Formula of Humanity, say, is covertly a second-order principle telling us how to choose our maxims. In which case the real, underlying structure of the Formula of Humanity would have to be

something like “Act only in accordance with those maxims that use humanity, whether in your own person or in the person of another, always at the same time as an end, never merely as a means”.

This would be a good second-order rendering of the Formula of Humanity. Only the problem is that it is too good: in fact, this second-order formulation is equivalent, practically speaking, to the first-order formulation that Kant actually uses. And these formulations would have to be practically equivalent in order for this line of response to work at all. If an apparently first-order principle is to turn out, on further analysis, to be a second-order principle, then the first-order and second-order formulations will have to require us to perform precisely the same actions and acts of volition.

So this is a general truth: if a second-order principle directs you to choose a maxim because of something about that maxim’s “practical content”, then that second-order principle will be practically equivalent to a first-order principle. “Act only according to those maxims that do not include lying” is practically equivalent to “thou shalt not lie”, for example. For both tell you to act the same way, only by different “routes”, so to speak, corresponding to the order of the principle. Instead, the only irreducibly second-order principles will be those that tell you to choose maxims not because of their content, but rather based solely on their forms: such as the Formula of Universal Law and the Formula of Autonomy.

So whatever normative properties a content-directed second-order practical principle has, its first-order practical equivalent must also have, and vice versa. (If the first-order Formula of Humanity is a valid categorical imperative, then the practically equivalent second-order Formula of Humanity must be a valid categorical imperative also, and vice versa.) After all, these normative properties are directed to our actions, so practically equivalent principles must have the same normative profiles. As such, there is simply no basis here for distinguishing imperatives from maxims based only on the order of the principles, through either their explicit structures or their underlying, “deep” structures.

Moreover, Kant's view is that *all* practical principles function to direct our choice of maxims, even if they do not do so through any part of their structure, explicit or implicit. Hence, we cannot separate imperatives from maxims by determining which practical principles function to constrain choice of maxims either.

For one thing, every maxim that we adopt constrains our ability to choose other maxims *eo ipso*. After all, fulfilling our new maxim is now one of our ends, which then serves as the basis for hypothetical imperatives constraining our ability to choose other maxims. For now we cannot choose any maxims that would be incompatible with the one we have just adopted. And none of this implies that the original maxim must itself be an imperative: recall the discussion of hypothetical imperatives in §2.3.

For another thing, Kant believes that the faculty of judgment plays an ineliminable role in our cognition in general (CR 268-269, A132-134/B171-174; CJ 64-66, 5:177-179). For one of Kant's deepest and most fundamental doctrines holds that immediate and mediate representations—i.e. intuitions and concepts—cannot be reduced to each other. As such, a separate faculty—judgment—is always necessary to apply concepts to intuitions and to subsume intuitions under concepts. See, for example (CJ 66, 5:179).

And principles, as I have been arguing, are a kind of cognition, for Kant. As such, we cannot act according to a principle without judging actions in terms of that principle. But even if principles are not themselves cognitions, they are still articulated or expressed in concepts, while actions are particular objects. And Kant's view is that an act of judgment is always necessary to apply concepts to objects.

Thus, every practical principle can always have others subsidiary to it, because there is no point where a concept transitions into or even determines a particular intuition. Instead, each concept invoked in a principle can always be determined further and further, indefinitely, through new principles. Every practical principle, then, implicitly constrains our choice of maxims, because they can always have subsidiary principles under them, and they will

necessarily constrain the choice thereof. All of this follows just from the fact that we can never act from a practical principle except through an act of practical judgment, and none of it implies that there are no subjective principles.

For example: if I make it my principle not to lie, I can still have, and I will presumably need, subsidiary principles about what lies consist in. But those principles will leave open the possibility of further subsidiary principles, and so on, *ad indefinitum*, Kant thinks. And my original principle will impose some logical constraints on the nature of these subsidiary maxims.

As a result, there can never be a “ground level” practical principle that implies an action directly, without the mediation of the faculty of judgment.⁹ Instead, each practical principle can always have further practical principles under it, because practical principles can always be rendered more specific or more precise, without ever directly specifying their application in our actions. As such, each practical principle must constrain our ability to choose other principles, just because their content will constrain how they can be rendered more precise or specific.

None of this is to challenge the existence of first-order practical principles; it is only to reinforce the point that every practical principle places logical constraints on which other principles we can accept, without thereby turning every principle whatsoever into an imperative. Thus, from Kant’s point of view, we cannot separate maxims from imperatives by dint of a functional role in constraining maxim choice, either, since all practical principles serve to constrain choice of maxims, but not all of them are imperatives.

Now, Allison recognizes that practical judgment is ineliminable for Kant when it comes to acting on maxims (1990, 90). As I have argued, that is already enough to make it impossible to distinguish maxims and imperatives by their order. But Allison does not think that judgment is required for action in general, or, at least, he does not think that there always has to be “scope” for practical judgment. Sometimes a principle wholly determines its application, he thinks.

⁹ This is another reason why Kant will not accept the precept/maxim distinction. More on that presently.

For Allison believes, like Bittner, that “it must be acknowledged that the distinction between genuine maxims and mere *Vorsätze* is an important one and that it helps to avoid many common misunderstandings and criticisms of Kant’s moral theory.” (Allison 1990, 92) Here we have our friend the *Vorsatz* or precept returning to the stage, once again in contrast to maxims. And, moreover, the main problem that Allison has in mind here is the very one Bittner raises also, which we already discussed. Allison, though, actually gives us an account of what precepts are supposed to be, instead of just contrasting them with maxims. As Allison puts it, a precept is supposed to be “inflexible” so that it does “not leave such scope” for practical judgment (Allison 1990, 91).

Bittner agrees that precepts have this property also, although he is less clear about whether this feature is to be constitutive of precepts (Bittner 1974, 490-491). So it will be worth digressing (or maybe regressing) for a moment to discuss their views here.

To begin with: Allison’s own examples of precepts all fail to exhibit their purported inflexibility. Rather comprehensively, in fact. He mentions three precepts: “I shall begin each day with 20 push ups”, “I shall never drink hard liquor before sundown”, and “I shall play golf every day at 3 pm” (Allison 1990, 91-92).¹⁰ *Contra* Allison and Bittner, though, these precepts obviously do not preclude scope for practical judgment, just as we saw for my maxim of not lying. To wit: What is a push up? How far apart do my hands have to be? How much range of motion in the descent? What angle must my body make with the ground? Do I have to do all 20 in one set? How much time can elapse between each push up? When does my day “begin”? What counts as “hard liquor”? Does proof matter, or just type? When is “sundown”? Does “drink” require swallowing? What counts as “playing golf”? Mini-golf? The driving range? How many holes? Do I start at 3:00 or do I have to be already playing at 3:00?

¹⁰ Bittner’s example, remember, is parallel: I will “get up daily at seven o’clock” (Bittner 1974, 490). So it is susceptible the treatment I am about to give, too.

I trust you get the idea. These questions obviously modulate what it is to follow these “precepts” and what their success and failure conditions are. In Kant’s mind, though, answering these questions will never prevent similar questions from recurring for the answers. In fact, the fundamental difference between concepts and intuitions implies that questions like these can always recur, indefinitely and ineliminably. According to Kant, the only thing that can ever foreclose such questions is an act of judgment, whereby we finally decide that something like *this* counts as starting my day with twenty push ups, or that something like *that* counts as playing golf at 3:00.

So Allison appears to miss the space for judgment that Kant holds is here, perhaps because, if Kant is right, the judgments in these cases would often seem slight and obvious, and we would usually make them immediately, effortlessly, and constantly.¹¹ But Kant still sees the room for judgment quite clearly: for him, the fact that concepts can be more and less general never implies that they can transcend the “boundary” between concepts and intuitions all by themselves, nor can they ever be so specific that they put judgment “on rails”, so to speak.

For Kant, then, there can be no “ground level”, “most specific” practical principle that is inflexible enough to determine an action without the aid of practical judgment. As a result, each practical principle can always have others under it, and so this pass at distinguishing maxims and precepts fails also.

Allison’s remarks suggest a couple of other possible reasons for treating maxims as first-order principles. One possibility is that maxims must be first-order because they are the principles that agents actually act upon (Allison 1990, 88). However, from my point of view this argument would involve a notion of action that is far too restrictive. Second-order principles are

¹¹ Bittner misses the crucial point, too: he agrees that action according to a maxim always requires judgment—or, as he puts it, “interpreting the current situation” (Bittner 1974, 490). But he goes on: “Action according to a precept is different. When I have made myself get up daily at seven o’clock *and I see it is seven o’clock...my getting up results as the necessary reaction to the legislated condition.*” (Bittner 1974, 490, emphasis added.) Here he simply runs right over the crucial point: seeing that it is seven o’clock already requires judgment and “interpreting the situation” in light of one’s principle. At least in Kant’s view, at any rate.

supposed to direct our choice of maxims. But choosing a maxim is an act of volition, and we have seen no reason to exclude acts of volition from the class of actions. So don't we act on a second-order principle when we choose the maxims that it requires of us? Nor have we seen why acting from a first-order principle cannot count as acting from a second-order principle at the same time. So wouldn't we actually act on second-order principles either way?

In fact, remember how the Second Critique "defines" practical principles in general in terms of a "general determination of the will" (CP 17, 5:19). So if I choose my first-order maxims according to a second-order principle, why aren't I acting on this second-order principle at the same time, since both principles contain a "general determination of the will"? Similarly, why don't I act on this second-order principle as well, when I act on the resulting first-order principles? Moreover, one of the thrusts of the previous discussion is that any constraint on actions is a constraint on maxims at the same time. So any first-order principle also functions as a second-order principle for choosing maxims. So if I then act from a first-order principle, why isn't this also acting on a second-order principle at the same time?

To be a little more precise, the last chapter distinguished two notions of "acting on" a principle: one notion, which only requires satisfying the principle, no matter how, and another notion, which only requires the conscious application of the principle, right or wrong. (We could combine these two notions into a third, but we don't need to do so now.) In the former sense, we can obviously act on second-order principles, since we do that whenever we satisfy the relevant terms. So if Allison wants to distinguish maxims from second-order principles by saying that maxims are the principles that we "act on", then he would need to deny that the conscious application of a second-order principle in our choice of maxims counts as "acting on" that principle. In that case, though, we lose any sense in which second-order principles count as

“practical”, for Kant. And what, then, would be the notion of “acting on”, such that we act on first-order principles, but not on second-order principles?¹²

Moreover, notice that Allison’s view has an immediate corollary: if maxims must be first-order practical principles, then it follows right away that there can be no such thing as a maxim for choosing other maxims. First, note that this consequence is very implausible, even antecedent to any philosophical or textual analysis. Why can’t I adopt a maxim for choosing my other maxims? Nothing in Kant’s examples or official explications shows any reason to doubt this possibility, and it is very difficult to see why an absolutely self-determining, free will should be incapable of having such maxims. A subjective principle for choosing other subjective principles hardly sounds like a logical or conceptual impossibility.

That being said, the main reason for rejecting this corollary, and Allison’s claim that maxims must be first-order principles along with it, is that Kant’s views about high-generality prudential reasoning allow second-order maxims. In Chapter 2.3 above, I mentioned how Kant holds that

One cannot therefore act on determinate principles for the sake of being happy, but only on empirical counsels, for example, of a regimen, frugality, courtesy, reserve and so forth, which experience teaches are most conducive to well-being on the average. From this it follows that imperatives of prudence cannot, to speak precisely, command at all, that is, present actions objectively as practically *necessary*; that they are to be taken as counsels (*consilia*) rather than as commands (*praecepta*) of reason; that the problem of determining surely and universally which action would promote the happiness of a rational being is completely insoluble, so that there can be no imperative with respect to it that would, in the strict sense, command him to do what would make him happy[.] (G 29, 4:418)

To begin with, notice how Kant here uses ‘precept’ (*praecepta*) in a way that is directly incompatible with Bittner and Allison’s usage. Be that as it may, however, the upshot of Kant’s discussion here is that if we want to deliberate about how to become happy at any very general level, beyond the specific desires we have at particular times, then we will have no imperatives to guide us, even if some do exist in principle (as I argued in §2.3). Kant does think that we have an

¹² Thanks to Tyler Burge for pressing me to be elaborate here.

indirect moral duty to care for our own happiness (G 12, 4:399), but this duty does not give us any specific guidance as to what we must do in order to become happy, except perhaps in very specific circumstances.¹³ Otherwise, there are no categorical imperatives of prudence. Nor are there any such hypothetical imperatives, either, because once we begin to abstract from our specific desires, we quickly run into terrain where there is nothing we can know to be “indispensably necessary” (G 28, 4:417) for our own happiness.

Hence, if I want to consider how to advance my own happiness in general, if I want to decide what maxims to have in order to become as happy as possible, I will have clear use for, and clear space to adopt, second-order maxims of happiness. Because I have no imperatives of prudence to rely on, nothing prevents me from adopting maxims about how I am to choose my maxims of prudence, and I will have a clear use for such second-order maxims. The “empirical counsels” that Kant mentions could easily be read this way. But for another example: I might conclude that I should only choose maxims that prioritize long-term happiness over immediate gratification. This is, of course, practically equivalent to just deciding to prioritize long-term happiness. But this second-order maxim has further utility in structuring and guiding my deliberation, and it is perfectly possible for me to follow it throughout my deliberation and subsequent actions. For all that, though, this second-order principle would not have to become an imperative, especially when Kant denies that it ever could become an imperative.

Hence, I can see no reason to suppose that maxims must be first-order principles, for Kant. For it is simply not the case that the notion of a second-order maxim is in any way incoherent, still less volitionally impossible or otiose. For it is perfectly possible to deliberate according to them, choose maxims according to them, and act as a consequence of them. And I have already explained how every principle constrains our choice of actions and maxims. But that clearly does not imply that every principle is an imperative, because some of these

¹³ This duty is indirect because it derives from how unhappiness can be a great hindrance to doing our duty, and a great temptation to transgressing it.

constraints can remain ultimately subjective. A maxim does not become an imperative just by constraining my choice of actions, so, by parity of reasoning, a second-order principle need not become an imperative simply by constraining my choice of maxims. (Again, recall the discussion of hypothetical imperatives in §2.3) Instead, if a second-order principle must be an imperative, it ought to be either a hypothetical or a categorical imperative. It surely need not be a categorical imperative. But why must I have an end or purpose that *requires* me, as a point of practical *necessity*, to choose my maxims according to its second-order dictates? What, then, prevents me from just adopting a second-order maxim in precisely the way that I can just adopt first-order maxims? Our example of prudential reasoning shows that nothing prevents this: it is perfectly possible, in Kant's view.

Still, Allison suggests one last reason for thinking that imperatives and maxims are distinguished by their order, in Kant's mind. This reason seems to be both the deepest and best grounded in Kant's texts. Here, the idea is that imperatives must be second-order principles because imperatives "specify the norms for maxim selection and action" (Allison 1990, 88). Or, to be a little more precise, we must separate imperatives and maxims in this fashion in order "to preserve the normative character of laws vis-à-vis maxims." (Allison 1990, 131)

My remarks so far already ought to cast serious doubt on the soundness of this line of argument, to say the least. (Kant's examples of laws and imperatives, such as "thou shalt not lie", never give any indication that Allison is correct here, primarily.) But Allison does point to a passage to support his view, so let's explore how his strategy here might go.

Allison's best hope for dealing with the issues canvassed above, it seems to me, is to draw a distinction between Kant's views as of the *Metaphysics of Morals* and those in his earlier texts. The idea here would be that the *Metaphysics of Morals* marks the advent, or at least the first explicit development, of the distinction between will (*Wille*) in the narrow sense and choice (*Willkür*), and that this distinction represents a refinement in Kant's views significant enough mark a break from his earlier views. Allison could then say that the earlier texts quoted in

Chapter 3.3 above, which do not seem to allow us to distinguish maxims and laws by their orders, are being imprecise in their treatment of the distinction between maxims and practical laws. Because this strategy emphasizes the *Wille/Willkür* distinction, it would mesh well with the overall spirit of Allison's approach to Kant's critical ethics. And Allison does rely on the *Wille/Willkür* distinction in his thinking on maxims (Allison 1990, 130-131).

There, Allison cites the best bit of textual evidence for his interpretation. As I just suggested, this text comes from the general introduction to the *Metaphysics of Morals*, where Kant elaborates the *Wille/Willkür* distinction by saying that

Laws proceed from the will, *maxims* from choice. In man the latter is a free choice; the will, which is directed to nothing beyond the law itself, cannot be called either free or unfree, since it is not directed to actions but immediately to giving laws for the maxims of actions (and is, therefore, practical reason itself). (MM 18, 6:226)

At first, saying that the will is directed "immediately to giving laws for the maxims of actions" certainly makes it sound as if these laws should be second-order practical principles directing our choice of maxims. So once Kant introduces the distinction between *Wille* and *Willkür* in this fashion, don't we need to treat imperatives as second-order principles in order to preserve their status as norms for maxims? Don't maxims then have to be first-order principles? So shouldn't we go back and try to understand the apparently conflicting passages and lines of thought in these terms?

This proposal might seem persuasive, were it not based on a plain over-reading of this text. For a law clearly does not have to be a second-order practical principle in order to give "laws for the maxims of actions". I have already illustrated several ways in which a law can direct our choice of maxims without mentioning maxims explicitly. In fact, I have been insisting that *any* practical principle directs our choices of maxims. Consider again the imperative "thou shalt not lie": this imperative plainly directs our choice of maxims, even without explicitly referring to how we must choose our maxims. After all, it plainly commands us to adopt a maxim of not lying, and to reject any maxim that involves lying. For conforming our actions to "thou shalt not

lie” must require us to adopt and reject these maxims, or other maxims that have the same practical consequences, at any rate.

Accordingly, we should conclude that there is no textual basis in the *Metaphysics of Morals* sufficient to reinterpret the passages cited in Chapter 3.3 that conflict with Allison’s interpretation of Kant’s concept of a maxim. Instead, these earlier passages should be used to interpret the *Metaphysics of Morals*, so that the refinements introduced by the *Wille/Willkür* distinction do not revise the underlying picture. Allison thinks the best thing to say on Kant’s behalf is that “it is *Wille* in the narrow sense that provides the norm and *Willkür* that chooses in light of this norm” (Allison 1990, 130). But *Willkür* still does not act *from* the norm, only *in light of* it, he thinks. So Allison concludes that we act from maxims, but only in light of, and never from, imperatives. This picture is right, of course, except for the last part. *Wille* might provide the norm, and *Willkür* might choose in light of that norm, but that does not at all imply that *Willkür* does not also act from the norm. Why should it? Especially when we have seen nothing in Kant’s texts to prevent maxims from being moral laws, nor *vice versa*. But then these two kinds of practical principles must have the same structure, as I have been contending.

In fact, I suspect that Allison’s proposal actually destroys “the normative character of laws vis-à-vis maxims”, rather than preserve it, because Allison appears to break the connection between laws and the autonomy of the will. I can only raise a few questions here, since a full discussion would require much more detail about Kant’s views on autonomy, and it would even require us to go a fair bit beyond what Kant says about autonomy as “being a law to oneself”.

Still, though, it might be rather hard to see how laws of freedom could ever be the laws of an autonomous will, if the best that an autonomous will can do is conform its maxims to those laws. Under Allison’s kind of view here, we get from Kant a sort of parallelism of principles: because laws and maxims are of different orders, the laws that are valid for an autonomous will are on one level, while the principles that this will actually acts upon are on another. And when

all goes well, morally speaking, these two levels conform to each other, but the two levels still never merge or intersect.

In what sense, then, would the will be a legislator with regard to the laws that are valid for it? It seems that the will could legislate its conformity to the law, but not the law itself. Can that be enough to make the law “ours” in the right way for Kant? If not, how can we say that the categorical imperative is the fundamental principle of an autonomous will (G 47, 4:440)? Won’t this will find its laws to be an external constraint on its own principles, since *ex hypothesi* its own principles can only be maxims, and never the laws themselves? How, then, could the validity of these laws come from the autonomy of the will? Mustn’t the form of moral law now turn out to be the object, the end, the good, the purpose that our volition aims at and is supposed to realize, rather than the form of our own lawgiving? Mustn’t these laws really then devolve into the laws of a heteronomous will? Or else won’t it turn out that our wills will still require their own principles and their own law? In that case, however, these laws will turn out to be the true laws of freedom and the true principles of autonomy.

4.4 Formalized Treatments of Maxims

In our post-Frege/Russell era of analytic philosophy, it has become very common to try to explain what maxims are by attempting to render their structure in the apparatus of symbolic logic. Here are some representative approaches, some of which I already mentioned at the end of the last chapter:

a maxim may be characterized as a self-imposed, practical principle or rule of the form: When in S-type situations, perform A-type actions. (Allison 1990, 89-90)¹⁴

¹⁴ O’Neill (2013, 98-99) takes a similar approach, though substituting an “agent description” for a situation type. Allison is actually a little unclear about what should be included in the form of a maxim. He correctly points out later that “every maxim reflects an underlying interest of the agent, which provides the reason for adopting the maxim. Consequently, a reference to this interest is implicit in every maxim, constituting, as it were, part of its “deep structure”; but it need not be and, in fact, usually is not made explicit.” (Allison 1990, 90) Allison does not explain why this “implicit reference” or “deep structure” should not be considered part of the *actual* form or structure of the maxim, however, even if it is not usually made explicit. He does not appear to want to include interests as part of the “situation”, though, since there he appears to be only concerned to say what the explicit form of a maxim has to be. Why there should be any specific emphasis on the explicit structure remains unclear to me. Given Allison’s otherwise

A maxim of action will therefore usually have the form “I will do Action-A in order to achieve Purpose-P.” (Korsgaard 1996, 57-58)

I will do *A* if *p* (Aune 1979, 15)¹⁵

I am to do *X* in circumstances *C* in order to bring about *Y* unless *Z*. (Here *X* is an action and *Y* is an end, a state of affairs.) (Rawls 2000, 168)¹⁶

There are many other examples, but these convey the general tenor of the formalized approaches to understanding Kant’s views on maxims. Generally speaking, the elements that are usually thought to constitute a maxim are some combination of a situation or a set of circumstances, an action, and an end or purpose for the action. Sometimes there are indications that a maxim might include something like a ground or a reason for an action, but this suggestion is often not incorporated into the statement of the form of a maxim, or else the ground or reason is not fully distinguished from either the circumstances or the end/purpose.¹⁷ Instead, the proposed formalization usually only involves two of these elements: either a situation and an action or an end and an action, although Rawls does include all three.

By the lights of my account, none of these standard lines of interpretation render Kant’s thinking quite correctly, since I have been arguing that a maxim requires three elements, related in a very specific way: a ground, an action, and an end, where the latter are necessary consequences of the former. From my perspective, the common formalized approaches tend to leave out one element that I would require, construe the ground of the maxim too broadly,

thorough and very salutary emphasis on the Incorporation Thesis, it is rather odd that interests, incentives, or grounds do not show up in his official statement of the form of a maxim.

¹⁵ Aune is a little unclear about what “if *p*” is to involve. He says that these “subjective conditions” are “reasons or motives” that “are acceptable to a particular agent because of contingent facts about him” (Aune 1979, 15). But he seems to think that these reasons or motives must actually turn out to be ends or purposes, or else just conformity to universal law just as such (1979, 122). So his view seems more similar to Korsgaard’s than to mine, even though he, too, explicitly connects maxims to the hypothetical form of judgment (1979, 122).

¹⁶ As I mentioned at the end of the last chapter, the “unless *Z*” clause seems otiose, given that maxims are already supposed to stipulate the circumstances of action. And Rawls drops this clause from his subsequent analysis, too.

¹⁷ See notes 14 and 15, just above.

underdetermine the relation between the elements, or some combination thereof. For example, Allison's reference to "S-type situations" could be construed as capturing the ground of the action, but only if we restrict it to the specification of the situation that is supposed to make a certain action practically necessary. And similarly for O'Neill's "agent description" and Rawls' circumstances. Likewise, Aune and Korsgaard fail to include anything specifying why an agent should adopt a certain end or purpose.

Relatedly, one issue with these lines of interpretation is that these accounts of the form of maxims often do not fully capture their author's own views about what maxims are supposed to be, for Kant. For everyone, or virtually everyone, agrees that maxims are to play a very specific role in practical reasoning: my maxim is to capture or represent the principle that is supposed to justify my action, even if only to myself. (Which is part of the reason why maxims are not a kind of intention.) In other words, my maxim is the principle according to which my action appears rational to me. For example, if I get caught telling lies in order to get a loan, and someone wants me to justify myself, the proper answer, true to the maxim of the lying promise, would be: "Well, I believed I need money, and when you believe that you need money, you tell lies in order to get loans."

But these kinds of formalized accounts of the structure of maxims tend to be good at conveying how actions can be derived from principles, while failing to convey how these principles are supposed to rationalize the actions derived from them. Even if we ignore these authors' broader readings of Kant, their descriptions of the form of maxims would still give us a clear sense of how we are to derive our actions algorithmically from our maxims. After all, they tell us quite clearly how to derive an action from a situation, a set of circumstances, or an end.

However, despite these authors' best intentions, these derivations would have the feeling of a purely formal operation, which would do nothing to explain why it might even just look rational to act according to such a derivation. A certain principle says what the end of my action is to be, or that certain types of agents are to perform certain types of actions, or that I am to

perform a certain type of action in a certain type of situation: but none of that is enough to know what it is about that end, agent-type, or situation-type that is supposed to make it rational to act as the principle specifies. In short, these principles do not tell us what it is about their components that is supposed to provide a reason or ground for the action. For that we need to specify a ground: something that is to make it rationally necessary to act according to the principle.

In many cases, that ground will also be the end. Perhaps even in most cases. Kant thinks that our agency works basically like this: we represent to ourselves the end that we are to accomplish, and then we act to realize that end. For the will, Kant thinks, is a component of our faculty of desire, which is “a being’s *faculty to be by means of its representations the cause of the reality of the objects of these representations.*” (CP 8n, 5:9n; MM 11, 6:211; MM 13, 6:213) So our ends will often also serve as the grounds of our actions. Kant even has a technical name for the principles of this kind of action: “All practical principles that presuppose an *object* (matter) of the faculty of desire as the determining ground of the will” are *material* practical principles, he says at (CP 19, 5:21-22). But even when the end of the action also serves as its ground, these roles remain functionally distinct, and so the form of maxims needs to keep them separated, in order to capture how a maxim is to rationalize the actions pursuant to it.

So there is a gap between showing how an action can be derived from a principle and explaining the (purported) rationality of acting under that principle. Kant does say that the will is practical reason because “*reason* is required to for the derivation of actions from laws” (G 24, 4:412). But whether these derivations make an action rational, either actually or in appearance, depends on the underlying laws. Which is why it is essential for these laws to include the grounds from which our actions are to follow.

To their credit, most commentators do fill in this gap between derivations that rationalize actions and those that do not, but only elsewhere in their accounts of Kant, and not

in their official formalizations of the structure of maxims. Which only underscores how these formalized approaches are themselves lacking as an account of Kant's concept of a maxim.

This is not to say that no formalized approach can work. We could, after all, use my analysis to account for the form of practical principles, in this fashion:

Grounds of type *G*, when they obtain, make it (subjectively, objectively) necessary to (1) adopt an end *e* of type *E*, and (2) to perform an action of type *A* in order to realize *e*.

Here, "subjective" and "objective" serve two purposes: to distinguish maxims from the principles representing practical laws, and to specify the domain of rational agents for whom *G* is to constitute a valid type of grounds for their actions.

We could, of course, still use such principles in a purely formal fashion, without grasping or caring about how they purport to rationalize our actions. But at least their form now requires a specification of what is supposed to make the consequent actions rational, regardless of whether one acts from that ground or only in accordance with it.

4.5 Conclusion

One of the great oddities of the commentaries on Kant's critical ethics, I find, is that they tend to elide or minimize his persistent and fundamental emphasis on the grounds from which we act and according to which we ought to act. Not in the aggregate, mind: this emphasis is always part of the broader picture, of course. But the account of practical principles is precisely the place where this emphasis is most necessary in order to fully capture the structure and spirit of Kant's approach to ethical theory in his critical philosophy. So the tendency to de-center the role of grounds exactly there remains rather odd.

I think this strangeness can be seen most clearly if we return to where we began the discussion in this dissertation: to Kant's concept of freedom. If freedom is to involve the capacity of the will to determine itself with absolute spontaneity, but still in accordance with some kind of grounds, then the principles adopted by such a will must inevitably include reference to the

grounds it chooses for its absolute self-determination. As Kant put it in the *Religion*, quoted in full previously, but here slightly edited:

Since the adoption [of a maxim] is free, its ground (e.g. why I have adopted an evil maxim and not a good one instead) must not be sought in any incentive of nature, but always again in a maxim; and, since any such maxim must have its ground as well, yet apart from a maxim no *determining ground* of the free power of choice ought to, or can, be adduced, (R 47n, 6:22n)

it will follow that the maxim of a will that is free, but not for that reason lawless or groundless, must contain the ground of the will's absolute self-determination.

With all of this in mind, we are now in a position to approach Kant's derivation of the Formula of Universal Law from the concept of a categorical imperative.

5. Deriving the Formula of Universal Law¹

Now that we have my account of Kant's practical principles fully on the table, we can finally turn to his moral theory itself. Kant's view, of course, is that the Formula of Universal Law gives voice the fundamental requirement of morality. For this Formula is to be, as Kant characteristically puts it, the fundamental or supreme principle of morality (G 5, 4:392).² So in this chapter, I want to explore and offer a defense of one of the arguments that is supposed to show that the Formula of Universal Law is the fundamental moral principle.

First, it will help to briefly recapitulate what the Formula of Universal Law requires: the Formula of Universal Law holds that you must "*act only in accordance with that maxim through which you can at the same time will that it become a universal law.*" (G 31, 4:421) By this Formula, Kant means to say that morally permissible action must be universal, in the sense that moral behavior must follow from principles that could both be universal laws, and be willed as universal laws (G 33, 4:423-424). To be a little more specific, Kant thinks that morally permissible actions must follow from principles that everyone could follow, without the possibility of exception, and which we could also will for everyone to follow, again, without the possibility of exception.

Moreover, the Formula of Universal Law directs our choice of maxims, as well as our choice of actions. Even if it is possible to act in a way that satisfies a principle that you can will as a universal law, but by following a principle that cannot be willed as a universal law, that would not be enough to put yourself in the good graces of the Formula of Universal Law. Because the Formula of Universal Law requires us to always act on a maxim that we can will as a universal

¹ In addition to my previous acknowledgements, I would also like to thank the UCLA Ethics Writing Seminar for many valuable comments on various drafts of this chapter.

² Of course, Kant thinks that there are other Formulae that are to be equivalent to the Formula of Universal Law, such as the Formula of Humanity and the Formula of Autonomy (G 43, 4:436). He does not explain the sense in which these Formulae are to be equivalent, but their equivalence still allows us to call the Formula of Universal Law "the fundamental" categorical imperative. And, in any case, Kant still prioritizes the Formula of Universal Law as the way to "proceed in moral *appraisal* by the strict method" (G 44, 4:436). So it remains "first among equals", so to speak.

law, “*act in accordance with*” must tell us which principles we are to adopt and follow, and not just which principles we are to satisfy in our actions.³

So in this chapter, I want to explain why Kant thinks that these requirements should be our fundamental moral requirements, and, in particular, I want to offer a new solution to one of the longest-standing and most fundamental problems for Kant’s position here.

To introduce this problem: the idea that morality must be universal, in the sense that everyone must be able to follow moral principles, is surely controversial. I will discuss a few problems that have been raised for this idea in the next chapter. However, this idea is not that strange. It is rather more striking, though, for Kant to argue that our fundamental moral obligation is not just to act on principles that everyone could follow, but to act on principles that we could *will* for everyone to follow.

Historically, this difference between principles that everyone could follow and principles that we can will for everyone to follow has been the locus of one of the oldest and most persistent questions for Kant’s moral theory: even if we grant that morality requires us to act on principles that everyone can follow, why isn’t that enough, morally speaking? Why do we also have to be able to will our principles as universal laws?

Now, the difference between a principle that can be a universal law, in the sense of “everyone could follow this principle”, and a principle that can be willed as a universal law, in this sense, might seem negligible at best. So at first it might be hard to see what the problem here is supposed to be. But Kant admits that many maxims could be universal laws, even if they cannot be willed as such at the same time. What’s more, the basic adequacy of the Formula of

³ Kerstein (2002, 17-18) claims that the Formula of Universal Law says nothing about actions performed on no maxim, though, to be sure, he also argues that Kant denies that there are any such cases. I disagree, however. The Formula of Universal Law tells us to act only on maxims that we can will as universal laws. So to get this Formula to leave out actions without maxims, Kerstein has to read it as a principle that is solely about which maxims are permissible. But I think Kant is clear that the Formula of Universal Law is supposed to determine which *actions* are permissible, and not just which maxims are permissible. After all, the Formula starts “*act only...*”. And how could the Formula of Universal Law be the fundamental moral principle otherwise? So this Formula really tells us to always act on maxims that we can will as universal laws. And then if we act on no maxim at all, we violate the Formula of Universal Law, at least if we are capable of acting on a maxim that we can will as a universal law instead.

Universal Law as a principle of morality hinges on whether it can impose the second, stronger requirement.

For example, if I make it my maxim to help others only when I expect to receive something suitable in return, then Kant will think that my maxim is perfectly capable of being a universal law. Everyone could always act that way, after all (G 33, 4:423). So if morality can only require us to act on maxims that can be universal laws, then it would always permit us to put our own interests first, at least in this sense.

From Kant's point of view, though, to admit that would be to admit that moral obligation doesn't exist at all. He thinks that genuinely moral obligations, especially those to others, must be able to take precedence over our individual interests, at least some of the time. That is just part of the absolute, unconditional, and categorical character that Kant thinks is distinctive of genuinely moral obligations. And without that character, morality would be nothing more than a "high-flown fantasy" (G 8, 4:394) or an "empty delusion" (G 15, 4:402), as he puts it.

Thus, Kant's whole moral theory really depends on the idea that we have to be able to will our maxims as universal laws, precisely as the Formula of Universal Law requires. Kant provides one argument for the Formula of Universal Law as the fundamental principle of morality in the Second Section of the *Groundwork of the Metaphysics of Morals*, and another at the beginning of the *Critique of Practical Reason*.⁴ In this chapter, though, I want to focus on the argument in the *Groundwork*, and I will contend that the key to Kant's reasoning there is ultimately quite simple.

Simple, but not by any means obvious: I will argue that Kant's reasoning in the *Groundwork* hinges on the distinction that he draws between imperatives and the laws that imperatives express. For example: "thou shalt not lie" is an imperative, which expresses the moral law forbidding lying. Thus, this distinction between imperatives and laws is a version of

⁴ The argument in the Second Critique hinges on the difference between practical principles that determine the will through their form and those that determine the will through their matter or content.

the distinction that I have been drawing between practical principles and the laws that they represent, and my argument here will be continuous with the account of practical principles developed so far. I think that these distinctions are often overlooked in the literature, so the work of this chapter will be to articulate the difference between imperatives and their underlying laws, and then to use that distinction to explain why the *Groundwork* thinks that the fundamental moral imperative must require us to be able to will our maxims as universal laws.

Why should Kant's concept of an imperative help explain his thinking here? In short, because Kant himself emphasizes it. As we will soon see, Kant's argument in the Second Section of the *Groundwork* uses the concept of a categorical imperative to try to prove that the Formula of Universal Law is the only principle that could be the fundamental categorical imperative (G 30-31, 4:420). Then, since Kant thinks that genuinely moral obligations have to be expressed in categorical imperatives, it would therefore follow that the Formula of Universal Law will have to be the fundamental principle of morality.

This argument is often called the *Groundwork's* "derivation" of the Formula of Universal Law, because it seeks to derive what morality is to be from the very concept of a moral imperative: the content of the fundamental moral imperative, and the rest of the structure of morality with it, derived from the general concept of a distinctively moral imperative. So if Kant's derivation is sound, "we shall at least be able to show what we think by it and what the concept wants to say." (G 31, 4:421)

But this derivation does not show, nor does it purport to show, nor can it show, that the anyone is actually subject to the requirements expressed in Formula of Universal Law. If Kant's derivation is sound, it shows that the Formula of Universal Law is the fundamental principle of morality. But that does not show that morality is real or that moral obligations are real, any more than identifying the formula for Newton's Law of Gravity would show that gravity is real. Further argument would be required to show that these formulae are *valid* in any way: $F_g =$

$(m_1 m_2)/r^2$ needs to be backed up with astronomical data, and Kant has to show that there is such a thing as moral or categorical obligation.

Kant himself insists on this point, because he insists that no amount of conceptual analysis can show that a concept applies to anything. So when his derivation for the Formula of Universal Law is supposed to work by analyzing the concept of a categorical or moral imperative, he will insist that this argument cannot show that there are any such things as categorical imperatives or categorical obligations. The “deduction” is the argument that seeks to show that morality is real, and that we really are subject to moral obligations and their categorical imperatives.⁵ And we will explore the *Groundwork*’s deduction for the Formula of Universal Law in a later chapter.

Without further ado, then, here is the *Groundwork*’s derivation of the Formula of Universal Law, or its concluding steps, anyway, in its entirety:⁶

⁵ These arguments have come to be called “deductions”, following Kant’s example in the First Critique’s Transcendental Deduction of the Categories (CR 219-220, A84-85/B116-117). In the *Groundwork*, the deduction is supposed to be the work of the Third Section and its “acting under the idea of freedom” argument, which we will examine in a later chapter. The *Critique of Practical Reason* gives a slightly different, but arguably complementary argument, in its famous “fact of reason” passage (CP 28-29, 5:31). Rawls points out, though, in his lectures on Kant (Rawls 2000, 261-264), that the *Critique of Practical Reason* explicitly denies that there can be any deduction for the fundamental categorical imperative, at least not in the way of the categories (CP 41-42, 5:46-47). There, by a “deduction” of “the supreme principle of practical reason”, Kant indicates that he means “the justification of its objective and universal validity and the discernment of the possibility of such a synthetic proposition a priori” (CP 41, 5:46). However, in denying that such a deduction is possible for the fundamental categorical imperative, Kant focuses on the latter part of this conjunction, and never denies that the former is possible. He may also mean to deny that a proof can be given for the objective validity of the fundamental categorical imperative, where a proof specifically means a deductive argument from principles, of the sort that the Transcendental Deduction provides for the categories. Or he may simply think that ‘deduction’ should not be applied to practical arguments. Be all of that as it may, however, we can still speak of a “deduction” for the fundamental categorical imperative, as long as we only mean the practical considerations that are supposed to disclose to us the objective validity of that imperative, such as the considerations Kant provides in the Third Section of the *Groundwork* and in the Second Critique. Rawls prefers to speak of Kant’s “authentication” of the fundamental categorical imperative, and obviously for good reason. I will stick with ‘deduction’, though, in order to preserve Kant’s own terminology and the parallelism with the *Critique of Pure Reason*, as far as that is possible.

⁶ It is, of course, a matter of considerable controversy where the derivation of the Formula of Universal Law actually begins in the *Groundwork*. For the sake of simplicity, though, I will consider this passage to be the derivation itself, and the prior remarks explaining the underlying concepts, for example, to be discussion preliminary to the actual derivation.

When I think of a *hypothetical* imperative in general I do not know beforehand what it will contain; I do not know this until I am given the condition. But when I think of a *categorical* imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary.

There is, therefore, only a single categorical imperative and it is this: *act only in accordance with that maxim through which you can at the same time will that it become a universal law.* (G 31, 4:420-421)

This argument is as striking as its conclusion, but it is also quite easy to see why Kant's reasoning here has been plagued by questions, and why the concept of an imperative should help solve those questions. The argument's strategy is bold, but its execution is quite brief, evidently because Kant thinks that its conclusion follows "at once". Yet this conclusion does not really seem to follow at all. The paragraph break marks the key step: the concept of a categorical imperative leaves nothing "left with which the maxim of action is to conform but the universality of a law as such". So Kant argues that "There is, therefore, only a single categorical imperative", which says that we must "*act only in accordance with that maxim through which you can at the same time will that it become a universal law.*"

We do not even need to know what Kant means by "conformity with the universality of law as such" to see the question about this crucial step in the derivation: if our maxims only have to conform to the universality of a law as such, why must we also be able to *will* them as universal laws? More pointedly, though, couldn't we really be subject to a universal law, even though we could not will it as such? And if we were subject to such a law, then wouldn't we be categorically obligated to follow it, and thereby obligated, categorically, to make that law the maxim for our actions as well? In which case, wouldn't we be under a categorical imperative to act according to a maxim that we cannot will as a universal law? So how can the concept of a categorical imperative require us to be able to will our maxims as universal laws?

As I mentioned before, these questions have been pressed in many different ways, and the earliest versions date back to the first critical responses to Kant's work. Allen Wood (1990, 155-156) points out that there are versions of these problems already present in Fichte's *Science of Ethics* (1897), and even in an earlier criticism of Kant by Gottlob August Tittel in 1786. Hegel is a major locus of this kind of objection as well (e.g. PR §135; PS ¶¶596-631).⁷ And more recent commentators like Bruce Aune (1979, Chs. 1.5, 3.3) and even Wood himself (1990, Ch. 9) have also criticized Kant's derivation on these scores. I will return to answer some of these specific problems in the next chapter. This chapter, though, will focus only on the general problem: why does Kant think that morality requires us to be able to will our maxims as universal laws?

For even this general problem presents serious difficulties. Primarily, the more complicated and subtle that we make Kant's argument look, the harder it becomes to see how he could have thought that its conclusion follows "at once". The literature on Kant tries to defend his argument in many different and illuminating ways. Allison (1996, Ch. 10), for example, argues that the concept of transcendental freedom is ultimately the key to answering these kinds of objections to the derivation.⁸ My goal here, though, is to complement the defenses in the literature by drawing out the power in Kant's argument, while preserving the immediacy in his presentation. So I will also pass over the other defenses without comment, along with the specific objections to Kant's derivation.

Instead, because Kant's argument starts with the concept of a categorical imperative, my strategy is to start with his overarching concept of an imperative in general. For Kant, imperatives express the demands that objective laws impose on imperfectly rational agents, just as "thou shalt not lie" is supposed to express an objective law forbidding beings like us from

⁷ See the list of abbreviations for the citations to Hegel.

⁸ To be sure, Allison agrees that the *Groundwork's* derivation fails (Allison 1996, 144). He holds that transcendental freedom saves the derivation in the *Critique of Practical Reason*.

lying. So laws and imperatives are different, just because imperatives express a law's demands. And this difference matters in two ways.

First, since imperatives have to express the demands that a law imposes on certain rational agents, if a law cannot impose demands on those agents, then it cannot be expressed in an imperative for them, either. So the need to impose demands on rational agents might provide additional constraints that laws must meet in order for them to be able to be expressed in any imperative at all.

Second, because imperatives express the demands that laws impose on certain rational agents, imperatives are also essentially addressed to those rational agents. The law's demands are themselves addressed or directed to certain rational agents, so the imperatives formulating those demands are, too. As such, the requirements expressed in an imperative can be considered from two perspectives: from the (objective) point of view of the law itself and of what it requires in itself, or from the (subjective) point of view of the rational agents to whom these requirements are addressed. No matter which point of view we take, though, I will argue that there really will be further constraints on which laws can be expressed in imperatives, so that even some possible universal laws cannot be expressed in imperatives.

Briefly: imperatives are addressed to rational wills, albeit to imperfect ones. So when we take the point of view of the rational agent who is subject to an imperative, we will find that imperatives must conform to the more fundamental principles of their reason, such as the principle of non-contradiction. So a universal law that would force us imperfectly rational beings to contradict ourselves cannot be an imperative for us, because that imperative would violate the standards internal to our reason. But when we take the point of view of the law itself and of the requirements that are to be expressed in an imperative, we will find that some principles can be universal laws, while still imposing incoherent requirements. Such a principle can be obeyed universally—every agent really could always fulfill it in their actions—but only if they make

incompatible judgments, for example. Incoherent requirements like this cannot be expressed in a coherent imperative for a rational will, however.

So either way, imperatives just as such will require consistency within our volition. The general point here is that imperatives are to express norms for imperfectly rational agents, because of how imperatives are to express the demands of the laws to which these agents are subject. As such, it is not enough if everyone can follow an imperative, because the norm that the imperative expresses must also be rationally coherent.⁹

Distinctively categorical imperatives will also be universal in their distinctive way, and so Kant argues that categorical imperatives will require us to act on maxims that can be universal laws. But as imperatives, categorical imperatives must also require internally consistent volition. Accordingly, categorical imperatives must require us to act on maxims that we can all follow without having to contradict ourselves. For Kant, though, to be able to will a maxim as a universal law is just to be able to follow it as a universal law without being forced to contradict yourself. So categorical imperatives will require us to be able to will our maxims as universal laws after all, just as Kant contends.

To spell out this argument in more detail, I will begin by examining Kant's concept of an imperative more closely. Then I will walk through his derivation of the Formula of Universal Law, from both the subjective and the objective points of view just described.

5.1 Imperatives in General

In the *Groundwork's* Second Section, Kant introduces imperatives while laying the foundation for the derivation of the Formula of Universal Law itself. To lay that foundation, Kant explains that his strategy is to "follow and present distinctly the practical faculty of reason,

⁹ Thanks to Tyler Burge for suggesting more emphasis on imperatives as norms of reason.

from its general rules of determination to the point where the concept of duty arises from it.” (G 24, 4:412)¹⁰ So let’s follow his text closely and in sequence from there.

Immediately following, we get Kant’s famous *dicta* that “Everything in nature works in accordance with laws”, and then his explanation that “Only a rational being has the capacity to act *in accordance with the representation* of laws, that is, in accordance with principles, or has a *will*. Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason.” (G 24, 4:412) So here we have the aforementioned presentation of “the practical faculty of reason” and its “general rules of determination”: practical reason is the “*will*”, which is just “the capacity to act *in accordance with the representation* of laws”, or, in other words, “in accordance with principles”. These principles themselves then turn out to be the will’s “general rules of determination”.

A few comments before moving on: first, Kant says here that principles are representations of laws, just as my account of practical principles would have it. Some commentators, though, think that Kant really means to say that principles are representations of laws *of nature* (Cramer 1972), and Bittner thinks that Kant really means to say that the will is a faculty to act according to maxims (Bittner 1974, 492). The details of these arguments needn’t detain us, since I think that the text immediately refutes them.¹¹

The will can, of course, act according to maxims and according to representations of laws of nature. Those are both things that the will can do, according to Kant, so the will is certainly a faculty for acting in accordance with both of those things, for him. But clearly not just those things, and not primarily or properly those things, either.

For why would the will be the capacity to act in accordance with representations of laws of nature, when the entire point of Kant’s critical ethics is to show that we can act according to pure practical principles and pure practical laws (CP 3, 5:3)? When we discussed the Second

¹⁰ Arguably, this remark marks the beginning of the derivation argument. See note 6 above.

¹¹ But see note 14 in Chapter 3 for some discussion of Bittner’s view on this point.

Critique's Typic of Pure Practical Judgment in Chapter 3.2, we did see how Kant thinks that judgment according to pure practical principles needs to be mediated by the concept of a law of nature. Even so, we also saw there how our actions pursuant to such judgments will still be based, ultimately, in a pure practical principle, and not just in the mediating representation of a law of nature. And if Kant meant for the will to be the faculty to act "*nach Maximen*", as Bittner holds, then when does he call it the faculty to act "*nach Prinzipien*"? Especially when Kant has already introduced *Maximen* several times by this point in the *Groundwork*, both formally (G 14n, 4:401n) and informally (G 13, 4:399; 15, 4:402)? Instead, the will is ultimately supposed to be just as much a faculty for acting according to objective principles in general and pure principles in particular, and not just according to maxims or according to representations of laws of nature. Thus, Kant is perfectly correct to describe the will as a faculty to act in accordance with principles in general. For this allows him to leave open the possibility of action according to pure and objective principles, but without presupposing that possibility, either.

Next, moving on from Bittner and Cramer: Kant says that principles are representations, so it follows that principles can be either accurate or inaccurate. Their object, which they represent as a law, might actually be a law, or it might not be. And genuine laws must be "valid" for at least some agents, meaning that the agents in question must obey these laws, or else they make a rational mistake. So these laws constitute norms of reason for those agents, at least when those agents are capable of violating those laws. (I will say more on that point in a moment.)

In turn, laws can be valid either subjectively or objectively: subjectively, if the law could be valid for only a single subject, and if failing to obey the law would be a mistake only subjectively, or objectively valid, if the law is valid for all rational beings of a certain kind, and if the error in deviating from the law would be objective. Objective validity also comes from the nature of the law, while subjective validity comes from something that could be idiosyncratic about the agent subject to it. So when a principle represents a valid law, then that principle can be called a "law", too, and that principle is also valid, but in its own distinct senses. A "law", by

virtue of representing a law rather than by being a law itself, and valid by virtue of representing a valid law, by being an accurate or valid representation of an actually valid law.

Last, it is also worth noting right away that hypothetical and categorical laws, obligations, and imperatives are all objective. As such, this account of objective validity applies equally to all of them.

All of these points become clear as Kant moves from “the practical faculty of reason” and its “general rules of determination” “to the point where the concept of duty arises from it”. So let’s explore that step next: how does Kant go from the will as the faculty to act according to principles “to the point where the concept of duty arises from it”?

This step in the argument drives the ultimate nail into the coffin of readings like Cramer’s and Bittner’s: to get to the concept of duty from the idea of the will as the capacity to act according to principles, Kant distinguishes rational beings that necessarily will act according to the rationally correct principles from those that need not. The rationally correct principles will then constitute binding norms for the latter kind of rational being, and Kant analyzes these norms in order to arrive at the concept of duty.

Thus, Kant needs the will to be a faculty for acting in accordance with objective principles, not just maxims, and for acting according to objective principles of reason, not just representations of laws of nature. The will is the faculty for acting in accordance with the representation of laws. Thus, if Kant wants to go on to draw a distinction between beings for whom “reason infallibly determines the will” and beings for whom “the will is not *in itself* completely in conformity with reason” (G 24, 4:412-413), then the principles and laws in question must be the objectively correct ones, rationally speaking.¹²

¹² We don’t need to read “infallibly determining the will according to objective laws” as excluding action from morally permissible maxims. We can act “according to” a principle in two main senses: a negative sense, where we merely do not violate the principle, and a sense where the principle positively determines our actions. This is related to the distinction I have drawn between satisfying and following a principle.

Hence, the will has to be a capacity to act in accordance with objective principles of reason, if Kant wants to use a concept of the will in order to distinguish beings that will necessarily act as reason determines from those that need not. (Which is just the distinction between perfectly and imperfectly rational beings.) And the concept of a rational being that can act as reason determines, but need not, then yields the concept of the “*necessitation*” involved in “the determination of such a will in conformity with objective laws” (G 24, 4:413).

Now, Section I of the *Groundwork* argued that “*duty is the necessity of action from respect for law.*” (G 13, 4:400) So this concept of “*necessitation*” represents the terminus of Kant’s presentation of “the practical faculty of reason, from its general rules of determination to the point where the concept of duty arises from it”. For Kant proceeds to introduce philosophical concepts for the different kinds of this necessitation, so that he can characterize the necessitation in moral laws and moral duties, and then use that characteristically moral necessitation in order to derive the actual content of the fundamental law of morality.

Here is where imperatives enter the picture. As Kant puts it, “The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an **imperative.**” (G 24, 4:413)¹³ Despite Kant’s use of ‘principle’ here, it is fair to read him as meaning objective laws also. For he then continues:

All imperatives are expressed by an *ought* and indicate by this the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation). They say that to do or to omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be good to do that thing. (G 24-25, 4:413)

¹³ Kant’s German disambiguates the ‘it’ in “insofar as it is necessitating for a will”: ‘it’ refers back to the principle, not its representation, because Kant’s neuter ‘*es*’ has to refer back the neuter ‘*Prinzip*’, not the feminine ‘*Vorstellung*’. For Kant’s other explications of imperatives, compare (CP 18, 5:20-21) and (MM 15, 6:222). The former pointedly rejects the idea that hypothetical imperatives can be or express practical laws, a point that I will incorporate by emphasizing the differences between the ways in which the two kinds of imperatives are objective and universal. The latter text treats imperatives as rules rather than as representations of rules. But it still distinguishes the two, and it still attaches necessitation to the rule itself (as well as to the representation of the rule). So the differences across these texts are of little importance. If a rule is an imperative, then you can call the representation of the rule an “imperative”, too, in the same way that you can call an objectively valid principle a “law”. In the same way, both the rule/law and its imperative will necessitate the will.

Now, Kant does go on to distinguish hypothetical and categorical imperatives. However, it is worth noticing right away that *all* imperatives represent objective laws; Kant is clear about this in both of these passages. As such, all imperatives are also universal, only in very different ways. Moreover, Kant is quite explicit about how imperatives are to express norms for imperfectly rational agent: all imperatives are expressed by “an *ought*”, while they also express how doing or omitting something “would be good”, but for a will that might not do that good.

I will say more about these points shortly, because Kant’s derivation of the Formula of Universal Law hinges on them. First, though, I want to emphasize how the concept of an imperative begins life, so to speak, as the concept of a formula addressed to an imperfectly rational being. In fact, imperatives are *essentially* directed to the wills of imperfectly rational beings. Kant’s next paragraph is clear that

A perfectly good will would, therefore, equally stand under objective laws (of the good), but it could not on this account be represented as *necessitated* to actions in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good. Hence no imperatives hold for the *divine* will and in general for a *holy* will: the “ought” is out of place here, because volition is of itself necessarily in accord with the law. (G 25, 4:414)¹⁴

The fact that imperatives in general hold only for imperfectly rational wills is fundamentally important.¹⁵ As I have been saying, this is why imperatives both constitute and express norms of reason. Moreover, though, because imperatives are essentially addressed to our imperfect rationality, we should already see conceptual space opening up for laws that everyone could obey, but which cannot be the basis for an imperative.

For, because imperfectly rational beings are always capable of rational mistakes, our rational defects also make us capable of following laws that violate rational norms. (Kant might prefer to view this as an incapacity, though, not a capability.) But if imperatives are essentially

¹⁴ (CP 70, 5:82) and (MM 15, 6:222) make the same point.

¹⁵ Thanks to Barbara Herman for pressing me to be clearer about these points at this stage of the discussion.

addressed to our rational nature, then we should already have serious doubts about whether a law that violates rational norms could really be expressed in an imperative for us, even if everyone could actually follow that law. Hence, we should already be prepared to find ourselves capable of following a universal law that cannot be an imperative for us.

Moreover, if imperatives are addressed to our rationality, then we should already expect imperatives, and the requirements that they express, to be subordinate to, or at least compatible with, the other principles of our rational nature. Primarily, we should already expect imperatives to be subject to coherence and consistency requirements. For, in Kant's view, the fundamental principle of rationality in general, at least in its negative dimension, is just the principle of non-contradiction (CR 279, A150/B189-190).¹⁶ Thus, an imperative addressed to reason ought to have to be compatible with the demand of non-contradiction.

In short, then, even at this early stage of the discussion, before Kant has even distinguished hypothetical and categorical imperatives, we should already expect that imperatives must be rationally coherent, so that following them cannot introduce any irresolvable conflicts into our wills. However, we will soon see that, for Kant, "able to will a maxim as a universal law" only means "able to follow such a universal law without having to contradict yourself". So if we cannot follow a universal law without contradicting ourselves, we cannot will it as a universal law. But then it cannot be an imperative for us, either.

5.2 Hypothetical and Categorical Imperatives

Again, I will return to these points in more detail, since they are crucial to the entire argument in this chapter. For now, though, let's resume following the text of the *Groundwork*, but move ahead a little to Kant's distinction between categorical and hypothetical imperatives. As Kant puts it,

¹⁶ What the positive dimension of the fundamental principle of reason might be is one of the greatest mysteries of Kant's critical philosophy. O'Neill has proposed that it is actually the Formula of Universal Law itself (O'Neill 1989, Ch. 1). I tend to think that this is close to right, at the very least.

all imperatives command either *hypothetically* or *categorically*. The former represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will). The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end. (G 25, 4:414)

Before we get to the difference between these two kinds of imperatives, let's return to what they have in common. Above, I said that all imperatives are both objective and universal. Here, though, Kant says that whether or not an agent is subject to the "practical necessity" in a hypothetical imperative—whether or not a hypothetical imperative is valid for an agent, as I said before—depends on whether or not they actually accept a certain end. So hypothetical imperatives might seem to be subjective.

But Kant also says that hypothetical imperatives represent the necessity in taking a certain means to a certain end, which is an objective relation. The character of the means/ends instrumental necessity here is crucial: the means really must be *necessary*, not just suitable or expedient, for example. As Kant puts it a little later, in explaining how hypothetical imperatives are possible: the means must be "indispensably necessary" to the end in question (G 28, 4:417).

Thus, the necessity in taking the indispensable means to my end is certainly not subjective in the way that a psychological compulsion is a subjective necessitation. That would be a necessitation that has to do only with the idiosyncrasy of an individual subject, so it might be indispensable for them, but not in general. Nor is the necessity in a hypothetical imperative subjective in the way that the choice among equally good means to my end is subjective. When several means are suitable to one of my ends, but none of them are indispensably necessary to that end, then any necessity in taking any one of these means can only be subjective. There could be an objective *general* necessity to take some one of these means, based just on the idea that "Whoever wills the ends also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power." (G 28, 4:417). But any *specific* necessity would be subjective because, *ex hypothesi*, it would have to depend on desires or inclinations that might be wholly idiosyncratic to the agent.

Instead, when the means really is indispensably necessary to the end, then the necessity between them remains objective. To illustrate why “Whoever wills the ends also wills” the means indispensably necessary to that end, Kant mentions that “in order to divide a line into two equal parts on a sure principle I must make two intersecting arcs from its ends” (G 28, 4:417). This is entirely a fact about lines in Euclidean geometry, and so it is wholly objective, wholly due to the nature of the objects represented in this hypothetical imperative. Whether or not I have to follow this imperative does depend on whether or not I set this end for myself. But the necessity in taking this means to this end remains entirely objective: if I try to bisect a line by another method, I make an error objectively, in terms of Euclidean geometry, and not just subjectively.

So, to tie these points together with another example, Kant says that hypothetical imperatives about happiness do “involve necessity, which, however, can hold only under a subjective and contingent condition, whether this or that man counts this or that in his happiness” (G 27, 4:416).¹⁷ So hypothetical imperatives are objective, but they hold for an agent only under a subjective condition. This makes them conditionally objective, not subjective, at least whenever they are actually valid for an agent. Moreover, as objective, hypothetical imperatives are also universally valid: valid for all rational beings, but only granted that they actually have a certain rationally contingent end.

Now, in contrast to hypothetical imperatives, Kant says that the “categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end”. Categorical imperatives can still contain or command ends, however, and the nature of those ends can still be relevant to whether or not an imperative is valid for us, too. For if ends are one of the basic components of maxims, as Kant indicates (G 43-44, 4:436-437) and as I have been arguing, then their nature can always help determine which maxims can be willed as universal laws. Moreover, Kant is clear that there are some objectively necessary, objectively obligatory ends, which do produce categorical imperatives. Humanity as an “end in itself”, for a

¹⁷ Again, compare (CP 18, 5:20-21) and (MM 15, 6:222).

prominent example (G 26-27, 4:428). And Kant even goes so far as to say that there could be no categorical imperatives at all if there were not some objectively necessary ends (G 37, 4:428), which he also describes as “ends that are also duties” (MM 147-152, 6:382-388).

Instead, Kant distinguishes hypothetical and categorical imperatives through how the “former represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will)”. So Kant’s point is only that the necessitation that a categorical imperative would place us under does not depend on whether or not we already accept the end contained in the imperative. In fact, the very essence of the distinction between hypothetical and categorical imperatives lies just in how the validity of categorical imperatives cannot depend *only* on an end that a rational being might or might not have, without thereby making a rational error either way.

So, in sum, hypothetical and categorical imperatives are both objective and universal, and both can contain ends. Only hypothetical imperatives are *conditionally* objective and *conditionally* universal, since their validity depends on whether or not an agent has already adopted a rationally contingent end. Categorical imperatives, by contrast, are *unconditionally* objective and *unconditionally* or *truly* universal. For each one is “objectively necessary of itself, without reference to another end”, and thus valid for all rational beings whatsoever, or at least for all rational beings of a certain kind. Which is why, when Kant describes the necessitation in a categorical imperative a little later, he says that categorical imperatives “would be most suitably named” by “*commands (laws)* of morality”, since “only law brings with it the concept of an *unconditional* and objective and hence universally valid *necessity*” (G 27, 4:416).

5.3 Categorical Imperatives and the Universality of Law as Such

So a categorical imperative turns out to be a formula that tells imperfectly rational wills what their unconditionally objective, universally valid obligations consist in. With that in mind, we can now approach the derivation of the Formula of Universal Law itself.

Let’s return to the crucial paragraph:

when I think of a *categorical* imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. (G 31, 4:420-421)

This part of the derivation is usually thought to be relatively straightforward and unproblematic, but it still needs walking through.

So: Kant starts off by asserting that the concept of a categorical imperative is sufficient by itself to determine what categorical imperatives actually require: “when I think of a *categorical* imperative I know at once what it contains”. He then proceeds to detail what a categorical imperative “contains”, which is just two things: “the law” and “beyond the law, only the necessity that the maxim be in conformity with this law”. So, to use our previous discussion to elaborate on his thinking here: an imperative is a formula expressing the command or requirements of an objective law. Thus, a categorical imperative contains the law that it formulates, by representing its requirements, and it expresses these requirements as the command of this law. This command thereby formulates the law’s necessitation—“the necessity that the maxim be in conformity with this law”—which the imperative then contains as well.

Now, Kant does not explain what he means by “conformity with this law”. He could mean that we must positively make the law into our maxim, or he could just mean that our maxim cannot violate this law. The second reading is weaker, however, and it will turn out that the Formula of Universal Law does not require anything stronger anyway. So let’s understand Kant in those terms.¹⁸

In any case, though, a categorical imperative may still command us to pursue an end, since an end could still be part of the law that the imperative contains. As a categorical imperative, though, its necessitation cannot depend on whether or not an agent already accepts anything to be brought about by conformity to the law. Hence, none of the agent’s rationally

¹⁸ Thanks to Barbara Herman for pressing for clarification here.

contingent ends can restrict what they must do pursuant to the law. So the “law contains no condition to which it would be limited”.

So it follows that a categorical imperative, by its very concept, “contains” only the law that it expresses and the categorical necessity in conforming one’s maxim to that law. Kant therefore concludes that “nothing is left with which the maxim of action is to conform but the universality of a law as such”. We don’t know the content of the law that categorical imperatives contain—that is precisely what we are trying to find out through the derivation. But we know that categorical imperatives command conformity to that law, in their distinctively categorical way. So if we can use the concept of a categorical imperative to determine the content of the fundamental categorical imperative, then all that is left for our maxims to conform to is the categorical character of the necessitation expressed in the imperative, with its distinctive kind of universality. (Which I will say more about in the next section.)

Here, Kant reveals that the derivation is giving an argument by elimination. His strategy is to consider the elements of a categorical imperative and reject the components that tell us nothing about what categorical imperatives require, until there is “nothing left” but something that tells us what all categorical imperatives must require. For if something in the concept of a categorical imperative contains a practical requirement or implies that we must act a certain way, then all categorical imperatives must contain that requirement, just by definition. From there, Kant then seeks to identify the fundamental requirement or principle of all categorical obligation.

So let’s begin again. As an imperative, a categorical imperative ultimately has two components: a content, which is the representation of an action, and a modality, because it represents this action as rationally required, as practically necessary. Thereby, a categorical imperative contains the law that it formulates and the necessity in conforming to it. But the content of the law, the required action, is precisely what we are trying to determine, and the mere fact that an imperative has content tells us nothing about what this content might be. We

might suppose that some of this content is objectively necessary—an objectively necessary end for the action, for example, such as pleasure, *eudaimonia*, or perfection have sometimes been thought to be.¹⁹ Kant actually does begin with the concept of an objective or categorically binding end when he tries to derive the Formula of Humanity (G 36-38, 4:427-429). However, the current question is about what follows from the concept of a categorical imperative alone, so we have to abstract from thinking about the content of the imperative entirely.

And again, an imperative is only an expression of the rational necessity in conforming to a law. So, “beyond the law”, a categorical imperative contains “only the necessity that the maxim be in conformity with this law”. But if we do not know what action we are to perform, knowing that we must necessarily perform it will not help to resolve that mystery. This necessity might be objective in some sense. But without knowing what sense, objectivity can only tell us that the necessity is based in the nature of the action required; it would not help us to determine what this action consists in at all.

Moreover, as I said before, hypothetical imperatives are also objectively necessary. So how can objective necessity alone tell us anything at all about what categorical imperatives require?

Unlike hypothetical imperatives, however, categorical imperatives are unconditional: a categorical imperative “contains no condition to which it would be limited”. But this just tells us that there is no restriction on categorical obligation; it tells us nothing about what a categorical imperative positively requires.

Ultimately, then, all that remains in the concept of a categorical imperative is their distinctive kind of objective necessity: their distinctive universal necessity, which Kant here calls “the universality of a law as such”. The concept of a categorical imperative implies that it exhibits this universality, and that it requires our maxims to conform to its universally valid content. So the concept of a categorical imperative requires our maxims to have “the universality

¹⁹ See Bentham (1988, 1), Aristotle (*Nicomachean Ethics* 1095a), and (G 49, 4:443).

of law as such” also, just because it forbids our maxims from violating this universality. So at least this much content follows just from the concept of a categorical imperative: all categorical imperatives must represent the necessity in conforming to the universality of a law as such, which is the necessity in making our maxims exhibit precisely this universality.

Therefore, anything else that a categorical imperative might require has to at least be compatible with the universality of a law as such. But Kant is actually after a stronger conclusion here: that “this conformity *alone* is what the imperative properly represents as necessary”, at least at bottom, so that “There is, therefore, *only a single* categorical imperative” (G 31, 4:421, emphases mine).

There could, of course, be many specific categorical imperatives, only they are all supposed to derive from conformity to the universality of a law as such. And this stronger conclusion is evidently supposed to follow from the argument so far. Indeed, in introducing the derivation, Kant says that “we want first to inquire whether the mere concept of a categorical imperative may not also provide its formula containing the proposition which alone can be a categorical imperative.” (G 30-31, 4:420)

But how does that follow from the argument just given? Kant has just given an argument by elimination, so he must be thinking that he has already eliminated everything else that could determine the content of a categorical imperative. How, though? Kant’s argument considered everything within the concept of a categorical imperative, but why couldn’t something outside of this concept determine or produce a categorical obligation, and thereby a categorical imperative as well? Perhaps the concept of a categorical imperative cannot tell us what all of our objectively necessary ends are, for example?

Well, to get his conclusion out of his strategy, Kant must be thinking that only the concept of a categorical imperative can tell us what categorical imperatives ultimately require. And given the kind of thing a categorical imperative is, this seems correct, or at least quite plausible. After all, a “categorical imperative would be that which represented an action as

objectively necessary of itself" (G 25, 4:414). So categorical obligation is intrinsic: if I want to know why I must not lie, the answer is because lying is intrinsically forbidden, since not lying is "objectively necessary of itself". In other words, there is something about the intrinsic nature of lying that makes it morally or categorically prohibited, and if you want to know why it is morally forbidden, the answer is just to examine its intrinsic nature more closely.

So, since categorical imperatives represent actions as necessary because of their own natures, they can ultimately only point out the intrinsic necessity in acting a certain way. This necessity, in turn, is just to conform to categorical, intrinsic necessity itself. If, in the final analysis, a categorical imperative said anything beyond "Do this, because its own nature makes it necessary for you to do it", then it would have to say that the action is required for something outside of that nature. But then the action could only be necessary as a means to this something else, and the imperative would become hypothetical.

So, ultimately, a categorical imperative can only say "obey categorical imperatives" or "conform to categorical obligation" or "perform actions that are intrinsically necessary". A categorical imperative that says only "obey categorical imperatives" does sound thoroughly empty, to be sure. Really, though, this self-reference or self-reflexivity in categorical imperatives means that the concept of a categorical imperative must be sufficient by itself to determine what categorical obligation consists in. For the underlying notion of intrinsic necessity is self-reflexive, too, in the sense that it refers our questions about it back to itself and its own intrinsic nature, which must therefore be sufficient unto itself.

Thus, the concept of a categorical imperative, as the concept of a formula expressing an intrinsic practical necessity, must be enough to determine the ultimate or fundamental categorical imperative by itself. And this is just what Kant's derivation by elimination aims to do: show us the intrinsic necessity in acting according to maxims that conform to the universality of a law as such, because nothing else either internal or external to the concept of a categorical imperative can direct our actions categorically.

5.4 The Subjective Derivation: Imperatives and Contradictions in the Will

Now, as I mentioned before, the derivation so far is widely viewed as sound, or at least as convincing. And it gets Kant far enough to say that our maxims must be able to conform “with the universality of a law as such”. But what does this universality consist in? Is it enough if our maxims could be universal laws, or must they actually be universal laws, as Aune, for example, suggests (1979, 29)?

However this issue gets settled, though, the next step in the argument is going to look highly problematic. As I noted already, Kant presents the Formula of Universal Law as following immediately from the derivation so far. But regardless of whether the universality of a law as such means possible, actual, or even necessary universality, Kant’s idea about willing maxims as universal laws seems new, and perhaps not warranted by the previous derivation. So what should we say on Kant’s behalf here, especially if our goal is to preserve the immediacy in the final inference to the Formula of Universal Law?

Let’s return to the fact that imperatives are essentially directed to the wills of imperfectly rational beings. As I said before, this fact allows us to consider the derivation of the Formula of Universal Law from two points of view: the subjective point of view, belonging to the imperfectly rational agents who are to be subject to a categorical imperative, and the objective point of view of the law whose demands on such agents are to be expressed in the formula of the imperative.

It is quite noteworthy, then, that Kant presents the derivation from the agent’s point of view. The derivation asks: if I know what kind of imperative I am under, can I work out its content? If the imperative is hypothetical, the answer is: clearly not, because I don’t know the end to be achieved. But if the imperative is categorical, then Kant thinks that I can work out its full content. Or at least its ultimate or fundamental content, which requires that our maxims conform to the “universality of a law as such”.

But what does this requirement really involve? The Formula of Universal Law is a claim about what we have to be able to will, and a maxim does not actually have to be a universal law

in order for us to be able to will it as such, at least not always. A maxim must be able to be a universal law, if we are to be able to will it as a universal law, but it need not actually be such a law. So the universality of a law as such only has to require our maxims to be able to be universal laws; it does not have to require our maxims to actually be universal laws, at least not always. Moreover, the weakest reading of “conformity with the universality of a law as such” only requires our maxims to be consistent with this universality. So, since the Formula of Universal Law doesn’t require anything stronger, we can maintain only this weak reading.

Willing maxims as universal laws is still a bit of a further step, but only a little. If we have to act according to maxims that can be universal laws, then we also have to will according to such maxims. So if there are any prior constraints on what we can and cannot will, these constraints will obviously apply here as well. And here is where the subjective point of view on the derivation becomes crucial.

Imperatives are essentially addressed to our wills, as I have been emphasizing. But our wills, as our faculties of practical reason, really are subject to a prior constraint: they must be internally consistent. The need for internal consistency or non-contradiction is at least part of what constitutes our wills as a faculty of reason.²⁰ And if Kant’s derivation is correct, then the need to conform to the universality of law as such will be part of what constitutes our wills as a faculty of *practical* reason of a specific kind. For if Kant’s derivation is right, then this requirement will be part of the fundamental categorical imperative for its subjects, which, as their fundamental categorical, unconditional principle, will also be the fundamental principle of practical reason in general, at least for rational beings of that kind. For it will be the fundamental principle from which they are to derive their actions, and “*reason* is required for the derivation of actions from laws” (G 24, 4:412).

So non-contradiction is a principle of reason in general, and a fundamental one, at that, while conformity to the universality of a law as such is a principle of a specifically practical

²⁰ As a negative principle, non-contradiction cannot be the whole story, in Kant’s mind. See note 16 above.

reason. Thus, non-contradiction is rationally more fundamental, rationally prior to, conformity to the universality of a law as such. So if categorical imperatives are to be principles for rational wills, they must already conform to reason's underlying demand for non-contradiction. And now the demand that our maxims be able to be universal laws transitions immediately into the demand that we be able to follow our maxims, considered as universal laws, without having to contradict ourselves. For the concept of a categorical imperative asks us to consider whether our maxims could be universal laws—whether everyone could follow them—while the prior principle of non-contradiction requires this 'could' to include its own rational requirements also.

5.5 The Subjective Derivation: Willing Maxims as Universal Laws

According to Kant, though, if a maxim can be a universal law, and if we can also follow it as a universal law without having to contradict ourselves, then we can also will it as a universal law. This is easiest to see if we take a step back and explore what Kant means by “willing something”.

As I explained in Chapter 2, Kant thinks that the will is a causal faculty (G 52, 4:446). Thus, to “will something”, in its proper sense, is to engage that agency or causality in order to bring that thing about. In this proper sense, ‘willing’ is keyed first to volition producing action, and then to these actions themselves.²¹ To be able to will something, in this sense, is therefore to be able to act to bring it about.

Therefore, when the Formula of Universal Law asks us to act on maxims that we could will as universal laws, the question it poses to us is ultimately about whether we are capable of making our maxims into universal laws. Of course, in one sense we are never capable of making our maxims into universal laws, because we totally lack the power to really accomplish that. However, that is not the sense that Kant is interested in.

²¹ Kant also gives at least some indication that he would accept a broader sense of “willing something”, wherein we also “will something” if we *would* bring it about, had we the power and motivation. In this sense, acts of will could also include a certain kind of wishing (MM 13, 6:213), and, arguably, all judgments about how the world ought to be. We will not need to rely on this broader sense here, however. Thanks to Barbara Herman and Tyler Burge for helpful pressure and examples on these points.

Instead, the Formula of Universal Law wants us to determine whether our wills would contradict themselves, were we to make our maxims into universal laws. That is evident from how Kant describes the circumstances in which we cannot will a maxim as a universal law: “it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself.” (G 33, 4:424) Thus, to know whether we can will a maxim as a universal law, we must consider it as a universal law—we must “raise” our maxim “to the universality of a law of nature”—and then determine whether our “will would contradict itself”. If there is no contradiction, then we can will our maxim as a universal law, and the Formula of Universal Law is satisfied.

Of course, many of the things that we will are not compossible. Moreover, universal laws, especially moral laws, will often introduce conflicts into our wills, since we often will things that conflict with their demands. Such conflicts need not raise problems under the Formula of Universal Law, however, so long as we can resolve them by prioritizing the universal law. In other words, as long as we can stop willing anything that conflicts with the universal law in question, we will have a way to will this law as a universal law without bringing our will into conflict with itself. So such universal laws will not come into conflict with the Formula of Universal Law.

Thus, we are unable to will a maxim as a universal law only when the “practical logic” of the maxim, so to speak, would *necessarily* lead our wills into an *irresolvable* conflict with themselves, were this maxim a universal law. In that case, the only way to restore coherence in our volition would be to give up on willing this maxim as a universal law.

So, in short, I can will a maxim as a universal law if and only if the corresponding universal law would not force my will into an irresolvable conflict with itself. Kant is not clear about whether these conflicts are supposed to follow from the law itself, or from an irresolvable conflict between the law and something else that we have to will on independent grounds. I argue elsewhere that both of these cases are genuine contradictions in the will, but that Kant

primarily has in mind the former case. This question will not matter here, though, since the nature of the contradiction is the same in both cases.

To wit: what does an irresolvable conflict in the will look like? Briefly: willing both one thing and its opposite. For example, to illustrate how to apply the Formula of Universal Law, Kant considers a man's "maxim of neglecting his natural gifts". There, Kant argues that this man "cannot possibly **will** that this [maxim] become a universal law", because "as a rational being he necessarily wills that all capacities in him be developed" (G 32-33, 4:423). Thus, the problem is that he would have to will both that his natural gifts fall into disrepair and that "all capacities in him be developed". For his maxim, considered as a universal law, requires him to will the former, while he must also will the latter, just because he is a rational being.

Similarly, Kant thinks that we cannot will a maxim of indifference to the welfare of others as a universal law, because "a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself." (G 33, 4:423) Here, the problem starts from the fact that a universal law based on this maxim of indifference would prevent us from receiving help from others, at least whenever we cannot expect to repay them sufficiently. So we cannot will this maxim of indifference as a universal law without willing to relinquish the disinterested aid of others. But abandoning the disinterested aid of our confederates is supposed to conflict with "the assistance that we wish for ourselves". To make this a contradiction within the will, Kant must be thinking that we rationally have to wish for this assistance, and that this wish is an act of volition.²² So, in the end, the problem with this maxim of indifference is that willing it as a universal law would ultimately force us to will both to receive the disinterested aid of others, since we rationally must will that, and not to receive that same disinterested aid, because the universal law of indifference would make such aid unconditionally unavailable.

²² See note 21 above, for some textual support for construing certain wishes as a kind of volition.

Putting aside whether these examples are convincing, the fundamental point is that willing both P and not-P is a rational mistake, just like believing both P and not-P. So, as we said, it doesn't matter why rational beings have to will that their capacities be developed or why they must will to receive the disinterested aid of their confederates. Their maxims, considered as universal laws, could commit them to both of those acts of will, or the commitment to one of them could come from another direction. But the nature of the contradiction is the same: either way they end up willing both P and not-P. Imperfectly rational beings are prone to such mistakes, of course, but they remain rational mistakes, and of a fundamental sort. Thus, I cannot will a maxim as a universal law if this universal law would commit me to a contradiction of this kind.

With that in mind, I can now tie the previous discussion together and finish the subjective derivation of the Formula of Universal Law. Commentators like Wood (1990, Ch. 9) point out, and Kant agrees, that there is no "inner impossibility" (G 33, 4:424) in a principle that everyone could follow, even though following it as a universal law would force us to contradict ourselves, in the way that we just described. As a result, nobody can will this principle as a universal law. But this principle can still be a universal law, so why can't it produce a categorical imperative that will also force our wills into irresolvable contradictions? And wouldn't that mean that the concept of a categorical imperative can't require us to be able to will our maxims as universal laws, as Kant's derivation of the Formula of Universal Law requires?

Well, as we saw in the last sections, all imperatives are essentially addressed to the wills of imperfectly rational beings. As such, all imperatives are already subject to the rationally prior demands of non-contradiction. Therefore, a genuine imperative cannot force our wills into an irresolvable contradiction. In other words, imperatives are to constitute norms of reason for imperfectly rational agents, because they are to express what imperfectly rational agents "ought" to do, just as rational agents. But the principle of non-contradiction is a more fundamental

rational norm for agents of this kind. So no imperative can express a genuine norm of reason, if it violates this more fundamental norm.

A categorical imperative, though, requires our maxims to be able to be universal laws. So if one of our maxims, as a universal law, forces our volition into a contradiction of this kind, then that maxim violates the prior constraint enforced by non-contradiction. Thus, that maxim also violates the concept of an imperative, and the concept of a categorical imperative *a fortiori*, even though this maxim could be a universal law, in the sense that everyone could actually follow it, so long as they make this fundamental rational mistake.

But all it takes to be able to will our maxims as universal laws is to be able to follow them as universal laws without having to contradict ourselves; that should now be clear. So if the concept of a categorical imperative really does require us to be able to follow our maxims as universal laws without contradicting ourselves, then it also requires us to be able to will our maxims as universal laws. Thus, Kant's derivation of the Formula of Universal Law is validated.

However, the objections to Kant's derivation usually do not proceed from the point of view of the agent and of the obligations that categorical imperatives impose on them, but rather from the point of view of the law that is to be formulated in a categorical imperative. This is not how Kant presents the derivation, as we have seen. But still, to give a complete defense of the derivation, we need to be able to answer these objections in their own terms. After all, their point of view on the derivation is equally valid. We can answer them by considering what it would be for a universal law to require our wills to contradict themselves.

5.6 The Objective Derivation

To reiterate the problem: Kant agrees that universal laws can force us to contradict ourselves. In describing the two ways in which we can be unable to will a maxim as a universal law, he explains that

Some actions are so constituted that their maxim cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* that it *should* become such. In the case of others that inner impossibility is indeed not to be

found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself. (G 33, 4:424)

The upshot of which is that there is no “inner impossibility”, no “contradiction in conception”, in thinking of a universal law that we cannot will as such. Wood, among others, thinks that this means that Kant’s derivation of the Formula of Universal Law fails, because Kant has not shown that such a universal law cannot be a categorical imperative. Has Kant implicitly condemned his own argument?

I have already given one answer to this question, but to look at the issue again from the point of view of the laws that categorical imperatives are to formulate, it will help to begin by determining more closely what this “inner impossibility” consists in, which prevents certain maxims from having the form of law.

The clearest case here is the famous example of the lying promise. Here, Kant states that the “maxim of action would go as follows: when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen.” (G 32, 4:422) So when we consider this maxim as a universal law, we are to imagine what would happen if everyone acted this way, for this end, whenever this ground obtains: what would happen if everyone lied, in order to get a loan, whenever they believe that they need money?²³ Could everyone actually fulfill these requirements?

No, Kant argues, because if everyone obeyed this maxim, that “would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him.” (G 32, 4:422)²⁴ If everyone who thought they needed money always made this

²³ I am going to rely on the account of the structure of practical principles defended in the last chapters, for ease of exposition and the sake of continuity. But the argument here will not depend on that account.

²⁴ Kant’s idea that the promise itself would also be impossible might be a bit puzzling. Surely one could still make the false promise, even if it would not be believed? I take it, though, that Kant’s point is that a promise that is not believed fails in two ways: it fails both to secure the end of the promise, and to be taken as a genuine promise. If you believe my promise, but still decline to loan me any money, my promise has still succeeded in a way. So Kant is thinking, I take it, that the promise would be guaranteed to fail in both ways, were this maxim to become a universal law.

false promise, then these promises would not be believed, and the loan would not be granted, unless some other factor, not mentioned in the maxim, intervened. Your word alone, based on nothing more than the belief that you need money, would not be enough to secure the loan, and so you could not satisfy the terms of the maxim. Some creditors might take pity on you, but the connection between *this* means and *this* end, as contained in the maxim, would still be broken, and the lie would no longer yield the loan.

Or so Kant seems to be thinking. At any rate, for this specific case, at least, Kant makes the relation between the action and the end the locus of the conflict, under the idea that the universal use of these means to this end would make the latter unattainable through the former. And such maxims cannot be universal laws, Kant thinks, if they cannot always determine what will happen. So, extrapolating from the case of the lying promise, these “contradictions in conception” consist in this instrumental conflict between the action and the end contained in a universalized maxim.²⁵

Certainly there are many objections to Kant’s picture here, and it is difficult to tell how to extend this account of contradictions in conception to other cases.²⁶ But even the commentators most critical of Kant’s derivation of the Formula of Universal Law are usually prepared to concede that the concept of a categorical imperative rules out these kinds of conflicts. Instead, the issues that are more specific to the derivation of the Formula of Universal Law come in later, with contradictions in the will.

So now, notice that if a universal law produces an irresolvable conflict in the will, then that law will face a problem analogous to a contradiction in conception. When a universal law

²⁵ Kant also uses the case of suicide to illustrate contradictions in conception (G 31-32, 4:421-422). This case looks hard for my interpretation, because the conflict there seems to include the maxim’s ground, as well as its ends and means. But it is not clear that this is a good case for Kant, either, and it is intended to illustrate a different kind of duty than the lying promise case. So we can restrict our account to the relevant kind of moral duty, if necessary.

²⁶ For example, Hegel (PR §135) is often understood to allege that any maxim can be a universal law, and sometimes also that any maxim contradicts itself as a universal law also. This is a problem for the next chapter.

can't always be followed, it can't be satisfied *naturally* or *in nature*. In contrast, a universal law that forces us to contradict ourselves, or to make some other rational mistake, can still be satisfied in nature. As imperfectly rational beings, we are, after all, eminently capable of rational error. But such a law is still not satisfiable in another way: since it cannot be satisfied without making a rational mistake, it is *rationally* unsatisfiable. Its demands are rationally impossible. For imperfectly rational beings like us, for whom the laws of reason are fundamentally norms, we cannot fulfill such a law without violating a norm to which we are fundamentally subject.

Accordingly, such a law is still deeply incoherent. It can specify what will happen in the world, but it cannot specify how what happens in the world should be judged or evaluated, rationally speaking. Remember that imperatives “are expressed by an *ought*” and “say that to do or to omit something would be good” (G 24-25, 4:413). So the law that an imperative formulates must imply or require that the actions pursuant to it are good and ought to happen, if its imperative is to be expressed by an “ought”, and if that imperative is to represent those actions as good. For the expression of the law can contain nothing that isn't already in the law itself. Also, recall that Kant says that “A perfectly good will would, therefore, equally stand under objective laws (of the good)” (G 25, 4:414). So here we are considering a law that says that two incompatible things are good and that both ought to be realized in the world.

To see the implications of this point, consider indifference to the welfare of others again. According to Kant, the man in this example cannot will his maxim as a universal law because “by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself.” (G 33, 4:423) There would be no contradiction here if he could give up wishing for assistance, and, as I said before, wishing has to be a kind of willing here. So the problem is that this maxim, as a universal law, ends up forcing him to will both to be assisted when in need and not to be assisted when in need.

Now, there are two ways to understand this contradiction: first, the maxim, as a universal law, could by itself imply that we must will both to be assisted and not to be assisted.

Alternately, this universal law could only require that we will not to be assisted, while something else requires us to will to be assisted.²⁷ In the former case, it is clear that the maxim, as a universal law, cannot produce a coherent representation of what would be good or of what ought to happen, since it will end up saying that we ought to will both to be assisted and not to be assisted. This is immediately incoherent.

But the second case is no better. Here, we must consider the need to will our own assistance as already established, since it is supposed to be independent of the maxim under evaluation. So we must suppose that it really is necessary to will our assistance, objectively speaking, and, correlatively, that it really is good to will this and that this really ought to happen. A maxim, then, which would require us to will the opposite, were it a universal law, still cannot provide a coherent take on what is good or what ought to happen, because it contradicts a principle already established about precisely this.

Now, again, none of these considerations calls into question whether this maxim could actually be a universal law. Nevertheless, such a universal law cannot be expressed in a categorical imperative, because it contradicts the concept of an imperative in general. An imperative is supposed to represent what ought to happen, what would be good to happen, and it is supposed to address this representation to a rational being. So if this representation is incoherent, the imperative is incoherent, and it cannot coherently command anything at all, still less can it coherently command anything of a rational being. A law with incoherent requirements would violate the concept and preconditions of norms of reason, in short.

Moreover, notice that there is an odd result here: if a universal law requires a contradiction in the will, then it can only be followed if an agent makes a rational mistake or violates a norm of reason. Therefore, only imperfectly rational beings can follow such a law. Perfectly rational beings are incapable of such errors, so they could not follow this law.

²⁷ A “true need”, for example, as Barbara Herman has persuasively argued, in her “Mutual Aid and Respect for Persons” (Herman 1993, 45-72).

Therefore, it could not be a law for them, and this time really in the same way that Kant supposes that the maxim of the lying promise cannot be a universal law in nature. But if a principle can't be a law for perfectly rational beings, how can it necessitate an imperfectly rational will? How can such a principle constitute a rational norm for them? Accordingly, it would be impossible for such a universal law to be expressed in a categorical imperative, and, crucially, because this law is incompatible with the concept of a categorical *imperative*. Therefore, the concept of a categorical imperative really does require our wills to cohere with themselves. Thus, Kant's derivation is validated here also, even when we consider the concept of a categorical imperative from the point of view of the law that it is to formulate.

5.7 Conclusion

To be sure, much more work remains to be done to defend Kant's derivation of the Formula of Universal Law from the many more specific objections that have been lodged against it. (Work which the next chapter will undertake.) However, I have still offered a general solution here. Two, in fact. Kant seeks to derive the Formula of Universal Law from the concept of a categorical imperative. So, by emphasizing the concept of an imperative and how they relate objective, universal laws to the wills of imperfectly rational agents, I was able to examine Kant's derivation from two points of view: the point of view of the agent subject to a categorical imperative, and the point of view of the law to be formulated in a categorical imperative. From the former point of view, I argued that a categorical imperative cannot force our wills to come into conflict with themselves. Our wills are rational and so they cannot be subject to an imperative that would force them to violate a fundamental, prior principle of reason, such as the principle of non-contradiction. Conversely, from the point of view of the law, I argued that a universal law could force our wills into internal contradictions, but such a law still could not be formulated in a categorical imperative, because what this law requires is rationally incoherent. Either way, Kant thinks that we can will a maxim as a universal law whenever we can follow it as a universal law without having to contradict ourselves. From both points of view, then, the

concept of a categorical imperative really does require us to be able to will our maxims as universal laws, and both presentations of the derivation preserve Kant's sense that the argument is immediate and direct, as we sought to do.

In short, the fundamental strategy of this paper has been to insist on Kant's distinction between imperatives and the laws that they represent, just as past chapters distinguished principles from the laws that they represent. Once we make this distinction, we learn very quickly that categorical imperatives are really laws in two ways: first, as categorical, they are to be universal laws, and so laws of what is to happen in nature. But, as imperatives, they are to be universal laws not just for nature, but for distinctly *rational* nature. As such, these laws have to constitute rational norms for imperfectly rational beings. And, as this latter kind of law, categorical imperatives immediately need to be rationally coherent, so that we always have to be able to will coherently, no matter what else we have to will. A universal law of nature can force us to will inconsistently. Kant himself recognizes this. But a law of rational nature cannot do likewise, so such a universal law can be no imperative at all, still less a categorical imperative.

In the end, this is why Kant can say that categorical imperatives can *only* require our maxims to conform to the universality of law as such, and still go on to say that we must be able to will our maxims as universal laws. Imperatives already require consistency in volition, so adding the former is enough for the latter.

In closing, then, I want to make an observation about the objections to Kant's derivation: they get the structure of the derivation exactly backwards. In fairness, Kant is easily misleading here. He presents the derivation as a kind of progression from the concept of a law, to the concept of a rational will, to the concept of an imperative, to the concept of a categorical imperative, to the idea that we must be able to consider our maxims as universal laws, to the idea that we must be able to will our maxims as universal laws. Seeing this progression, the objections target the last step, and accuse it of invalidity. But when we notice that the derivation's fundamental concept is not universality, but rather the concept of an imperative, we

can see something that perhaps should have been apparent from the very beginning of the *Groundwork*: Kant is, and has always been, looking for the fundamental principle of a good, albeit imperfect will (G 14-15, 4:402). As such, that principle always had to be addressed to the will. Hence, this fundamental principle always had to be an imperative, and so it always had to be subject to the demands of non-contradiction.

Commentators like Wood ask: why couldn't a universal law necessarily lead to conflicts in the will? But Kant's question was never about what the universal laws are *simpliciter*, but about what the universal laws for wills like ours might be. As such, the fundamental principle that Kant is after always had to be a principle about what we can and cannot will. And, since the will is to be a faculty of reason, what we can and cannot will always had to be subject to consistency requirements. The project of the derivation then, is just to fill in the content of what a good will is supposed to will, and the answer is: maxims that can be universal laws.

So the ultimate moral of our discussion here is that the entire *Groundwork* has always been operating within the parameters of a requirement that our wills cohere with themselves. Perhaps paradoxically, then, the real philosophical difficulty in the derivation lies just in the part that Kant's critics are usually prepared to concede. In fact, when we put the derivation my way, we can see that our fundamental requirement is really just to will consistently, and the derivation just applies this general, fundamental requirement to the specific content that morality requires us to will. If morality exists at all, anyway. Thus, the need for our maxims to be coherent as universal laws is actually posterior to the requirement to will our maxims as universal laws, not prior to it, as Kant's order of discussion suggests, and as the objections require. And, again, perhaps this need not come as much surprise: if our fundamental requirement were not to will consistently, then it would become very hard indeed to see how Kant could ever hope to show that our autonomy as rational agents consists in following whatever else this requirement turns out to be. For a law that ineluctably leads our rational wills into irresolvable conflicts with themselves must be a law that dominates and enslaves our wills,

not one that expresses their fundamental freedom, autonomy, and authority over all of our actions.

So, in the final analysis, Kant's derivation of the Formula of Universal Law is just this: As beings with a will, we must will consistently, and as moral beings with a will, we must will maxims that are consistent as universal laws. The derivation simply combines these two requirements, one of reason in general, and one of distinctively practical reason, into the Formula of Universal Law. Nothing more.

6. An Empty Formalism?

I have now explained Kant's derivation of the Formula of Universal Law in the *Groundwork of the Metaphysics of Morals* from two points of view: from the point of view of the imperfectly rational agents that are to be subject to categorical imperatives, which is the point of view that Kant himself adopts for the derivation, and from the point of view of the laws that are to issue in categorical imperatives for such agents, which is the point of view usually taken by the general objections to Kant's derivation. But, as I mentioned before, there are many more specific objections to Kant's derivation, as well as to the cogency of the Formula of Universal Law itself, considered as the fundamental principle of morality.

I have not yet addressed any of those specific objections, so in this chapter I want to address a family of those objections, which is loosely organized around the idea that Kant's derivation fails due to its emptiness. For example, Hegel and Allen Wood allege, in different ways, that the basic concepts involved in Kant's derivation are empty of content, or at least too empty of content to yield the fundamental principle of morality, as Kant intends them to do. So these objections allege that the premises of Kant's derivation are empty, or at least too empty to serve Kant's theoretical purposes. But Hegel also alleges that the derivation's conclusion—the Formula of Universal Law itself—is empty as well, in the sense that it systematically fails as a criterion for separating morally permissible action and volition from morally impermissible action and volition. The Formula of Universal Law is “normatively empty”, we might say.

Relatedly, there has been a long-standing set of objections claiming that we can always find a maxim that we can will as a universal law, no matter what we propose to do. These objections allege that we can will any maxim whatsoever as a universal law, so long as the principle is framed so specifically that we can guarantee that it can only ever apply to a single case. In other words, these objections would have it that we can will anything that we want as a universal law, so long as we tailor or “gerrymander” our maxims so that they can only ever apply to the unique case in which we propose to act. In which case the Formula of Universal Law also

turns out to be normatively empty, at least in the final analysis, because its strictures are always easily defeasible.

Last, there is a more particular problem about the kind of universality that the Formula of Universal Law requires. This problem is perhaps not best described as an emptiness problem *per se*, but it bears enough of a “family resemblance” to the other problems to be worth discussing alongside them. Namely: the Formula of Universal Law requires our maxims to have a certain kind of universality. But some intuitively permissible maxims rely on a certain specific kind of social coordination or cooperation, which would appear to be destroyed if everyone followed these maxims. Thus, it looks as though these permissible maxims could not be universal laws, in the way that the Formula of Universal Law requires. At the same time, this same universality sometimes seems to project universal social practices that would be intuitively immoral, but possible to endorse, accept, or will as universal all the same. So the universality that the Formula of Universal Law requires seems to give both false positives and false negatives, once we take into account how this universality interacts with certain kinds of social coordination and patterns of social practice.

For example, a maxim like “I will play tennis every Sunday morning at 10:00” seems morally permissible, at least when we assume that some people will want to play tennis at other times, or not at all. So there seems to be a backdrop of social coordination or cooperation that enables us to follow this maxim. However, if this maxim were to become a universal law, it seems as though it would destroy that very social cooperation. After all, if everyone tried to play tennis at that exact time, then it seems as though there would not be enough courts to go around, and then some of us would be unable to follow this law. So this maxim seems to fail the Formula of Universal Law, even though it also seems clearly permissible, morally speaking.

Similarly, a maxim like “I will pursue a career in academic philosophy” looks as though it would, were it a universal law, undermine the division of labor that we rely on in order to feed, clothe, and shelter ourselves, among other things. This maxim also looks permissible, but, again,

it seems to be impossible for everyone to follow it, because then nobody would grow food, etc. So here again, the Formula of Universal Law seems to violate our moral intuitions, by asking us to adopt a frame of analysis that undermines the social coordination that many of our plans and projects relies on.

Last, a maxim like “I will avenge the insults that I suffer”, when considered as a universal law, seems to project an entirely possible universal social practice of policing insults through acts of vengeance. And couldn’t we also accept this practice, so long as we are prepared to bet that we will never offend anyone, or at least if we are willing to have vengeance wreaked upon ourselves, from time to time? This maxim looks as though it should be immoral, but it also seems to satisfy the Formula of Universal Law, because of the kind of social practice that it creates or projects, when we consider it as a universal law.

Social practices are at issue in Hegel’s objections to Kant’s derivation as well, and the gerrymandering problem is ultimately about how maxims are properly formulated. Both sets of issues recur in this problem, as we will see later on. Hence, all three issues are of a piece, even if this last one is not specifically about whether or not the Formula of Universal Law or its derivation is empty in some sense.

To finish our defense of Kant’s derivation of the Formula of Universal Law, then, all that remains is to offer answers to these more specific objections. This will certainly not answer all of the specific objections to the Formula of Universal Law, or to its derivation, but they are all that we have space for here. I will treat these issues in three broad sections: Hegel and Wood on the emptiness of Kant’s derivation of the Formula of Universal Law, the tailoring or “gerrymandering” problem, and the problem of social coordination and cooperation.

6.1 Hegel: *Moralität* and *Sittlichkeit*¹

Hegel raises complicated objections to Kant’s practical philosophy on several different levels, and these objections are often quite hard to separate from Hegel’s broader philosophical

¹ Thanks to Tyler Burge for many helpful suggestions throughout this section and the next.

project and system. In fact, I will ultimately want to conclude that Hegel's objections have no force apart from that broader system.² If I am right, then the issues between Hegel and Kant will really be about whose system is better, all things considered, rather than about whether a specific objection is sound.

To reach that conclusion, however, we will have to start out by trying to locate and explain specific objections in Hegel's writing. Before I discuss his specific texts, however, I would like to begin with what his basic objection to the Formula of Universal Law is usually understood to be.³ It goes something like this:

Suppose that I make it my maxim to steal whatever I want, whenever I think that I can get away with it. By the lights of my account of Kant's practical principles, my maxim would then be something like "Whenever I want something, and believe that I can get away with stealing it, I will steal the thing that I want." The standard analysis of this maxim under the Formula of Universal Law would then hold that this maxim cannot even be a universal law, to say nothing of whether or not I can will it as such. If everyone stole whatever they wanted whenever they thought that they could get away with it, then we would destroy the institution or social practice of private property. Thus, this maxim is not supposed to be able to be a universal law in the first place.

But Hegel is understood to allege that there is really no such problem here at all. He would point out that we have to understand this maxim as committed to private property in order to think that it would undermine itself in the fashion just described, were it to become a universal law. Why should it be committed to private property, though? Even if I adopt this

² In fact, Hegel will presumably agree that at least the ultimate force of his criticisms of Kant's ethics cannot be separated from his broader project at all. After all, the Preface to the *Phenomenology of Spirit* famously pronounces that "The True is the whole" (PS ¶20), and even begins with a critique of the project of a preface to a truly philosophical work, leading up to the idea that "The true shape in which truth exists can only be the scientific system of such truth." (PS ¶15) This makes it quite difficult to discuss his disagreements with Kant, at least if we wish to avoid having to compare their two systems as two independent wholes (and we most assuredly do want to avoid that).

³ Thanks to William Bristow for pressing me to say more here.

principle as my maxim, why should I have to have any views on the institution of private property at all to go along with it? Or, even more strongly, what if I positively want to destroy private property, and universal theft is my proposal for taking private property down “from the inside”, so to speak?

Either way, though, it seems as though I could accept, or even positively will, the destruction of private property as a consequence of universal obedience to my maxim. How would my will end up having to contradict itself, if I did accept this consequence? But Kant thinks that if a maxim cannot be a universal law, then I cannot will it as a universal law, either. So, by contraposition, if I can will my maxim as a universal law, then it must also be coherent as a universal law. As, in fact, this universal law seems actually to be: if I can will the destruction of private property as a consequence of my maxim, taken as a universal law, then an actual universal law with this as its consequence would be working exactly as my volition would have it do. Such a law does not really seem to undermine itself after all, then. Quite to the contrary, in fact.

The Hegelian point here is then usually supposed to extend to any maxim whatsoever: to consider a maxim as a universal law is to consider it as the basis for a complex social program and practice. To imagine every human being always acting according to a certain principle is, after all, by definition to consider the maxim as setting the agenda for a complicated and thoroughgoing social arrangement. So why can't I will for this universal social practice to gradually destroy itself? And if I can do that, then wouldn't my maxim work exactly as specified, were it to become a universal law?

This reasoning is supposed to illustrate Hegel's general point, which is supposed to be that we can always just adjust our volition in order to accept the consequences of our “universalized” maxims. There is no logical contradiction in any actual state of the world—how could there be, if the world is actually that way, *ex hypothesi*? Instead, whatever contradictions there may be must be in us, or in our attitudes. Therefore, as long as we are prepared to accept

the worlds that our universalized maxims would create, can't we will our maxims as universal laws, no matter what these maxims say? So can't we really will any maxim whatsoever as a universal law, and then isn't the Formula of Universal Law empty, normatively speaking?

In the next section, we will look at some passages from Hegel that offer some support for reading him as objecting to Kant in this way. For now, though, I want to focus on this "standard Hegel" as an objection to the Formula of Universal Law's "contradiction in conception" test—as an objection to the idea that some maxims cannot be universal laws, because there is a contradiction in how they would be conceived of, as universal laws. "Standard Hegel's" idea here is that any maxim can be a universal law, as long as its agent is prepared to accept the actual consequences of that law, whatever they might turn out to be.

One possible response on Kant's behalf would admit that we posit a universal social practice whenever we consider a maxim as a universal law. But Kant could think that in positing this practice as *universal*, we are committed to its persistence. If I posit a *universal* practice of theft, then it sounds as though I am committed to the universal persistence of this practice. That commitment would appear to be violated, though, by how this practice undermines itself. I sounds as if I cannot steal anything if there is no private property, so a universal practice of theft seems to both require universal private property, while destroying it at the same time.

However, a commitment to the persistence of a social practice seems wholly compatible with willing for that practice to undermine itself eventually. My maxim of theft, considered as a universal law, might commit me to private property for a while, but indefinitely? This maxim says that when someone owns something that I want, I must steal it, assuming that I think that I will not be caught. However, the dissolution of private property only guarantees that the antecedent or ground of this conditional will stop obtaining. So how would my will contradict itself, if I willed this maxim as a universal law, and accepted that the law would stop coming into effect after a certain time, because its conditions of application stop occurring? Couldn't I hold that the law remains valid, while recognizing that these conditions will stop being met?

As we will see in the next chapter, the Formula of Universal Law might ask us to evaluate our volition from an atemporal, *noumenal*, or purely rational point of view. From that point of view, the fact that a social practice undermines itself over time may be of no consequence: that point of view abstracts from temporal processes and relations, so the fact that the practice undermines itself *at all* might be enough to show that we cannot will that practice as a universal law, from this point of view. Nevertheless, though, that result, if it is one, might not show that there is any problem with willing a self-undermining social practice in temporal nature. Accordingly, thinking of the maxim as a universal law *of nature* would not seem to be able to help Kant any, either.

So why must there be a contradiction in conception wherever the universal social practice would undermine itself? Can't the agent always get around these supposed contradictions simply by making the dissolution of the social practice part of what they will? To put the question back into the terminology of Kant's derivation of the Formula of Universal Law: why should the concept of a categorical imperative prevent us from willing a social practice that will eventually destroy itself? And, anyway, how can the question about willing our maxims as universal laws have any real meaning when we are considering a thoroughgoing social program in total isolation from how and whether it would ever be embodied in an actual society?

In fact, examples like these are supposed to illustrate the differences between Kant's and Hegel's entire approaches to ethics. By considering the individual's volition in isolation, Kant is alleged to remain within the limited position of *Moralität* (morality), which provides only partial and one-sided guidance for our actions. The true position, according to Hegel, is supposed to be that of *Sittlichkeit* (usually translated as "ethical life" or "ethical substance"), wherein the individual's actions are judged within the context of their broader social structures

and ideologies.⁴ Conflicts within these social structures and contradictions within or between these ideologies are supposed to reveal their limitations and drive them towards better and truer forms of *Sittlichkeit*. But such conflicts need not reveal that any particular individual is acting wrongly, at least not in any very straightforward and uncomplicated sense.⁵

In fact, Hegel thinks that ethical tragedy occurs when individuals embodying the contradictory ideals within a form of *Sittlichkeit* come into conflict with one another (PR §140n, 180n), as Hegel thinks the characters of Antigone and Kreon from Sophocles (1982) illustrate perspicuously. Moreover, an individual who drives towards the undermining, dissolution, and transcendence of a form of *Sittlichkeit* could ultimately be an ethical or a moral hero. Indeed, even one of world-historical proportions. (Think of the abolitionist movement, for example.)

But the trick here, of course, has to do with how we understand the maxim in question. It is surely possible for us to will for a social practice or institution to undermine itself over time. It might even be rational to will that, in certain cases—systematic racism and other forms of bias and oppression might be good candidates. But it is also possible for us to will the indefinite persistence of that very same practice or institution. And because we are imperfectly rational, our capacity for rational error “enables” us to will both of these contradictory things at the same time. In fact, because we are imperfectly rational, there simply can never be a commitment that we can undertake without the risk of putting ourselves into contradiction with ourselves.

Along those lines, attentive readers may have already noticed that the maxim in question, as formulated above, is not entirely clear about the ends and the means that it is to include. Moreover, recalling the discussion of practical judgment from Chapter 4, a lot hinges on how “stealing” is to be understood. For example: under one interpretation—quite probably the

⁴ I can, of course, only give a quick and dirty version of only one interpretation of Hegel’s ethical thought, which is intended primarily for illustrative purposes. The next section will look at a passage mentioning this difference between *Moralität* and *Sittlichkeit*, however.

⁵ The picture is complicated, because individuals will know that adhering to one ideal requires violating the other, as Sophocles is about to illustrate, on Hegel’s behalf.

normal one, adopted by actual agents—this maxim, as stated, really is indefinitely committed to private property. If “stealing” is to mean taking something that does not belong to you, in order to make it into your own private property, then this maxim really would undermine itself, were it to become a universal law, in precisely the way that the standard Kantian response to the problem would have it do. For your end is to acquire private property, but you would be unable to have private property if everyone tried to acquire private property following your example. You want to be stable and assured in your possession of the thing, as the institution of private property is supposed to ensure, but that is precisely what your maxim, as a universal law, would deny to you. And whether or not “stealing” does involve the acquisition of private property hinges on how the agent conceives of what they are doing and what they hope to accomplish by it.⁶

So the standard Hegelian objection here does not really apply to the maxim at issue, conceived along those lines. Willing the dissolution of private property is no help in resolving the contradiction here, because the maxim is already committed to its indefinite persistence. But if we understand the maxim to be indifferent to private property, or actively hostile to it, even, then those are simply separate questions. Can I will, as a universal law, to take what I want, when I think that I won't be caught taking it, and to use what I take for my own purposes, but without thereby seeking to make anything my own property? (Perhaps that is conceptually incoherent, too? Perhaps my purposes with the thing commit me to thinking of it as my property, or perhaps they constitute a claim of private property? Or, at any rate, perhaps my purposes for the thing would require me to object if someone tried to follow my example and take it from me? And why am I worried about getting caught? Is that a tacit recognition that my action is wrongful, in some sense, or from someone's point of view? Can you will for others to try

⁶ You might think that it is simply conceptually impossible to acquire private property through theft. Private property implies rightful ownership, one might assume, while theft implies wrongful acquisition. This conceptual incompatibility is precisely what the Kantian response I am trying to develop wants to suggest. Yet, imperfectly rational beings that we are, it is entirely possible for us to will something that is conceptually incoherent.

to get away with things, in that fashion?) Can I will a maxim of destroying the institution of private property by acting in this fashion? (If the former is incoherent, perhaps this is also, by parity of reasoning?)

At minimum, these are complicated questions, which would require much more tinkering with maxim formulation and much deeper Kantian analysis. In any case, though, these are such different interpretations of “stealing” that they ought to be viewed as representing different maxims, which we might even conclude do not really involve stealing at all, strictly speaking. So I have shown at least this much, *contra* “standard Hegel”: that there are some maxims that really cannot be universal laws, because they are both self-undermining and committed to their indefinite persistence. That result does not resolve all of the questions about the Formula of Universal Law here, but it is enough to show that it is not normatively empty.

Really, though, the ultimate Hegelian counter-response would just be to insist that willing a contradiction is often not an unambiguous rational or ethical mistake at all, except in some ultimate, absolute, “final analysis” sense, which is usually not available to agents when they act. Instead, relative to the context of certain forms of *Sittlichkeit*, these kinds of contradictions are precisely what agents ought to will, given that their form of *Sittlichkeit* is actually at odds with itself, and also because such contradictions are how forms of *Sittlichkeit* evolve, improve, and transcend themselves.

For example, from the point of view of ancient Greek *Sittlichkeit*, the ethical conflict between Kreon’s political obligation to the city of Thebes and Antigone’s filial obligation to her brother is irresolvable. (Hegel thinks.) Her brother led a rebellion against the city, so he must be denied proper burial as punishment and as an example to others. But the filial obligation to perform a proper funeral recognizes no such limitation, and their *Sittlichkeit* includes no way for them to reconcile the imperatives of the city and the imperatives of filial piety. So Kreon and Antigone both exhibit moral nobility and an admirable obedience to their respective ethical

ideals, and neither of them needs to view the other as acting wrongly, except to the extent that they recognize these ethical ideals to be in conflict with one another.

Thus, each of them really ought to act in a way which is at once the highest expression of one ethical ideal and the deepest transgression of another. This conflict is both ethical and tragic, precisely because those ideals, when properly applied to their situation, really do imply that her brother should both be buried and not buried. For these ideals are part of a single *Sittlichkeit*, only one part of this whole requires burial, while the other requires punishment and a political example. So this *Sittlichkeit* really does contradict itself, and both of them would follow its logic correctly if they willed for her brother to be both buried and not buried.

Furthermore, Hegel thinks that forms of *Sittlichkeit* only evolve through some kind of real, lived collision between their conflicting and irreconcilable ideals. So, from the point of view of more developed forms of *Sittlichkeit*, Kreon and Antigone do make unambiguous rational and ethical mistakes, but those mistakes also appear necessary to produce the more developed, subsequent forms of *Sittlichkeit*. So these later, more developed points of view are able to see how their forebears were unambiguously mistaken, but not relative to their own *Sittlichkeit*, or relative to the development of *Sittlichkeit* itself.

So when we look at the disagreement between Kant and Hegel this way, we find that it is really a matter of broader ethical frameworks and, ultimately, even a matter of logic. Kant's classical, Aristotelian logic holds that willing both P and not-P is always an unambiguous rational mistake: Antigone always violates a fundamental principle of reason if she wills both for her brother to be buried and not buried.

Hegel's dialectical logic, though, need not recognize this contradiction as an unambiguous rational mistake. For Hegel, two things—not necessarily propositions, but realizing them—are put into conflict with one another, and that conflict is ultimately to produce a third thing, different from both of the elements in the original opposition, but combining the proper elements of each. In Antigone's specific circumstances and ethical context, two outcomes

are possible if she wills both for her brother to be buried and not buried: ethical tragedy, as in the play, or she can violate the parameters of the *Sittlichkeit* that leads to this contradiction, leading the charge to a new form. The former is what her form of ethics requires, while the latter is something like the core nature of logic, reason, and world-history for Hegel. Either way, following that process out is the opposite of a rational or ethical mistake, at least in many cases.

In this way, the issue about what we can and cannot will without making a rational mistake turns into a disagreement about logic, not one about morality itself. It is not my purpose here to evaluate the extent to which Hegel intends to, or succeeds in, offering anything resembling a plausible or compelling alternative to Kant's Aristotelian logic. So far, we certainly have not seen any reason at all to prefer Hegel's picture, and, as we will see more fully over the course of the subsequent discussion here, Hegel's ideas about logic are so different, and so deeply embedded in his entire metaphysical, epistemological, and ethical project, that it would be quite fair to wonder about the extent to which his "logic" deserves the name in the first place. In fact, this systematic divergence is the whole crux of my argument here. And in any case, I certainly have no interest in challenging Aristotelian logic myself, especially when my project here is to offer a defense of Kant. Moreover, given the history of logic from Aristotle to Frege to the present, the burden is clearly on Hegel and his followers to dethrone that logic, if that really is their intention.

Notice, though, that we started off with what looked to be an objection to Kant's stance on whether or not we can will a particular maxim as a universal law. So if the issue is now about fundamental logic, then we've lost something of the sense in which Hegel is objecting to Kant's moral theory *per se*. The simplest case, though—the one that we started out with, the one that would be most clearly devastating to Kant, and the one which, I daresay, captures what most thieves actually have in mind—that case can actually be answered quite straightforwardly, and in the terms of the standard response. Beyond that, we would need to understand whether Hegel is right about the differences between *Moralität* and *Sittlichkeit*, and about his dialectical logic.

6.2 Hegel and the “Empty Formalism”

As I indicated before, the objections coming more directly out of Hegel’s texts are far more complicated. For example, besides rejecting the Formula of Universal Law as an “*empty formalism*” (PR §135), he also criticizes the idea that moral worth depends on acting from duty or for the sake of duty (e.g. PS ¶637ff), as well as the entire moral standpoint that he understands Kant to be operating from (e.g. PS ¶616ff). However, it is often quite difficult to extract specific arguments for these criticisms, and especially to find arguments that can be isolated from Hegel’s broader, systematic project. Despite these difficulties, though, we can draw out a few lines of reasoning, at least as far as the “empty formalism” objection is concerned.

Even here, there are many ways to understand the objection; I will attempt to track only a few. To begin with, the strongest version of the objection holds that the Formula of Universal Law fails to provide any criterion at all for separating moral and immoral actions, since, this objection alleges, the maxim of any action whatsoever can be shown to satisfy its requirements.⁷ The previous section covered the standard reading of this objection, which, if some version of it is correct, would mean that the requirements that the Formula of Universal Law imposes on our maxims will turn out to be totally empty: hence, an “empty formalism”.

This objection is obviously decisive against Kant’s whole ethical theory, if it can be shown to be sound, and it challenges Kant’s derivation of the Formula of Universal Law by challenging its conclusion: if this objection is sound, then the derivation would fail, because it is supposed to produce a non-empty principle.

⁷ The problem can be stated in a few different ways even here. I will focus on the version that says that any maxim can satisfy the Formula of Universal Law, but Hegel sometimes claims that any maxim can be shown to fail it as well (e.g. PS ¶431). Weaker versions of the objection hold that the Formula of Universal Law can draw some moral distinctions, but not a fully adequate list or system of them (Wood 1990, 154-155), or that the distinctions that the Formula draws cannot be applied in action fully, adequately, or even at all, to any extent (O’Neill 2013, 43ff). These latter issues are not quite Hegel’s, but are related enough to deserve to fall under the same general heading of “empty formalism” problems.

However, it is very difficult to tell how this “empty formalism” objection actually engages with Kant’s picture of how maxims satisfy and conflict with the Formula of Universal Law. Over the previous chapters, we have developed a detailed picture of what maxims consist in, according to Kant, and we have discussed a few examples of how the Formula of Universal Law is supposed to rule out certain maxims. How does Hegel actually object to this account, in order to show that this entire picture is really an empty formalism?

Let’s start with Hegel’s remarks in the *Elements of the Philosophy of Right*, since this text is a key source for the phrase “empty formalism”. It is also one of his latest texts, and it represents much of the tenor of his earlier thinking on these points as well.⁸ Moreover, Hegel addresses Kant more directly and specifically here than in many other places where he raises similar issues. So the passage is worth quoting at length:

However essential it may be to emphasize the pure and unconditional self-determination of the will as the root of duty – for knowledge of the will first gained a firm foundation and point of departure in the philosophy of Kant, through the thought of its infinite autonomy (see §133) – to cling on to a merely moral point of view without making the transition to the concept of ethics reduces this gain to an *empty formalism*, and moral science to an empty rhetoric of *duty for duty’s sake*. From this point of view, no immanent theory of duties is possible. One may indeed bring in material *from outside* and thereby arrive at *particular* duties, but it is impossible to make the transition to the determination of particular duties from the above determination of duty as *absence of contradiction*, as *formal correspondence with itself*, which is no different from the specification of *abstract indeterminacy*; and even if such a particular content for action is taken into consideration, there is no criterion within that principle for deciding whether or not this content is a duty. On the contrary, it is possible to justify any wrong or immoral mode of action by this means. – Kant’s further form – the capacity of an action to be envisaged as a *universal* maxim – does yield a more *concrete* representation of the situation in question, but it does not in itself contain any principle apart from formal identity and that absence of contradiction already referred to. – The fact that *no property* is present is in itself no more contradictory than is the non-existence of this or that individual people, family etc., or the complete *absence of human life*. But if it is already established and presupposed that property and human life should exist and be respected, then it is a contradiction to commit theft or murder; a contradiction must be a contradiction with something, that is, with a content which is already fundamentally present as an established principle. Only to a principle of this kind

⁸ In fact, the crucial section of the *Elements of the Philosophy of Right*—§135—directs us to paragraphs 596-631 of the *Phenomenology of Spirit*, which we will cover in the course of fleshing out the objection in (PR §135).

does an action stand in a relation of agreement or contradiction. But if a duty is to be willed merely as a duty and not because of its content, it is a *formal identity* which necessarily excludes every content and determination. (PR §135)

This passage is remarkably lucid for Hegel, but one could still be forgiven for thinking that there is no argument at all here, or at least no positive argument. We do get a few statements of Hegel's position: it is essential to emphasize "the pure and unconditional self-determination of the will as the root of duty", as Kant does. But we will be left with an "*empty formalism*", unless we transition from this "merely moral point of view" to "the concept of ethics". Hegel continues: no "immanent theory of duties" is possible from the merely moral point of view; "it is possible to justify any wrong or immoral mode of action by this means"; there is no contradiction inherent in the supposition that "*no property* is present"; and so on. Hegel also implicitly provides a few attempts to deflect objections to his position, such as "What if we "bring in material *from outside*" the moral point of view?", or "Doesn't Kant give us a "*concrete* representation" of the moral point of view, so that it isn't just a "specification of *abstract indeterminacy*?" But if we want a positive argument for Hegel's conclusion, we are left with some reconstruction to do.

In that spirit, if we work our way through this passage, we can find a few remarks that might be able to serve as the basis for fleshed out arguments. Here are passes at four such arguments, in the order in which Hegel presents them:

(1) It "is impossible to make the transition to the determination of particular duties from the above determination of duty as *absence of contradiction*, as *formal correspondence with itself*, which is no different from the specification of *abstract indeterminacy*": The Formula of Universal Law tells us to act according to maxims that will not put our wills into irresolvable conflicts with themselves, when we consider them as universal laws. But absence of contradiction within the will and the formal correspondence of the will with itself are really just two kinds of "*abstract indeterminacy*". And just because of what abstract indeterminacy consists in, nothing particular can follow from it. So we cannot determine any particular duties from the idea that morality ultimately only requires internal consistency in our volition or formal correspondence with ourselves.

(2) "The fact that *no property* is present is in itself no more contradictory than is the non-existence of this or that individual people, family etc., or the complete *absence of human life*": Denying property is not intrinsically incoherent, nor is

denying the existence of anyone and everyone. In fact, whatever would result from a maxim, considered as a universal law, is internally consistent “in itself”. So, by extrapolating from cases, no action, act of will, or will in general need ever be incoherent. Hence, we can will any maxim we like as a universal law, without ever contradicting ourselves.

(3) A “contradiction must be a contradiction with something, that is, with a content which is already fundamentally present as an established principle”: By the very nature of contradictions, something can only contradict an established principle. So it is impossible for any will or act of volition to conflict with itself alone, only with an independently established principle. Hence, no contradiction is possible under the Formula of Universal Law. Therefore, we can will any maxim we like as a universal law, without ever running into a contradiction.

(4) If “a duty is to be willed merely as a duty and not because of its content, it is a *formal identity* which necessarily excludes every content and determination”: Duty is a form or nature that many actions and kinds of actions have in common. So willing an action just because it is your duty is to choose your action only on the basis of its form, not on the basis of its specific content. You choose not to lie, for example, not because of the specific nature of lying, but just because it is a duty not to lie. However, to select principles based only on their form must be to choose them regardless of their content or their specific determination: otherwise you would be choosing by something other than their form alone. But if content is irrelevant to the selection of principles, then you can choose any content whatsoever, without choosing something other than duty for its own sake.

To begin with, notice that none of these lines of argument address the specific details of Kant’s account of how the Formula of Universal Law is to be applied. For example, in no instance here does Hegel actually take a clearly immoral maxim and try to show, after Kant’s fashion, how to make it compatible with the Formula of Universal Law.⁹ The closest he comes to that is in (2), in the assertion that there does not have to be any internal contradiction in denying property. He does not, however, do anything to *show* that there can be no contradiction here.

Instead, Hegel is giving us considerations that are far more general and abstract. Let’s work through them in turn, saving the concrete cases in (2) for last, however. Consider argument (1) first. Here, the argument is that the Formula of Universal Law only requires a certain kind of “*absence of contradiction*” in our volition, or that it only requires our wills to

⁹ In (PS ¶¶430-431), Hegel does try to show that certain concepts of property and non-property both contradict themselves, but not in Kant’s fashion, as we’ll show a little later.

exhibit a certain kind of “*formal correspondence with itself*”. But that is supposed to make the requirements of the Formula of Universal Law into an “abstract indeterminacy”, which is to be incompatible with the idea that particular, concrete duties can be derived from this Formula. Is this a sound line of argument against the Formula of Universal Law?

To begin with, it is not at all clear that this argument ever actually engages Kant’s moral theory in its own terms. The Formula of Universal Law surely does require “absence of contradiction”, at least in some sense. But why should this mean that there is anything indeterminate about the Formula of Universal Law? Especially when this Formula does not ask about whether our wills contradict themselves only in their own subjective terms, or even just in general, but rather when our maxims are considered as universal laws? That appears to be a determinate version of the question about non-contradiction, and one which is posed in terms of a standard that is supposed to be objective, and not merely down to the idiosyncrasies of any particular will.

Hegel does agree that “Kant’s further form – the capacity of an action to be envisaged as a *universal* maxim – does yield a more *concrete* representation of the situation in question” (PR §135). But he still thinks that “it does not in itself contain any principle apart from formal identity and that absence of contradiction already referred to.” (PR §135) So his view, evidently, is that the Formula of Universal Law is still an “abstract indeterminacy”, despite how it is “a more *concrete* representation of the situation in question”. He does not explain why, however.

So perhaps the “abstract indeterminacy” is supposed to come from the “formal correspondence with itself” that the Formula of Universal Law is supposed to require? Again, the Formula of Universal Law certainly does require some kind of “formal correspondence with itself”, just because it requires our volition to be internally coherent when we consider our maxims as universal laws. That is, after all, a kind of formal coherence or “correspondence with itself”.

Only Hegel does not explain what “formal correspondence with itself” is supposed to mean in this passage, either. Elsewhere, though, he appears to mean something like “mere self-identity”. For example, in making the “empty formalism” objection in the *Phenomenology of Spirit*, Hegel writes that “the essence of ethics consists just in law being identical with itself and through this self-identity, i.e. through having its ground in itself, it is unconditioned” (PS ¶430). Our passage from the *Elements of the Philosophy of Right* had “formal correspondence with itself” as the essence of the ethical theory in question, too, so it sounds as if “formal self-correspondence” and “being identical with itself” mean the same thing, for Hegel.

Moreover, Hegel also adds that “It would be strange too, if tautology, the maxim of contradiction, which is admitted to be only a formal criterion for the cognition of theoretical truth, i.e. something which is quite indifferent to truth and falsehood, were supposed to be more than this for the cognition of practical truth.” (PS ¶431) I take it that we would call Hegel’s “maxim of contradiction” the “maxim of non-contradiction” instead. After all, contradictions are surely not “a formal criterion for the cognition” of any kind of truth at all, while our principle of non-contradiction surely meets this description. So Hegel’s equating of “tautology” and “the maxim of contradiction” suggests that the “formal correspondence with itself” involved in the latter is supposed to be the same as the self-identity in the former.

But self-identity in what sense? In Kant’s own “general” logic, and in standard Aristotelian logics generally, it is trivial to say that everything is self-identical. So if Hegel were right in alleging that Kant proposes mere self-identity as “the essence of ethics”, then Kant’s ethics really would be an “empty formalism”, because it would follow trivially that every principle is identical with itself, no matter how we consider it. So if that self-identity is all that it takes to make a principle morally respectable, then any principle at all would pass Kant’s test, no matter how depraved.

Put this baldly, though, Hegel would be obviously misrepresenting Kant’s actual position. Kant’s tests are aimed at determining whether our wills conflict with themselves when

we consider our maxims as universal laws, and Kant gives no indication that this calls anything's self-identity into question. Indeed, how could it, since for him everything is identical with itself, no matter how we consider it?

And we must admit that it would be truly bizarre to accuse Kant of proposing bare, logical self-identity as the fundamental principle of morality, just as it would be strange to accuse him of being in error as to whether or not a maxim or a will could be identical with itself, in that trivial sense. The only way such an allegation would make any sense at all would be if you adopted a logic wherein internal inconsistency precludes self-identity. In such a logic, to show that something contradicts itself would be to require it to transform into something else, violating its self-identity. And then requiring internal consistency would also require self-identity. Hegel arguably has such a logic, but Kant does not. The essential program of the *Phenomenology of Spirit* is to illustrate how things turn into their dialectical opposites—masters into slaves, slaves into masters (PS ¶178-196)—thereby “negating” themselves, as Hegel puts it (PS ¶79), and forcing transformations into higher configurations or forms of themselves.

Now, Hegel's notion of “*determinate*” negation is intentionally and self-consciously quite different from the negation operator in Aristotelian logic, both in its function and in its range of application (PS ¶79). As is his notion of “contradiction”, as will come out a little more fully soon. Accordingly, these notions might be better thought of as metaphysical concepts, rather than as logic ones, and perhaps even from Hegel's own point of view.

Still, in Hegel's sense, too, self-identity is empty. In his dialectical logic, “self-identity” means remaining with a single side of a dialectic, without opposing anything to it. Without a dialectical opposition, however, nothing at all follows, in Hegel's view. Hence, the Formula of Universal Law would still be empty, if it requires only “formal correspondence with itself”, construed to mean only self-identity in the sense of Hegel's logic. For it would lack the dialectical opposition required for it to have any consequences at all.

But, again, this is far from Kant's picture of how general logic works, because it is far from the Aristotelian logic that Kant took to be established, for multiple compelling reasons, and which he took to be established in its own terms and in its own right, independently of its role in the project of Kant's critical philosophy. So if Hegel is objecting to the Formula of Universal Law because he thinks that it only requires a "formal correspondence with itself" in his dialectical sense, then he is not really objecting to the Formula of Universal Law *per se*, but rather to the whole logical framework that Kant is working in—he is disagreeing about the very meaning of "formal correspondence with itself" and "self-identity", as these terms occur in Aristotelian logic, and not just in ethical and moral theory. For Hegel's logic says that nothing can follow from our volition, considered in isolation from a dialectical opposite, while Kant's logic emphatically does not require that.

So no matter what case might be made for Hegel's dialectic, we can ignore it as not germane to our specific issue. For now, we should just conclude that whether we take self-identity in Kant's sense or in Hegel's, the accusation that the Formula of Universal Law requires only self-identity misrepresents Kant's actual position, either about what the Formula of Universal Law says, in the former case, or about logic itself, in the latter case. And an objection to Kant's moral theory is on rather awkward footing, to say the very least, if it requires us to resolve issues in logic in general, and particularly if it requires us to give up on Aristotelian logic.

The other remarks on "formal correspondence with itself" just quoted from the *Phenomenology of Spirit* exhibit a similar divergence in broader framework (§430). There, Hegel equates self-identity with "having its ground in itself". Hegel is right that, for Kant, a maxim's moral status is grounded in its form, so that a maxim's moral status counts as self-grounded, too. We can even say that a maxim "has its ground in itself", as long as that is supposed to mean that the ground of its validity or invalidity lies in itself. In the last chapter, I made this point by saying that the moral status of a maxim is "intrinsic" to it.

But even so, to say that a maxim's validity or moral status is grounded in its own form is not the same thing as saying that the Formula of Universal Law requires only self-identity. It requires our maxims to have a certain form, so that their corresponding validity or moral status counts as intrinsic or self-grounded, too. But that isn't the same as requiring our maxims to be identical with themselves, still less is it to require only that. It does not even say that the *maxim* is grounded in itself, not *per se*, only that its validity or moral status is grounded in itself. A maxim's identity and self-identity is to be the ground of its validity and moral status, but the Formula of Universal Law calls on us to choose maxims because they have a certain identity, not because they are self-identical. Choosing maxims that are self-identical would be to choose them because they have *an* identity, not because they have the required identity.

So, here again, Hegel seems to be misunderstanding Kant. In fact, if the *Phenomenology of Spirit* intends a remark like "the essence of ethics consists just in law being identical with itself and through this self-identity, i.e. through having its ground in itself, it is unconditioned" (PS ¶430) to be a criticism of Kant, then it seems to get Kant exactly backwards. Kant does want the validity of moral maxims to be intrinsic or self-grounding, but Hegel seems to say that Kant either tries to get this self-grounding, or that he can only get this self-grounding, by basing ethics in self-identity. Again, that follows in Hegel's dialectical logic. There, something can be immediately or intrinsically self-grounding only if we leave it totally in isolation, without a dialectical opposite. So that kind of self-grounding can only rely on mere self-identity. That really would be totally empty and unilluminating, however. Genuine self-grounding, in Hegel's logic, requires the mediation of the dialectic: a thing has to return back to itself, to uncover its real or true form, through the long process of dialectical opposition and resolution, until the progression of this dialectic comes back around to where it began, or, perhaps better, until this process finally arrives at the true expression or realization of its starting point.

So, once more, Hegel's criticism either misrepresents Kant, or presupposes a logic that Kant does not accept, and which we have not yet seen any reason for him to accept. We get the

same result if we look at Hegel's idea that the Formula of Universal Law only requires "absence of contradiction" or "formal correspondence with itself" through the lens of his equation of "tautology" and "the maxim of [non-]contradiction". As Hegel puts it, "It would be strange too, if tautology, the maxim of contradiction, which is admitted to be only a formal criterion for the cognition of theoretical truth, i.e. something which is quite indifferent to truth and falsehood, were supposed to be more than this for the cognition of practical truth." (PS ¶431) Hegel does get Kant right here, at least to the extent that non-contradiction is only a "negative" criterion of truth, according to Kant (CR 279, A150/B189-190). But for Kant this only means that non-contradiction is "quite indifferent" to *contingent* "truth and falsehood", not to any truth and falsehood whatsoever. In Kant's mind, internal consistency means that a proposition could be true, if only so far as general logic is concerned, while internal inconsistency means that it must be false, no matter how it is considered.

In contrast to this, Hegel appears to think that non-contradiction cannot show that anything is false: the criterion of non-contradiction is "is quite indifferent to truth and falsehood", as he puts it. It is hard to know what, exactly, Hegel means by this, especially since he himself accepts his own version of the principle of non-contradiction, as logically fundamental, even. Hegel certainly adopts a very broad notion of contradiction, broader than in Aristotelian logic, because the dialectical oppositions that we have been discussing count as contradictions in his sense, even when the terms of the opposition are not themselves anything like concepts or propositions. Aristotelian logic would not recognize the opposition between slaves and masters as itself a contradiction, for example, while Hegel's dialectic treats it as such. So it might be better to treat these oppositions as involving or realizing contradictions. In any case, though, in Hegel's broad (and arguably over-broad) sense, the entire dialectic is supposed to work by generating contradictions, in his sense, because of the oppositions that it both involves and produces. Hegel's principle of non-contradiction then requires these oppositions to

be resolved, in the way already illustrated. (Which, it is worth emphasizing, is decidedly not the way in which Aristotelian logic treats contradictions.)

So Hegel's broad, non-Aristotelian principle of non-contradiction is "active" at every moment of the dialectic, because it is always violated throughout, and its demands drive the dialectic forward, until non-contradiction is finally satisfied only at the end of this process. It is then a bit hard to see how the principle of non-contradiction could be "quite indifferent to truth and falsehood", even from Hegel's own point of view. Hegel says that the dialectic is ultimately to present the truth in its true form (PS ¶15)—so putting aside whatever 'truth' and 'true' are to mean here, how, then, can his principle of non-contradiction be "quite indifferent to truth and falsehood"?

Hegel says that "It would be strange too, if tautology, the maxim of contradiction, which is admitted to be only a formal criterion for the cognition of theoretical truth, i.e. something which is quite indifferent to truth and falsehood, were supposed to be more than this for the cognition of practical truth." To square this remark with the role that non-contradiction is supposed to have in the dialectic, Hegel's emphasis must be on the idea that non-contradiction is "admitted to be only a formal criterion". When the dialectic is realized in the world in the process and progress of human history, Hegel thinks that it is not only a "formal criterion", but a "real criterion", too. As such, Hegelian non-contradiction is supposed to be able to serve to drive towards the realization of ultimate, absolute truth.

Hegel would object, though, that the Formula of Universal Law treats non-contradiction as only a "formal criterion", because it applies only to our volition, only "internally", in our decisions about how to act. So in Kant's use, Hegel would think that non-contradiction really is "quite indifferent to truth and falsehood"—to real truth and falsehood, whatever that might amount to, in Hegel's terms. The Formula of Universal Law then turns out to be an "empty formalism", as alleged, since it cannot tell us what our "true", "real" duties are.

Here again, though, we see Hegel understanding and evaluating Kant in the terms of Hegel's dialectical logic, and not engaging with Kant's logic in its own terms, to say nothing of his moral theory. We do not even need to understand what, exactly, Hegel means by a "formal criterion" to understand that. All we need to do is notice, once again, that Hegel's logic is not Kant's, because Kant does not require non-contradiction to be realized in the world outside of volition in order for it to be a criterion for what is true and false. Nor is it at all clear why Kant should require that, in the absence of an argument against Aristotelian logic.

So, again, Hegel does not seem to be interpreting the Formula of Universal Law fairly when he says that the Formula only requires "absence of contradiction" in the will or the "formal correspondence [of the will] with itself". At least not when "fair" means "as Kant would have understood the Formula from his own point of view, and not from Hegel's". Hegel is just engaged in a very different project: he is trying to reveal something about what is going on under and around Kant rather than in Kant, so to speak, or to show what is going on in Kant *properly* understood, not understood in Kant's own terms. That is totally fair by Kant's own lights and practices, we should add (CR 395-396, A313-314/B370). But much more work would need to be done to show why we should prefer Hegel's project and system to Kant's, particularly when Hegel wants us to adopt an entirely different system of logic, which has not been able to establish itself in its own terms, just as a logic, apart from Hegel's broader philosophical system.

In any case, though, we have been trying to understand Hegel as arguing that the Formula of Universal Law is an "abstract indeterminacy", from which nothing particular can follow, just by definition. If Hegel is right about that, then no particular duties could ever follow from Formula of Universal Law, just by virtue of the kind of principle that the Formula is supposed to be. Hegel's argument, as I have tried to reconstruct it, tried to show that the Formula of Universal Law is an abstract indeterminacy of this kind, because this Formula ultimately only requires "absence of contradiction" within the will, or for the will to exhibit "formal correspondence with itself". But we have just seen that this argument can only be sound

within Hegel's broader framework, since otherwise it misunderstands Kant's actual position. But is the Formula of Universal Law an "abstract indeterminacy" anyway, even if Hegel's argument has not yet shown this?

Again, it is difficult to know what "abstract indeterminacy" is supposed to mean, for Hegel. The Formula of Universal Law is certainly abstract in a number of ways: it is a conceptual representation, and so abstracted from all intuitions, as we saw back in Chapter 1.1. This Formula is also a pure representation, so that it abstracts from all empirical or sensible content (G 22, 4:410). And so the Formula of Universal Law is indeterminate in corresponding ways as well. But that does not mean that the Formula of Universal Law is *absolutely* indeterminate, or indeterminate in a way such that nothing particular can follow from it or conflict with it. Far from it: Kant is consistently of the view that pure concepts do have determinate, particular consequences. In fact, this is one of the fundamental claims of his whole critical project, since he could not explain the possibility of synthetic *a priori* theoretical or practical cognition if he could not show how pure concepts could have determinate consequences, even and especially in the empirical world. In earlier chapters, we looked at some specific moments where Kant tries to prove precisely this: the schematism of the categories and the topic of pure practical judgment.

For Hegel, however, to say that something is pure often appears to mean that it is totally indeterminate or devoid of specific content, just perforce. Major parts of his discussion of morality, for example, include remarks like "the fact is that pure duty consists in the empty abstraction of pure thought, and has its reality and its content only in a specific reality" (PS ¶637). Hegel also says that "This *pure* conviction is, as such, as empty as pure *duty*, is pure in the sense that there is nothing in it, no specific content that is a duty" (PS ¶643), and that "whatever content it be, it contains the *blemish of determinateness* from which pure knowing is free, determinateness which pure knowing can disdainfully reject, or equally can accept. Every content, because it is determinate, stands on the same level as any other" (PS ¶645). See also (PS ¶606), (PS ¶611), (PS ¶630), and (PS ¶659).

So if purity implies emptiness, for Hegel, then he appears to be arguing that the Formula of Universal Law is practically indeterminate just because it is to be pure or abstract. Whatever his reasons for thinking that, it is a criticism of the Formula of Universal Law only by way of a far broader criticism of Kant's entire epistemology, and even of his entire critical project.

So it looks as though our first reconstructions of Hegel's "empty formalism" objection give us nothing that we can isolate from Hegel's broader philosophical project. If we reject his logic and epistemology, as Kant does, then he has not yet given us any reason to think that the Formula of Universal Law is empty. Thankfully, we can cover the third and fourth reconstructions more briefly, and then cycle back to the second.

Our third possible line of thought started from Hegel's point that "a contradiction must be a contradiction with something, that is, with a content which is already fundamentally present as an established principle." (PR §135) And we read this simply as a definition of what contradictions are supposed to consist in, which definition is supposed to show that acts of will can never come into conflict with themselves, but only "with a content which is already fundamentally present as an established principle." If that is right, then it looks as though we can will anything we want as a universal law, without ever being at risk of contradicting ourselves. So the Formula of Universal Law would be empty, as alleged.

However, it is quite hard to know what to make of this remark, read as an attempt to define contradictions. What is "established" supposed to mean, for example, and why should a contradiction have to require an "established" principle? The concept of contradiction in Aristotelian logic requires nothing this stringent.

Moreover, it is also hard to see why the contradictions within the will that Kant envisages cannot satisfy Hegel's definition here. Hegel insists that a contradiction must be a contradiction "with something", and, more specifically, with "a content which is already fundamentally present as an established principle." Under my reading of Kant's practical principles, though, it sounds as though we should be able to take an individual's maxim as given, as an "established

principle”, with its three general elements—its ground, action, and end—all “fundamentally present” and “established” within in. These elements are then separate “principles” also: principles for how we are to work out what follows from the maxim considered as a universal law. And nothing prevents one component of the maxim from coming into conflict with another, as we saw in the previous chapter, when the end invoked in the maxim of the lying promise came into conflict with its means, or when the judgments under the maxim of indifference came into conflict with each other. Despite how these elements or moments of the maxim are related to one another within the maxim, they remain “separate” elements, “external” to one another, even within the context of their overarching maxim. So why are they not also apt for Hegel’s definition of contradictions, even though they remain unified under the overarching maxim?

To rule out an analysis like that one, “fundamentally present as an established principle” would have to mean something on the order of “really established”, in something like the way in which the ethical principles of Greek *Sittlichkeit* are supposed to be realized and established in Greek social relations, social structures, ethical ideologies, and behavior. Read this way, Hegel’s definition would allege that Kant cannot show that there are any *real* contradictions within the will, because a real contradiction has to be a conflict with a “really” established principle of this kind. An individual’s volition alone cannot establish such a principle, as Kant himself would admit. So conflicts within the will alone would have to be of no real moral significance, unless they involve a conflict with a “really established” principle, too.

However, for one thing, the Formula of Universal Law appears to ask us to consider our maxims as “really established” principles, in the relevant sense here. After all, as I said earlier, to consider our maxims as a universal laws is to consider them as the basis for universal social practices. So any conflict that arises within the scope of such a law would still seem to count as a conflict within a “really established” principle, too.

Thus, it remains rather hard to see how Kant’s picture violates this definition of contradictions. Even if it does, though, why should we accept Hegel’s definition, once we

understand it this way? Hegel would have to rely either on his background views of logic, or on his background views of what immoral action consists in. Either way, though, his objections to Kant devolve back to his broader system.

The fourth line of thought, then, held that choosing a principle only because of its form “necessarily excludes every content and determination” (PS §135) just by definition. For if you take the content into account, then you choose on the basis of something besides *only* the form. But then content is irrelevant to a choice based only on form, and so you can choose any content at all on that basis. Thus, the Formula of Universal Law would be “an empty formalism”, that determines no content.

This argument has a superficial kind of plausibility, and even cleverness, to it, but it clearly forgets that a form has to be the form *of something*. In Kant’s picture, this will be the form of a specific maxim, which under our interpretation consists in the way in which its three components relate to each other. So to choose a maxim on the basis of its form is really inseparable from choosing it on the basis of its content. To include the content into your evaluation of the form is not to choose solely on the basis of that content, nor is it to make the content alone the determining ground of your will. It is, rather, to choose on the basis of the form *of that content*. So if Hegel’s argument is to be persuasive at all, it can only be because he is relying on other aspects of his system again: for example, on an interpretation of form wherein form is intrinsically indeterminate as to content, or wherein form is always essentially empty, at least until it comes to be the form of a concrete content in the world, rather than of a thought or volition. But Hegel just begs the question against Kant, if he asserts that the form of volition is essentially or intrinsically empty, before it gets embodied, realized, or projected into the world. Or, at minimum, Hegel transplants the disagreement between their ethical theories to their accounts of form and content and the relation between them.

So far, then, Hegel’s empty formalism objection neither directly engages Kant’s account, nor relies on any premises that Kant would accept, nor gives anything in the way of positive

reasons why Kant should accept those other premises, independently of Hegel's entire system. Hegel does develop a couple of examples in order to substantiate his claim that the Formula of Universal Law is empty, though, so let's now consider those.

In our passage from *Elements of the Philosophy of Right*, Hegel claims that "The fact that *no property* is present is in itself no more contradictory than is the non-existence of this or that individual people, family etc., or the complete *absence of human life*." (PR §135) His point being, I take it, that we can suppose or imagine any of these states of affairs without having to contradict ourselves. Instead, these scenarios can only involve contradictions if we assume property or human life as pre-established values or principles for these states of affairs or for actions like "theft or murder" to conflict with. No actual configuration of the world is, or even can be, a contradiction "in itself". So we have to posit pre-established values or principles, otherwise there would be nothing at all for actions, volitions, or states of affairs to conflict with or to contradict. So we find, again, that there is no such thing as a will that must conflict with itself just in its own terms, and the Formula of Universal Law is "an empty formalism".

Here again, though, it is rather hard to see right away how this point is supposed to engage Kant's actual views. We can concede that Hegel is surely right to point out that we can assume or imagine a world without property or without humanity, but that proves nothing by itself, according to Kant. For Kant, the question is not about whether we can coherently *imagine* a world without property. Instead, Kant's question is rather about whether we can coherently consider such a world to be the product of a universal law for our wills. Ultimately, Kant wants to know whether we can *will* such a world without contradicting ourselves, and, if we can, in what circumstances, or under which maxims.

For example, the previous section described the main Kantian line on theft, which was something like the following: I can't steal whatever I want just because I want it, because if everyone did that, the social institution of property would be destroyed. Thus, it would be impossible for me to make anything my property through theft, and my maxim cannot be

conceived as a universal law, still less can it be willed as a universal law. Theft itself is only possible against the backdrop of private property, so a universal law of theft is incoherent, even though we can imagine a world without private property, and even though we can imagine making such a world through our wills by other means.

So we need to do a little more work in order to see how these examples might tell against Kant's views, at least as we have been interpreting them. We discussed this a little bit already in the previous section. But Hegel also tries to explain how social arrangements involving either property or its absence are both self-contradictory and internally coherent, depending on how we consider them.

“Suppose the question is: Ought it to be an absolute law that there should be property?” (PS ¶430), Hegel asks, and then tries to work out whether either this law or its opposite would contradict themselves. If we consider either the institution of property or its absence as an “*isolated*” or a “*simple* determinateness” (PS ¶430/¶431), then Hegel claims that social arrangements can involve either without contradicting themselves. For Hegel, to consider these institutions this way is just to prescind from whatever internal structure that either institution might have, rather than to allow either of them to be “resolved into its moments” (PS ¶431) or basic components. So it follows trivially for him that neither institution can contradict themselves when considered in this way, because here they both lack the internal structure necessary for such a conflict to arise.

Instead, when we do “resolve them into their moments”, we should see that both institutions conflict with themselves, at least when we consider them as the requirement of an absolute law. For an absolute law forbidding property, Hegel argues that

the idea of a thing being arbitrarily allotted to the first self-conscious individual who comes along and needs it, does not accord with itself. In a society based on a common ownership of goods, in which provision would be made in accordance with a universal fixed rule, either each receives as much as he *needs*—in which case there is a contradiction between this inequality and the essential nature of that consciousness whose principle is the equality of individuals—or, in accordance with that principle, goods will be *equally* distributed, and in this case

the share is not related to the need, although such a relationship alone constitutes the very notion of 'sharing'. (PS ¶430)

Here, Hegel presents a dilemma: the idea behind an absolute law of “non-property” is that goods should be distributed according to need—“arbitrarily allotted to the first self-conscious individual who comes along and needs it”. Here, the need is the important part: goods are to be distributed to each individual, in the order in which they “come along” and genuinely need the thing. Ideally, this “arbitrary allotment” to each individual as they happen to come along will eventually lead to everyone getting what they genuinely need. The background idea here is supposed to be “a common ownership of goods” based on “the equality of individuals”. Each individual has an equal share in the ownership of everything, which is then to be expressed or realized in the distribution of those goods according to need.

But Hegel thinks that these two ideas—equal ownership and distribution according to need—are in tension with one another. If we make a universal law—“a universal fixed rule”—requiring that “each receives as much as he *needs*”, then these goods will not be distributed equally. Needs differ, and so the distribution of goods to satisfy those needs must also be unequal. But this inequality contradicts “the equality of individuals”, which was supposed to be the basis for giving everyone what they need. Instead, “the equality of individuals” really requires that “goods will be *equally* distributed”. An equal distribution, though, would require giving up on the idea of distribution according to need.

As stated, this argument has some obvious holes that would need to be filled in in order to make it philosophically convincing. Mainly: why isn't a distribution according to need an equal distribution in the relevant sense? You might think that equality as individuals yields no claims on a good, at least not by itself. Rather, our equality as individuals gives our individual needs equal standing in the distribution of goods. Thus, distributing goods according to need would be exactly what the equality of individuals requires. So why wouldn't that distribution be coherent, even as required by an “absolute law”?

But even if Hegel's argument here can be made to work, notice that he is still giving us the wrong kind of analysis, from the point of view of Kant's moral theory. To show how the idea of non-property conflicts with itself, considered as an "absolute law", Hegel asks us to consider "a society based on a common ownership of goods, in which provision would be made in accordance with a universal fixed rule". So he is asking us to consider "non-property" as a kind of social distribution of goods: common ownership, distributed according to an established universal rule.

Under the Formula of Universal Law, however, to consider a maxim as a universal law is not to consider a "society based on" that law, except in the sense that we are to consider a society wherein everyone individually always follows that maxim. As I said before, the Formula of Universal Law does ask us to consider our maxims as the basis for social practices. But those social practices are to emerge from how every individual will always adhere to the maxim in question, not from any official social institution or coordination. According to Kant, the Doctrine of Right in the *Metaphysics of Morals* is to determine which social institutions we are to have. That is not the job of the fundamental moral theory immediately, directly, or by itself.

So the Formula of Universal Law does not require us to imagine any social or institutional coordination in implementing our maxims as universal laws. In fact, we ought not to, since its question is about what we can will as individuals. Yet Hegel's whole argument here presupposes that we are talking about a socially coordinated distribution of goods, albeit one with an element of arbitrariness to it. As such, Hegel's argument simply doesn't speak to any questions about individual morality just as such, about what I can will as an individual, or about what follows when we consider my individual acts of will as universal laws. So even if the idea of non-property would conflict with itself, when we consider it as the basis for a socially coordinated distribution of goods, it is far from clear what that would tell us about what maxims we can and cannot will as universal laws.

In order to translate Hegel's argument here into the proper Kantian frame of analysis, we would have to stipulate a maxim, perhaps something like: I will distribute goods in my control to others according to their need, in order to satisfy that need, because of their equal moral status. Arguably, this maxim is not a maxim of "non-property", since it appears to presuppose that I have property to distribute: "goods in my control" seem pretty close to property. But there is no conflict in this maxim in any case, because this agent is working under a certain interpretation of equal moral status, which we do not have to suppose is the one that Hegel is utilizing in order to make his objection.

We can make similar points in response to Hegel's argument that the notion of property also contradicts itself, once it is "resolved into its moments." (PS ¶431) Here is his argument:

The single thing that is my property is held as such to be something universal, solidly established, and permanent; but this contradicts its nature, which consists in its being used and in *vanishing*. At the same time, it is held to be *mine*, something which everyone else acknowledges, and lets alone. The fact, however, that I am acknowledged implies rather my equality, my identity, with everyone, and that is the opposite of exclusiveness. What I possess is a Thing, i.e. something which is for others in general and is only for *me* in a quite general, undefined way; that *I* possess it, contradicts its universal thinghood. Consequently, property is just as much an all-round contradiction as non-property (PS ¶431)

In short, this is a metaphysics of property, not an analysis of a maxim under Kant's Formula of Universal Law. So even if Hegel's analysis here is correct, in order to make it tell against Kant in any way, we would need a story explaining how a specific maxim is committed to these highly specific views about what property involves—what its "moments" are. In fact, Hegel gives the entire game away, so to speak, in stating his own conclusion: "Consequently, property is just as much an all-round contradiction as non-property; each contains within it these two opposed, self-contradictory moments of individuality and universality." (PS ¶431) In reducing the purported contradiction to "moments of individuality and universality", Hegel is importing the basic categories of his dialectical logic, which again shifts the disagreement with Kant away from the Formula of Universal Law as such, and to their broader philosophical frameworks.

Hegel makes a few more points in connection with the “empty formalism” objection, which we can deal with fairly briefly. Specifically, he claims that any action can be performed from the sincere conviction that it is a moral duty, so that “pure duty” is “utterly indifferent to any content and tolerates any content.” (PS ¶644) Moreover, and relatedly, he holds that any action can be properly described in ways that make it both conform to and conflict with the Formula of Universal Law. For example, we may suppose that one person

perceives that he fulfills this duty in this particular case. Others, perhaps, hold this specific way of behaving to be humbug; *they* hold to other aspects of the concrete case, *he*, however, holds firmly to this aspect, because he is conscious of the increase of property as a pure duty. Thus, what others call violence and wrongdoing, is the fulfillment of the individual’s duty to maintain his independence in face of others; what they call cowardice, is the duty of supporting life and the possibility of being useful to others; but what they call courage violates both duties. (PS ¶644)

And, last, Hegel argues that general ethical principles can’t be made determinate enough to guide action without destroying their original meaning. For example,

‘Everyone ought to speak the truth.’ In this duty as expressed unconditionally, the condition will at once be admitted: *if* he knows the truth. The commandment, then, will now run: everyone ought to speak the truth at all times, according to his knowledge and conviction...But with this correction, what the proposition wanted to enunciate as universally necessary and intrinsically valid, has really turned round into something completely contingent. For speaking the truth is made contingent on whether I can know it, and can convince myself of it; and the proposition says nothing more than that a confused muddle of truth and falsehood ought to be spoken just as anyone happens to know, mean, and understand it. (PS ¶424)

Hegel thinks that each of these cases yields up the emptiness of the Formula of Universal Law, but we can answer these objections pretty straightforwardly. To begin with, perhaps Hegel is right that anyone can perform any action from the sincere conviction that it is a moral duty. (Perhaps he isn’t, also, but we can grant the point, for the sake of argument.) But this is actually no argument for the emptiness of the Formula of Universal Law. According to Kant, the Formula of Universal Law is a canon for determining when actions really are and are not moral duties. As such, this Formula is simultaneously a test for when our convictions about our moral duties are accurate and inaccurate. So which convictions are volitionally or psychologically possible for us

provides no independent argument for the emptiness of the Formula of Universal Law whatsoever. It does not even matter if we broaden the discussion here to include moral worth. I take it that the standard interpretation of Kant's views on moral worth holds that actions that are contrary to duty cannot have moral worth. I tend to support that reading, though it is controversial. (See (Kerstein 2002, Ch. 6), for example.) But even if we grant that actions that are contrary to duty can sometimes have moral worth, when performed from the sincere conviction that they are duties, that does not make the Formula of Universal Law empty. For we have still not conceded that the sincere conviction of duty is sufficient to make your action *conform to duty*.

Thus, here again, Hegel's point would have to be supplemented with other considerations, otherwise it does not tell against the Formula of Universal Law. We can treat Hegel's other two points together. For Kant, the proper locus of moral analysis is the agent's maxim. This maxim then provides us with the morally canonical or privileged description of their action. We may never be able to determine an agent's maxim to any very high degree of certainty, but their maxim remains the canon for describing their action. So even if other descriptions are equally accurate as a matter of logic or of semantics, and even if these other descriptions yield different moral evaluations, this would have nothing at all to do with the proper moral analysis of the action in question. Perhaps we will feel that prioritizing the agent's maxim in this way is problematic, or even unacceptable, but unless we have some reason to reject this priority, the fact that other action descriptions are possible has no bearing at all on whether the Formula of Universal Law is empty with respect to the specific maxim in question.

Likewise, we need to distinguish between an agent's maxim and their formulation of their maxim—between their actual representation of their ground, action, and end and how they formulate that representation in language or in thought. Kant is clear that an agent can be mistaken about their own maxims, especially when it comes to their grounds (e.g. G 19-20, 4:407). Still, Hegel isn't really speaking to the point, when he tries to illustrate how trying to

make a formulation of a maxim more determinate leads to a conflict between what the formulation says and what it means: “It *said*: everyone ought to speak the truth; but it *meant*: he ought to speak it according to his knowledge and conviction” (PS ¶424). This isn’t quite to the point, since what the formulation “means” is what is relevant for moral analysis, in Kant’s picture. And only what it “means” insofar as this “meaning” gets expressed in the agent’s actions. So infelicities or tensions between an agent’s maxim and its formulation are surely important to suss out, if we are to evaluate the maxim properly. But even if such tensions are inevitable and persistent, they do not show that the Formula of Universal Law is unworkable or empty, either in practice or in principle. For moral evaluation will only rarely, if ever, require a perfect formulation of the maxim. Not every tension between what is said and what is meant is morally relevant at all, to say nothing of morally relevant in every case that might come up in practice.

On this score, the previous chapter also pointed out that Kant does not think that there can ever be a point where conceptual representations are determinate enough to specify actions immediately, without an intervening act of judgment. So an agent always needs to make a judgment in order to act according to a maxim, no matter how general or specific their maxim is. Moreover, we also saw that concepts can be specified and precisified indefinitely, and any stage of this further specification will always risk either creating or revealing tensions with the underlying principle. For Kant, the problems of precisification and application are simply endemic structural features of our agency, because of the nature of the cognitive apparatus that he thinks that we bring to our actions. But those problems do not defeat moral analysis or render us unable to apply our principles in the world, precisely because we also have faculties for navigating these problems. (Judgment and understanding, primarily.) We do have to know what the relevant concepts mean to their agent, at least in general. But we do not need to fully specify how those concepts are to be applied in action and judgment, either to act according to a principle, or to evaluate that principle morally.

So Hegel's analysis here cannot really get off of the ground, both because we must focus only on what the formulation of the principle "means", and because we don't need to worry about fully specifying how the principle is to be applied in action. We only need a sufficient formulation. That may be hard enough to come by, to be sure. But, nevertheless, we don't need to resolve every possible unclarity or indeterminacy in advance of either moral analysis, or in advance of acting on a principle. On Kant's view, all action and practical evaluation would be totally impossible, if we did.

6.3 Allen Wood: An Equivocation in the Derivation of the Formula of Universal Law?

So we have now examined many different ways of reconstructing Hegel's argument for his "empty formalism" objection, and they all turned out to really target other aspects of Kant's critical philosophy, where they succeed in engaging Kant's thought at all. In turn, then, none of these lines of reasoning can be separated from Hegel's broader philosophical project or system, and we do not get a specific, isolable objection to the Formula of Universal Law just as such, and hence no objection to its derivation, either. So we can pass by these objections, as involving too many issues outside of the scope of our study here.

Allen Wood, however, does give us one way of developing a criticism specific to the Formula of Universal Law out of Hegel's broader lines of criticism. Wood agrees that the Formula of Universal Law is not empty, *contra* Hegel's objection (Wood 1990, 161).¹⁰ But he suggests that we can understand Hegel as challenging the derivation of the Formula of Universal Law directly, and not just through the adequacy of its conclusion. "The *Philosophy of Right* criticizes Kant for remaining at the moral standpoint without going on to the standpoint of ethical life" (Wood 162), Wood avers, confirming points we made in the previous sections. So "Hegel's deeper and more interesting thesis is that Kant is not *entitled* to employ a principle with any content beyond the criterion of noncontradictoriness" (Wood 162). Meaning, specifically, that Kant is not entitled to a principle that would require us to be able to will our

¹⁰ All parenthetical citations to Wood in this section will be to this text.

maxims as universal laws. Wood then concludes that “If Hegel’s emptiness charge says that Kant is unable to provide an adequate deduction for a principle with any content and if the FUL is not empty, then we might pose Hegel’s emptiness charge as a challenge not to the FUL but to its derivation.” (Wood 162)¹¹

Wood finds some precedent in Hegel for understanding his “empty formalism” objection as a challenge to the derivation of the Formula of Universal Law (EL §54A; Wood 163). But, in any case, here is how Wood presents the problem:

Kant’s argument may be paraphrased as follows: A practical principle is objective, or a practical law, only if there is a ground for following it which is necessarily valid for every rational being as such. But a ground of this kind can consist in nothing but the agreement of the agent’s maxim with the concept of a principle that is objectively valid. This agreement, however, can consist only in the possibility of willing that the maxim be universally followed. Consequently, the only possible objective principle is: Act according to a maxim that you can will to be universally followed.

This argument is fallacious. An objective ground is universally valid, and likewise an objective principle is one that every rational being has a reason (the same reason) for following. The universal validity of an objective principle thus consists in what we earlier called its “universality of applicability.” But it does not consist, as Kant seems to infer, in the rational desirability (for someone, perhaps for anyone) of the state of affairs that would result if everyone followed the principle. Kant is arguing from the premise that it is rational for each person to follow a certain principle to the conclusion that it is rational (for someone, or anyone) to will that everyone follow the principle. But that inference is not valid. From the fact that it is rational for each of us to do something, we cannot derive any conclusion about the rational desirability of the state of affairs that would result if we all did that thing. (Wood 164)

Now, I don’t want to nitpick the way that Wood sets up Kant’s argument. In the last chapter, I did present Kant’s derivation as hinging on the concept of a categorical imperative, rather than on “the concept of an objective ground or an objectively grounded principle” (Wood 164). I also tried to explain why these concepts are not equivalent: hypothetical imperatives are objective, too. Mainly, though, I tried to show why the concept of an imperative is crucial to understanding the derivation. Wood does not include this concept, or at least he does not

¹¹ The “FUL” is, of course, the Formula of Universal Law. Wood also says “deduction”, where we have been using “derivation”. The deduction could be viewed as including the derivation, however.

emphasize it. But overall he presents the starting point of the derivation fairly and reasonably, and our differences here do not affect his subsequent argument.

Instead, the problem with Wood's interpretation lies in its understanding of the conclusion of the derivation: Wood claims that the Formula of Universal Law requires our maxims to exhibit not just "universality of applicability", but also a kind of collective rationality, which is to consist in "the rational desirability of the state of affairs that would result if we all" followed a certain maxim. As Wood puts it,

although Kant's conception of a practical law contains universality of applicability, it does not contain anything like collective benefit or collective rationality.¹² The concept of a practical law involves the idea that taking each rational being individually, it is rational for that being to follow the law; but this leaves entirely open the question whether the result of everyone's following the law would be (collectively) desirable, or whether it would be something that a rational agent could will without contradiction. A priori we have no way of ruling out the possibility that a principle with universal applicability might dictate to each individual a course of action which she cannot will that others should follow too. (Wood 165-166)

Thus, Wood concludes that Kant's derivation fails because of an unjustified slide from universal applicability to collective rationality, based on an equivocation in the meaning of a "universal" law: "universal" because applying to each rational being separately, or universal because applying to all rational beings collectively?

¹² Wood is careful to distinguish collective benefit from collective rationality, to be sure, lest the categorical imperative start to sound like a utilitarian or consequentialist principle. He does, however, think that the two are similar enough for his purposes (Wood 165), and indeed similar enough that he is prepared to speak of things like "rational desirability". In fact, Wood is sometimes a little vague about what collective rationality is supposed to consist in. Sometimes it sounds like a terminological variant for being able to will a maxim as a universal law, but it seems clear overall that he means, and needs to mean, something richer, roughly like collective benefit or desirability, only rational and non-consequential (Wood 165-166). I actually think this comparison is misleading, though, and in a way that is important to the success of Wood's argument. Wood's thought seems to be that a world under maxims that can be willed as universal laws has less conflict in it between rational agents. The world of universal beneficence, for example, just has less conflict in it than the world of universal non-beneficence, though both are equally conceivable. So we appear to behave more rationally as a collective in the world of beneficence, and the conflict in the world of non-beneficence may even make it collectively irrational. Hence, the world of beneficence has a "collective benefit", rationally speaking, which makes it more "rationally desirable". Or perhaps the underlying idea is that contradictions in the will are really contradictions in something like the Rousseauvian "general will" (Rousseau 1997). But I have been arguing that contradictions in the will do not consist in these conflicts between agents, or in a will as representative of the general will, but rather in what each individual has to will under such maxims, when they are considered as universal laws. See, for example, the discussion in Chapter 5.5 above.

The problem with this interpretation, however, is that the Formula of Universal Law, as we have been understanding it, simply does not require our maxims to exhibit collective rationality at all. When the Formula of Universal Law requires us to be able to will our maxims as universal laws, Wood understands that to imply that it must be rational for all rational beings, taken collectively, to always follow that maxim. This is a plausible interpretation, and natural. But it is still mistaken, despite Wood's plausible explanation for how Kant could make the sort of equivocation that Wood alleges.¹³

Consider the case of beneficence, as discussed in the previous chapter. The problem there was not that everyone would be worse off under a universal law of indifference, nor that it would be collectively irrational in some other sense, still to be determined. The problem wasn't even that the particular agent in question wills both their own satisfaction and a world wherein that satisfaction would (arguably) be harder to come by, if not actually impossible. Instead, the problem was that when we take their maxim as a universal law, they actually end up having to will both to receive beneficence and not to receive it. This analysis does not hinge in any way on whether universal beneficence would make everyone better off or on whether it would be in some way more rational, speaking collectively. For all it does is show that strict obedience to this maxim, taken as a universal law, necessarily leads the individual will into an irresolvable conflict with itself. A world without such conflicts is surely more rational than one with them, but that

¹³ Wood's explanation is plausible, but, I think, unlikely. As he puts it, "Kant seems to have been misled by the connotations of his terms "practical law" and "universal law"" (Wood 165), because good legislators attend to whether their laws can both apply to everyone and advance some form of collective good. So, Wood seems to be saying, Kant slides between the universality to which his concepts entitle him and a richer notion implicit in his idea of the will as universal legislator. Likewise, Kant's "mistake is an easy one to make because terms like "universally valid" and "universally rational" can be taken either individually or collectively." (Wood 166) One can imagine how Kant could have made these errors. But they are nevertheless pretty striking errors for Kant to have made with such fundamental concepts and arguments, especially given his focus on the individual good will and his deeply rooted anti-consequentialism, which should be inimical to any kind of analysis in terms of collective good, rational or otherwise. Yet, in Wood's defense, Kant is plausibly read as suggesting that we consider ourselves as representatives of the general or collective will, under the influence of Rousseau (1997), perhaps. Plus it is a little hard to see what else willing a maxim as a universal law might consist in, or add to the contradiction in conception test. And Wood also draws a natural lesson from Kant's discussion of beneficence. It is primarily our account of maxims that helps clarify the issue, at least as far as I have been arguing here.

“collective rationality” is just a side benefit, not the reason for rejecting the maxim. The underlying contradiction is.

We can even extend this analysis to Wood’s case: the rational egoist. As he puts it,

Suppose I am an egoist, who thinks that one objective principle is the principle of self-love: “Always seek your own happiness.” In the second critique, Kant claims that if everyone followed this principle, the result would be a conflict between people’s ends and the destruction of everyone’s happiness. Let us suppose that Kant is right and that I (the egoist) agree with him. I admit I cannot will that everyone should follow the principle of self-love, since this would destroy my own happiness (along with everyone else’s). But does this tend in the least to show that the principle of self-love is not an objective principle? No, it does not...

If Kant’s argument were valid, it would be impossible to maintain without contradiction that a principle is objectively grounded and also that one cannot will that that principle be universally followed. But though it is doubtless false that the principle of self-love is an objective principle, the supposition that it is one leads to no contradiction. Hence Kant’s argument is not valid. (Wood 165)

To begin with, notice that collective rationality plays no role in Wood’s argument here, except perhaps only incidentally. Even by Wood’s own account here, when we consider the maxim of self-love as a universal law, the problem in the agent’s will is not supposed to arise from the fact that this maxim is collectively irrational, at least not *per se*, but from the fact that they end up willing both their own happiness and the destruction thereof, precisely as my account would have predicted. “I admit I cannot will that everyone should follow the principle of self-love, since this would destroy my own happiness”, as Wood puts it. So Wood is mistaken to think that Kant’s derivation of the Formula of Universal Law fails because Kant does not prove that it must require our maxims to exhibit a kind of collective rationality. For the Formula of Universal Law was never supposed to require collective rationality in the first place.

But what of Wood’s central, deeper claim, that on Kant’s view it should “be impossible to maintain without contradiction that a principle is objectively grounded and also that one cannot will that the principle be universally followed”? This does seem to be perfectly possible. Maybe Wood is right about this, even if willing a principle as a universal law does not require collective rationality?

Wood does make an important point here, but my interpretation of Kant allows us to see its mistake. If a practical principle really is objectively grounded, then it cannot necessarily lead to an irresolvable contradiction within the wills of the agents who adopt it. For that would just be to say that all rational beings must act a certain way, and not act that way at the same time, at least to the extent that they must will both one thing and its opposite. Or, to put the point a little differently, if a practical principle is objectively grounded, then a rational agent who does as they ought will follow it meticulously. Therefore, if following this principle necessarily leads their will into a conflict with itself, then we cannot consider this conflict to be a rational error, under the supposition that this principle really is objectively grounded. But that would be absurd. Their will just is their faculty of practical reason, so a conflict within their will must be rational mistake. Surely, then, no principle leading inevitably to such a conflict could be valid for all rational beings just as such, any more than a principle could be a genuine axiom of logic, if applying it properly inevitably lead us into theoretical contradictions. Wood said that an objective ground “is necessarily valid for every rational being as such”, and that “An objective ground is universally valid, and likewise an objective principle is one that every rational being has a reason (the same reason) for following.” (Wood 164) But no ground or principle can be valid for a rational being “as such” if it requires them to contradict themselves, and no rational being can have “a reason” to contradict themselves, just by definition. So, *contra* Wood, if a practical principle is really objectively grounded, then we must also be able to will it as a universal law, and Kant’s derivation of the Formula of Universal Law is valid.

To see this more concretely, remember that maxims must include the ground for the action they require. To suppose that a maxim is a universal law, then, is just to suppose that all rational being always act a certain way for the sake of a certain end whenever a certain ground obtains. For our rational egoist, their maxim would then be something like: self-love requires me to always act in order to satisfy or maximize it. Now if we consider this maxim as a universal

law, it then imposes these requirements on all rational beings, just in general. But these requirements now become incoherent.

Suppose I find myself in circumstances where the only way to satisfy my self-love would be through your disinterested beneficence. My self-love, as the ground of this maxim-cum-universal-law, would then require me to do what I can to receive your aid. So, willing being a form of acting, I will have to will for you to help me, and I will have to act “externally” in order to try to make that happen. In general, my maxim says that my self-love should be satisfied, so, if that requires your disinterested aid, then that is what I should get, it says.

But considerations of self-love just as much require you to refuse to help me, since we are supposing that there is nothing in it for you. Accordingly, when I evaluate the case from the point of view of the principle of self-love, I have to will your beneficence, but at the same time recognize that you are perfectly correct, rationally speaking, in denying it to me. Not just correct to deny me aid, but positively required to, even, under the very same principle. So if I will this maxim as a universal law, I must also will for you to refuse to help me. In which case I end up having to will both for you to help me and for you to refuse to help me. As a result, I cannot will this maxim as a universal law without putting my will into an irresolvable conflict with itself.

On Wood’s behalf, one response to this analysis would argue that there is no distinctively rational conflict here, because really we are just looking at the conflicting points of view of two agents. From my point of view, you should help me, but from your point of view, you shouldn’t. But that isn’t a contradiction within either will. Since each of us is correctly applying the principle in question *relative to our own self-love*, there is no genuine conflict here. Or, to be more precise, there is no genuine, internal, rational conflict, despite how our actions may come into conflict externally. Instead, the conflict lies ultimately in our incompatible self-interests, which is just a bare fact about the world.

This response, while plausible, simply ignores my original argument, which did not analyze the case only from our subjective points of view, but from the point of view of the maxim

itself, considered as a universal law. In fact, we are constrained to take the point of view of the law itself by the very supposition that this principle is objectively grounded—that it is valid for all rational beings, just as rational beings. From that point of view, to be able to will a maxim as a universal law, I must be able to will all of the consequences of that law. So when I apply that law to my own self-love, I find that I must will for you to help me, and when I apply it to your self-love, I find that I must will for you not to help me, because those are the consequences of this universal law. To will this universal law as such, then, my will would have to contradict itself. So there really is a conflict within my own volition under this supposed objective law, and one which will necessarily follow if I am to follow this principle of self-love as scrupulously as befits an objectively grounded practical law.

So, in short, if we focus just on the requirements of this supposed objective principle, we still find that its requirements are incoherent. Your disinterested aid is in my interest, but not in yours, so the objective, universally valid ground of the principle both requires me to receive your aid and not to receive it. Thus, the grounds of the principle conflict with what follows from them, and so it cannot specify a single, coherent state of the world, which is to be brought about through the actions of all rational beings pursuant to it—even when we consider them just as isolated individuals, and not as a collective.

In effect, then, Wood's principle of self-love says: let each pursue their self-interest to the best of their ability, and whenever individuals come into conflict, let nature settle it. The prize goes to whoever is stronger, smarter, more persuasive, better able to elicit sympathy, or whatever. However it works out is a matter of indifference, as far as the principle is concerned, since it specifies no rational basis for deciding the question. According to the principle, each is to do whatever it takes to secure their greatest self-interest—for everyone's self-interest matters equally to the principle; that is precisely what its universal validity and its universality of applicability consists in—and the principle itself says nothing about how the resulting conflicts

are to be resolved. So can we really accept that such a principle could be objectively grounded, when it necessarily conflicts with itself, and provides no rational basis for resolving this conflict?

Wood (166-167) does consider whether we could solve his issue with the derivation by allowing it to start with a notion of objective grounding and universal validity that includes collective rationality. Wood points out that this would make the derivation look trivial, but that this might not be an objection to this solution. For Wood, though, the problem is that starting the derivation with stronger concepts invalidates the subsequent deduction of the Formula of Universal Law's validity for us. "If we employ the stronger concept of objectivity in the first stage of Kant's deduction of the FUL, then we save that stage by trivializing it, but we only doom the deduction as a whole to failure." (Wood 167)

My answer to Wood here ought to be fairly obvious: since the Formula of Universal Law does not require collective rationality at all, the concepts needed to derive it are in no need of strengthening. And this is not what I have done in order to answer Wood. In any case, though, I will attempt to explain the *Groundwork's* deduction of the Formula of Universal Law in the next, and final, chapter.

So we can conclude that there is no "gap" or invalidity in the *Groundwork's* derivation of the Formula of Universal Law, as Wood argues there is. Wood offered us two takes on what the mistake is to be: either the derivation illegitimately slides from the concept of universality of applicability to the concept of universal or collective rationality, or else Kant improperly supposes that an objectively grounded principle—one which is valid for all rational beings just as such—can only be one that we can also will as a universal law. But, on the one hand, the Formula of Universal Law does not require collective rationality, so there can be no equivocation between universality of applicability and universal or collective rationality. And, on the other hand, contradictions in the will really are incompatible with the idea of a principle that is valid for all rational beings, *just as rational beings*.

Now: suppose that a universal law were tailored so narrowly that it could only ever apply to a single action, performed by only a single agent. For example, suppose that there were a law saying that “Everyone with this name, born on this date, with this physical description, must tell this lie to this loan officer at this bank, at this time, on this date, in order to get this loan.” As a matter of form, this law purports to apply to everyone, universally. But the content of the law is constrained so narrowly that only a single person could ever satisfy its conditions. And, moreover, the law is so narrow that it could only ever require a single action from that lone individual as well. So doesn’t it follow, trivially, that everyone could obey this law, so long as that single person can? And wouldn’t it also be trivial for this person to will that action as a universal law? It seems as though they could very easily agree that everyone should obey this law, given how narrowly it is constrained. And then—lucky them!—it also looks as though they can make this false promise to get their loan, without having to violate their Kantian moral scruples.

Now we seem to have a procedure that we can use to make any action whatsoever conform with the Formula of Universal Law. No matter what we propose to do, it looks as though we can always use this example to generate a universal law that everyone could actually follow, and that we could also will as a universal law, albeit only because the law is “gerrymandered”, “tailored”, or “particularized” so that it can only ever apply to a single, unique case. So now it looks as if the Formula of Universal Law cannot actually forbid any actions whatsoever, so long as we are specific enough about the corresponding universal laws. Quite the loophole, that.

This “tailoring” or “gerrymandering” problem is quite venerable and hoary—versions of it came out almost as soon as Kant published the *Groundwork of the Metaphysics of Morals*. And the responses to this problem seem to be of two general kinds—the unsympathetic responses, even on the part of some devoted and sophisticated scholars of Kant’s work, tend to think that cases like these are more or less immediately fatal to the Formula of Universal Law. The sympathetic responses, in contrast, tend to have a hard time taking these kinds of cases very

seriously. They tend to think that the language that we've used to describe the maxims in these cases doesn't really capture the volition in question. We don't usually will anything as specific as "Anyone with this name, with this date of birth, must tell this lie...", for example. After all, if I want to lie to get a loan, then things like my name, my date of birth, and my physical description normally just aren't relevant to my volition. They don't normally appear in my deliberation, for example, or in my intentions, because what's important to me—what makes this way of acting seem attractive to me in the first place—is usually just that I want money and this lie seems like a convenient way to get it. So making the universal law that specific looks to be a pretty transparent attempt to disguise my actual volition in order to evade the demands of the Formula of Universal Law.

I'm in basic agreement with this sympathetic line of response to these cases: the fundamental problem here is about the relationship between volition, as Kant understands it, and the language that we use to express or formulate volition. And I agree that the reasoning I just sketched is essentially persuasive for the normal run of cases, where the agent is trying to give their actual volition the gloss or mask of morality. But it is harder to see how this argument can answer every case of gerrymandering.

To wit: the sympathetic argument so far holds that the gerrymandered, "particularized" law cannot correspond to the agent's actual volition, because their volition just isn't sensitive to anything that specific. For, the thinking goes, the factors that show up in their intentions, their deliberations, etc., are always more general, and so their volition must perforce be more general, too.

But this argument hinges on an analysis of the agent's actual volition. It argues that, as a matter of fact, the particularized law does not correspond to the agent's actual volition, because the particularities mentioned in the law are irrelevant to what they actually will. To really defend the Formula of Universal Law, though, it is not enough for this argument to say only that the agent's volition *is not in fact* that particular; instead, the argument really needs to be able to say

that the agent's volition *cannot be* that particular. For as long as my volition can be that particular, it still looks as though we have a procedure for gerrymandering volition in order to make it satisfy the Formula of Universal Law, which then still looks to be in rather dire trouble.

For example, suppose that my volition actually starts out as Kant imagines, when he says that the "maxim of action would go as follows: when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen." (G 32, 4:422) Even if I start out with a maxim that is general like that, why can't I then change my maxim to the gerrymandered one, which I seem to be able to will as a universal law?

As long as the particularized maxim is really a maxim—which is far from obvious—Kant's views on freedom and autonomy appear to allow me to make this change. As an absolutely self-determining causality, as autonomy, as the property of being a law to one's self or of giving laws to oneself, freedom appears to involve a fairly radical capacity to choose one's own principles. In any case, though, in order to explain why I can't just adopt a gerrymandered maxim, the argument under consideration will have to say something like: my intention cannot be that specific, or the reasons for my actions cannot be that specific. Why not, though, either from Kant's point of view, or as a more general philosophical matter?

Now the questions start to get far harder than the sympathetic position might have hoped. Intentions, for example, are a matter of serious philosophical controversy themselves. Moreover, Kant hardly ever speaks of intentions, and has no obvious theory of them. So what can we say here, if we don't want to rest the case entirely on bare intuitions about intentions? And, as for practical reasons, why can't I sincerely believe that my situation and the reasons that it provides to me are wholly unique? Perhaps I have some kind of particularist view about practical reasons, or maybe I just think this one situation has no implications for how I or anyone else should act in any other situation. How would Kant answer someone with that kind of view? Why can't their maxims capture their particularist beliefs?

I think that this version of the gerrymandering case is a genuine problem for the Formula of Universal Law, because it challenges Kant to answer someone with a very different and incompatible set of ideas about practical reasons. And standard boilerplate about how reasons have to be general will not suffice as an answer here, because the generality of reasons is part of what is at issue in this case: the competing particularist view either denies that reasons have to be general at all, or, as I will understand the view, it contends that “particularized” reasons *just are* fully general, at least in the sense that Kant is interested in, precisely because we can will for everyone to act on them, as a universal law.

So now we do need a systematic answer to the gerrymandering problem, and we should expect to find it in Kant’s views on maxims. After all, we will need to be able to show why a maxim cannot be gerrymandered or particularized, as in this example, or else we will have to show why such a maxim cannot be willed as a universal law after all. (Or both, as I think is actually the case.) And, to be specific, the question is about what can and cannot go into a maxim. So the question also invokes what I have been calling the “basic structure” of a maxim: their basic parts, along with how these parts are to be related to one another. So now I want to use my account of the basic structure of Kant’s practical principles in order to answer this “particularist” version of the gerrymandering problem.

To do so, we’ll need to go back to the beginning of the account: to Chapter Two, where we showed how Kant thinks that the basic structure of practical principles is an instance of the basic structure of causal relations, and thereby ultimately an instance of the ground/consequence relation in his general logic, too. There, we saw how practical principles are causal principles, and how Kant thinks that causation, at its most general level, consists in a certain instance of the logical relation between grounds and consequences, where causes consist in a certain kind of ground, and effects in a certain kind of consequence. The Schematism of Causation made the crucial point: “The schema of the cause and of the causality of a thing in general is the real upon which, whenever it is posited, something else always follows.” (CR 275, A144/B183)

And we also saw how this schema makes causation into the real instance of the logical ground/consequence relation: causes are real grounds, and effects are real consequences. Moreover, the schema also illustrates how Kant thinks that it is of the utmost importance for causal relations to be necessary relations. Partly this derives from the logical necessity in the ground/consequence relation, but the schema of causation also indicates how “The schema of the cause and of the causality of a thing in general is the real upon which, *whenever* it is posited, something else *always* follows.” (CR 275, A144/B183, emphasis added)

The answer to the gerrymandering problems starts with this necessity inherent in Kant’s concept of a ground and of ground/consequence relations, and in how this necessity carries through into causal relations, and then ultimately into practical principles. Because the relation between a ground and its consequence is necessary, Kant thinks that it follows that ground/consequence relations must always take place according to a rule. Again, this is quite perspicuous in the schema of causation: “The schema of the cause and of the causality of a thing in general is the real upon which, whenever it is posited, something else always follows. It therefore consists in the succession of the manifold insofar as it is subject to a rule.” (CR 275, A144/B183) The rule here, I take it, has the general form of “relevantly similar causes produce relevantly similar effects”. I take it that this follows both from the nature of the underlying logical relation—similar grounds have similar consequences—and from the necessity that is supposed to inhere in this kind of relation. That is how Kant presents it in the passage just quoted, in any case. Because the effect always follows, whenever the cause is posited, causality “therefore” consists in this succession “insofar as it is subject to a rule”, Kant says. So the nature of the ground has a necessary consequence, I take it, and whenever this nature recurs, the consequence must recur as well. And so likewise for causes and effect, and for the grounds and consequences contained in a practical principle.

To finally solve the gerrymandering problem, then, I will want to argue that genuine causal laws—Newton’s Law of Gravity, for example—and Kant’s version of the maxim of the

lying promise involve the right kind of rule, while gerrymandered maxims do not. Material bodies of equal masses exert equal quantities of gravitational force, and whenever I believe I need money in the same way, I shall tell the same kind of lie to get it. Relevantly similar grounds, relevantly similar consequences. But in a “particularized” maxim of false promising, I could believe I need money in the same way in different circumstances, but without needing to lie in both cases. A particularized maxim is one that is rendered specific enough to guarantee that it can only ever apply once. So identical circumstances could even obtain at different times, and a particularized maxim indexed to one time would still not need to say anything at all about how to act at the other time. So, I want to say, the underlying ground/consequence logic allows the former as a genuine maxim, but not the latter.

We could still resist Kant’s account of logic and causation, of course. Why must the nature of the ground be repeatable? Why can’t grounds be wholly unique, never to recur again, or at least without any implications for other cases? Why can’t the rule linking ground and consequence be essentially particular and unique? Kant seems to think that something about the nature of logic itself prevents this, but it is not clear what this should be. Why should the mere logical concepts of ground and consequence rule out grounds and rules that are essentially particular? This is going to be the crucial point for answering the gerrymandering case, so I will return to it later. It is worth highlighting now, though, if only to properly locate the crux of disagreement about the gerrymandering cases.

First, though, let’s see how we can solve the gerrymandering problem, if we grant Kant that ground/consequence relations have to be rule-governed in this fashion. As a reminder, the problem is that no matter what we propose to do, it looks as though we can always fashion a maxim that we can will as a universal law, so long as we make the maxim extremely specific or arbitrary. How does Kant’s thinking about rules in ground/consequence relations help solve this problem?

Let's return to the beginning of the problem, in order to show how my account can carry through the sympathetic response to the problem that we started with here, but in a more systematic way, which can ultimately answer someone who believes that practical reasons can be essentially particular.

To begin with, then, the whole gerrymandering problem hinges on the idea that we can formulate principles however we please, at least within the bounds of the correct basic structure, and then make any such principle our maxim. For example, I start out with a maxim like: because I want money, I need to tell a lie in order to get a loan. But I suspect that this maxim is immoral, so I try to find an adjacent principle that will allow me to do the same thing, at least in substance, but which I can will as a universal law. So I formulate a principle like "because I have my name, I must tell this specific lie at this specific time to get this specific loan", and then I make that principle my maxim instead.

Perhaps I really can change my maxim this freely. I do not find that to be obvious, however. I have occasionally emphasized the difference between an agent's volition and how that volition gets expressed in language. Because of that difference, a change in the linguistic formula for my maxim need not imply a genuine change in volition. So notice that the tailored formula will not have the same practical requirements. Under my original maxim, if I'm on my way to the bank to try to get a loan under false pretenses and I receive news that a wealthy relative of mine has died and left me a substantial inheritance, then I am released from the need to tell the lie. If I no longer want money, then the ground of my original maxim no longer obtains, and so I am released from its consequences, too. (So long as there is no other ground with the same consequences, anyway.)

However, under the tailored formula, I must still proceed with my original plan. I still have my name, after all, so I must still proceed to tell the specific lie I proposed at the specified time nevertheless. When I set out to evade my moral scruples about lying, did I really commit myself to this? If not, then I did not really change my maxim, I merely misrepresented it. Or, at

best, I changed my maxim, but then abandoned it, or reverted back to my original maxim, once the tailored maxim became inconvenient. Either way, this reveals something about the underlying structure of my volition throughout this whole episode.

The basic idea driving my account here is that practical principles are to express rational requirements, be they subjective, in the case of maxims, or objective, in the case of practical laws. For the ground is supposed to make it rationally necessary to act a certain way, to achieve a certain end. So the tailored formula proposes that my name rationally requires me to tell a specific lie at a specific time. Perhaps I really do believe this, and perhaps I really do commit to acting this way—I will consider this possibility presently. But in most cases, framing my principle this way is a fairly transparent attempt to mask my true volition in a formula that appears to satisfy the Formula of Universal Law. In reality, the ground of my action is almost certainly still my desire for money. This is really what makes me believe that I have to tell the lie, as we can see when we investigate how I would act when this ground no longer obtains. And there is usually nothing about this specific lie or this specific time that is supposed to necessitate my volition, save that they are convenient for my purposes. Would I really still tell the appointed lie at the appointed time if another were more expedient?

So in most cases, this tailored formula can hardly be said to capture my actual volition. As we can easily determine when we stop to probe the actual ground, end, and means adopted in my will. What do I think actually requires me to act as I propose to do, rationally speaking, and what do I take to be actually required of me?

So far, all I've really added to the reasoning that I gave earlier is a clearer method for determining an agent's actual volition. For my account doesn't ask about intentions, and it has a specific interpretation of what practical reasons consist in. So let's move deeper into the problem now.

What if someone really does believe that their name requires them to lie in order to secure a loan? It is a little hard to take that suggestion literally. People do often believe that their

identity requires them to perform certain actions, to be sure, such as their identity as a parent, a friend, a citizen, or a member of a religion.¹⁴ We sometimes even think that our name requires us to act a certain way, but where our name is really a metonym for something like our honor or our reputation. That is clearly not the same as believing that our name *as such*, as a title or as a piece of language, literally requires lying: it is rather hard to interpret that in a way that makes it intelligible as a rational requirement. Speaking just intuitively, it is hard to understand how a rational being, in possession of their faculties, could genuinely believe that their name, just as such, required them to tell a lie in order to get a loan. That belief looks irrational on its face.

Still, though, human psychology is knotty, mysterious, and prone to odd twists and compulsions. So I am certainly not in a position to deny that someone could really believe this. And perhaps we can still adopt tailored formulae as our maxims no matter what we antecedently believe. Psychologically speaking, I seem capable of adopting and following the tailored formula, even once its full requirements are spelled out to me. Kant's views on freedom seem to allow this also, as I already suggested. Moreover, I have to admit that this tailored formula appears to have the form that my interpretation of Kant requires: it includes a ground—my name—and it states that this ground requires an action of me—telling a certain specific lie at a certain specific time—in order to secure an end—obtaining the loan. Nevertheless, though, my account can still show how such maxims do not, in fact, exhibit the proper form, because the proposed rational requirement cannot be genuine.

For there are two ways that we can understand volition under a principle like “my name requires me to lie to get a loan”: either this principle is an instance of a more general maxim, which explains why this ground requires this action, or else the necessity here is just supposed to be brute, irreducible, or not subject to further explanation. Essentially particular, as I have been putting it.

¹⁴ Thanks to Barbara Herman for encouraging me to be clearer and say more here.

But my maxim cannot be formulated properly, if it is also to be an instance of a more general maxim, which is to make it rationally intelligible. For example, we just considered some aspects of my identity that could more plausibly be taken to require lying. But if I think that my name requires me to lie because I need to protect my reputation for wealth or generosity, or because I need to buy medicine for my child or my friend, then my actual volition is not well described by saying that my “name” requires me lie. I might sincerely believe that, but my sincere belief does not guarantee that I express my maxim correctly. I might still fail to accurately report or formulate the actual ground of my volition. So where is the more general principle about parts of language, considered just as such, that will explain why some of them require lying?

The harder case is the one where I believe that my name requires me to lie, but where I think that this requirement is brute or irreducible. The problem here, though, is that I am now describing a psychological compulsion or a whim, not a potential rational requirement. Above, I emphasized that Kant thinks that causal relations always take place according to a rule, which has the general form of “relevantly similar causes have relevantly similar effects”. I also explained how this follows from the fact that causation has its origin in the logical relation between ground and consequence, where logically similar grounds have logically similar consequences. But to treat the relation in this case as a brute necessity is precisely to deny that it takes place according to a rule of this form.

The gerrymandered maxim is supposed to connect my specific name and a specific lie at a specific time, for this is precisely the maneuver that is supposed to allow me to will this maxim as a universal law. But a brute relation between these specific, singular objects cannot possibly take place according to a rule, specifically not one of the form that Kant requires. To restrict the relation to the specific case is exactly to deny that relevantly similar grounds must have relevantly similar consequences.

However, even if I believe that the necessity in the gerrymandered maxim is brute in the actual world, it might still seem to be open to me to believe that the relation in the maxim is rule-governed in the right kind of way either counterfactually or across different possible worlds.¹⁵ Counterfactually or even *per impossibile*, it seems as though I could admit that if there were to be anyone else in these exact circumstances, then they should act the same way, or I could allow that anyone in identical circumstances in other possible worlds ought to act the same way, too. In each case, the necessity in acting this way would still be brute and essentially particular, but it would seem as though there are still relevantly similar grounds and relevantly similar consequences, only across counterfactual cases and different possible worlds.

Except that, from Kant's point of view, it is still logically arbitrary and artificial to deny that relevantly similar grounds in any actual world must have relevantly similar consequences. The rule that is to be involved in the logical relation between grounds and consequences is to apply to *all* relevantly similar cases, not just relevantly similar counterfactual cases, or relevantly similar cases in other possible worlds. Therefore, the ground/consequence relation in Kant's logic objects to the entire idea of brute logical necessities, because they do not exhibit the right kind of rule connecting their ground and their consequence. Giving a brute necessity a counterfactual or possible-worlds modal profile does not change this basic situation.

Accordingly, now it would be incoherent to describe the necessity here as a *rational* necessity, since it cannot have the proper ground/consequence form that logic itself requires. So it must instead be a purely psychological compulsion or necessitation, or simply a whim. If I genuinely cannot resist this compulsion, then my action is not morally evaluable to begin with, since it is not a product of my free or autonomous agency. I do not really even act, in this case; my behavior is no more subject to the Formula of Universal Law than my cats' is when they wake me up for breakfast. But if I can resist the compulsion, but simply don't, then my maxim is

¹⁵ Thanks to Tyler Burge for pressing me to address these cases.

really of the form “act as you feel compelled to” or “follow your whims”. Such maxims pose no special problems with regard to the Formula of Universal law.

So the problem with gerrymandering maxims, on my view, is how the tailoring of the formula obfuscates, and is sometimes designed to obfuscate, either the actual ground of the agent’s volition, or the nature of the necessity that they propose to act under, or both. By insisting on the idea that practical principles have to impose rational requirements, that they have to present what makes a certain way of acting seem rationally necessary, my interpretation of Kant gives us the tools we need to see through this confusion. Like the reasoning we started out with, this account can show how a tailored formula does not really express the agent’s actual volition, at least in the normal range of cases. But this account goes farther, too, because it can also show how, strictly speaking, a gerrymandered formula cannot be a maxim at all, or, at best, it can only be a defective maxim, because such formulae cannot express well-formed rational requirements.

But now: the particularist. What about someone who denies Kant’s views about the ground/consequence relation in logic, or who, maybe more to the point, wants to know why the grounds and rules that Kant invokes can’t be essentially particular. As I have been interpreting him, Kant says that relevantly similar grounds require relevantly similar consequences. But why can’t there be a ground that is totally unique, totally particular, so that there could never be any relevantly similar ones?

As a matter of fact, perhaps there could be a ground that would have to be totally unique. I don’t think that we can rule that out. But in Kant’s logic, there cannot be any grounds that are unique or particular *in principle* or *essentially*, and here’s why: Kant thinks that logical relations are inherently conceptual, and concepts are inherently repeatable.

Kant explains his views on logic in several places, throughout his critical corpus. To help explain his notion of “transcendental logic” in the *Critique of Pure Reason*, he says that “logic in turn can be undertaken with two different aims, either as the logic of the general or of the

particular use of the understanding. The former contains the absolutely necessary rules of thinking, without which no use of the understanding takes place, and it therefore concerns these rules without regard to the difference of objects to which it may be directed.” (CR 194, A52/B76) Here, the idea that general logic abstracts from “the difference of objects to which it may be directed” already suggests that Kant’s logic is general in precisely the sense that I have claimed it to be. The Jäsche *Logic* gives the same general account of logic, saying that “this science of the necessary laws of the understanding and of reason in general, or what is one and the same, of the mere form of thought as such, we call *logic*.” (JL 528, 9:13) And even the Preface to the *Groundwork of the Metaphysics of Morals* starts off by saying that “All rational cognition is either *material* and concerned with some object, or *formal* and occupied only with the form of the understanding and of reason itself and with the universal rules of thinking in general. Formal philosophy is called **logic**” (G 1, 4:387). In both texts, the idea that logic is essentially “formal” also suggests that logic is general, in the way I am now proposing.

But we can also see this generality of Kant’s “general” logic quite perspicuously, if we carry the analysis of his views on logic through to his thinking about thought. Since logic “contains the absolutely necessary rules of thinking”, this is the natural next step for the analysis anyway. And thought, for Kant, is essentially discursive or conceptual: “Thinking is cognition through concepts” (CR 205, A69/B94), as the Transcendental Logic has it, while the Jäsche *Logic* explains that “Cognition through concepts is called *thought (cognitio discursiva)*.” (JL 589, 9:91) Putting these explanations together, then, we find that general logic contains the necessary rules for discursive cognition, for cognition through concepts.

And we already discussed the nature of Kant’s concepts a bit way back in Chapter One, but the point is worth revisiting now. Here, again, are Kant’s main explications of concepts:

A **perception** that refers to the subject as a modification of its state is a **sensation** (*sensatio*); an objective perception is a **cognition** (*cognitio*). The latter is either an **intuition** or a **concept** (*intuitus vel conceptus*). The former is immediately related to the object and is singular; the latter is mediate, by means of a mark, which can be common to several things. (CR 399, A320/B377)

All cognitions, that is, all representations related with consciousness to an object, are either *intuitions* or *concepts*. An intuition is a *singular* representation (*repraesentatio singularis*), a concept a universal (*repraesentatio per notas communes*) or *reflected* representation (*repraesentatio discursiva*). (JL 589, 9:91)

A concept is opposed to intuition, for it is a universal representation, or a representation of what is common to several objects, hence a representation *insofar as it can be contained in various ones*. (JL 589, 9:91)

The essential point here, of course, is that concepts are inherently general: they represent an object “by means of a mark, which can be common to several things”; they are “universal representations”, or “*repraesentatio per notas communes*”, that is, representations by common marks or signs; and they are “a representation of what is common to several objects”. And this generality is in precise contradistinction to intuitions, which are essentially singular and immediate.

So general logic, then, contains the necessary rules for cognition through general representations, according to Kant. In other words, Kant’s view is that logical relations are always at the level of types, and never only at the level of tokens. So if I judge that my particular name requires me to tell this particular lie, then this can only be true if there is a more general relation between this type of name and this type of lie. My particular name cannot require anything except as an instance or a token of a certain type of name, and my particular name can require a particular lie only as an instance or token of a certain type of lie. The logical relation itself has to be between the types, and types are essentially repeatable; they are always repeatable in principle, even if in fact there would or could only ever be one token of them. For concepts can only represent objects through that “which can be common to several things”, which marks those things out according to types, not just as tokens.

So there can never be any essentially or irreducibly particular logical necessities, according to Kant’s understanding of general logic. And, since rational requirements must conform to general logic, it will follow that maxims cannot include anything essentially or irreducibly particular, since maxims are to express rational requirements.

And now we can even see why tailored, “gerrymandered” maxims cannot actually satisfy the Formula of Universal Law after all, if they even count as maxims at all in the first place: tailored maxims cannot actually express a properly rational requirement, so we cannot will for all rational beings to follow such maxims. For these maxims cannot even be universal laws for all rational beings. Following tailored maxims would destroy or relinquish rational agency in these instances, even if they are unique, so they cannot constitute laws for all rational beings, and neither can we will them as such.

At minimum, then, we can conclude that the gerrymandering problem devolves into a disagreement with Kant’s views on logic, and not with his moral theory *per se*. So we can safely leave the discussion there.

6.5 Sunday Morning Tennis and Avenging Insults

Now, having answered Hegel, Wood, and the gerrymandering problem, we can conclude by discussing our last range of problem cases: maxims like “Avenge your insults”, “Play tennis at 10:00 on Sunday mornings”, and “Become a philosophy professor”. As a reminder, the issue here is supposed to be that the Formula of Universal Law gives the wrong results for these maxims: it improperly permits the first, while incorrectly forbidding the latter two. For the universality that the Formula of Universal Law requires is supposed to create a universal social practice that we can will as such in the first case, while destroying the social cooperation that the actual permissibility of the latter two depends on. To evaluate these maxims correctly, though, we once more need to look past their formulae and probe the underlying volition. When we use my account of Kant’s practical principles to do this, we can see that Kant’s account yields the correct results after all.

First, tennis. The problem here is supposed to be that if everyone tried to play tennis at 10:00 a.m. on Sundays, there wouldn’t be enough courts to go around.¹⁶ So it looks as though

¹⁶ There are a lot of things that we could fuss over here. Under the assumption that everyone is to be playing tennis at that time, maybe we would guarantee that there are enough courts to go around? Or

this maxim can't be a universal law in the first place, still less can it be willed as such, because not everyone can follow it. But surely our moral intuitions say that there is nothing at all wrong with following this maxim? We coordinate court times precisely to avoid such distribution problems—so why should moral analysis prescind from that coordination? And then doesn't the Formula of Universal Law give a false negative here, by forcing an improper frame of analysis?

Well, my account of practical principles requires maxims to specify a ground, an end, and an action as a means to that end. But "Play tennis every Sunday at 10:00 a.m." only specifies an "external" action, which, I argued before, we should really understand as an end to be achieved, because playing tennis every Sunday morning at 10:00 actually represents the satisfaction or success conditions for the maxim. By itself, then, the principle does not yet tell us what must be done to accomplish this end, nor why this end is a rational requirement. We could simply interpret the principle as stating its end as a categorical obligation, as its surface form might suggest, and as I did before in order to show that principles must include some ground. But when so much of its structure is left implicit, it would be better to explore what an actual agent might will by or under this principle.

For example, one agent might take themselves to be required to play tennis every Sunday at 10:00 a.m., but only when they are not subject to an overriding obligation, moral or otherwise. So, for example, such an agent would not break a promise to a friend in order to get their weekly game in, nor would they try to forcibly remove other players in order to gain access to a court. Such agents can still have "Play tennis every Sunday at 10:00 a.m." as one of their maxims, only they interpret this formula in such a way that it does not require them to violate their other commitments.

In contrast, imagine an agent that really did view their weekly tennis game as something approaching a categorical obligation. As something that they absolutely have to accomplish,

maybe folks would find other ways to play? This draws us into hard questions about contradictions in conception very quickly, so I will just grant that this maxim really wouldn't work as a universal law. Thanks to Pamela Hieronymi for pressing these questions.

whether they want to or not, no matter the consequences. Compatibly with that, they might think that there are some things that they cannot do in pursuit of that end, but they also might not. To be sure, they might never actually do anything morally questionable in order to play tennis at the appointed time. But if they still view themselves as required to, then they will think they have acted wrongly, or they will think they have been treated wrongly if their weekly game is denied to them.

This second kind of agent, too, can have “Play tennis every Sunday at 10:00 a.m.” as their maxim, though their volition is really quite different. This example is a particularly clear illustration of a recurring point in the study here: that there is no neat correspondence between linguistic formulae and volition. As such, if we are not careful, we can easily confuse ourselves in trying to match our moral intuitions with the results of the Formula of Universal Law. Our first case here, where a weekly tennis game is treated as subordinate to other moral considerations, is clearly permissible, both according to our intuitions and according to the Formula of Universal Law. The second case, however, is the opposite: that way of willing a weekly tennis game is clearly problematic, if not outright forbidden, even by our moral intuitions. Moreover, though, this is the interpretation of the formula and its maxim that runs into trouble with the Formula of Universal Law: if everyone viewed their tennis game as a categorical obligation, or tantamount to one, in any case, then there really wouldn’t be enough courts to go around. Either way, though, the Formula of Universal Law is back into alignment with our moral intuitions.

So I conclude that this purported false negative case rests on an equivocation. When we have an intuition that this maxim is permissible, I submit that it is really because we implicitly judge the formula of the maxim in the context of “normal” human volition. We tacitly assume that a person with a maxim of playing tennis at 10:00 a.m. does not view it as a categorical obligation, but rather adopts this maxim in the context of normal human pro-social and moral commitments about sharing court space, waiting your turn, skipping a week or rescheduling in the face of other commitments and obligations—none of which a person who viewed their tennis

game as a categorical obligation would be prepared to countenance. Yet this is how we interpret the formula when we think that it violates the Formula of Universal Law.

So, in short, it is perfectly possible for us to understand “Play tennis every Sunday morning at 10:00” as already including and respecting a socially coordinated distribution of resources, just as I have been suggesting we do when we view this maxim as representing an intuitively permissible way to behave. Thus, when we look at this case through the lens of my account, and focus on specifying what the agent takes to be rationally necessary and what they take to be rationally contingent, we can resolve this false negative problem fairly immediately.

“Become a philosophy professor” succumbs to a parallel analysis: either the agent understands the maxim to be constrained by considerations of division of labor, or they do not. Would they be willing, in times of sufficient scarcity, to grow food, or to perform other kinds of labor, if socially necessary? If so, then there is no longer any problem about willing this maxim as a universal law, since the whole problem was supposed to be that this maxim, considered as a universal law, would be incompatible with the socially necessary division of labor. But if the agent understands their maxim in a way that does not respect that division, then our moral intuitions would reject their principle also.

The vengeance case is similar: it seems difficult if we do not press too hard on what it would be to will according to the specified formulae. Kant himself mentions this example, but he does not explain why this maxim cannot be willed as a universal law. He says only that “someone can make it his maxim to let no insult pass unavenged and yet at the same time see that this is no practical law but only his maxim—that, on the contrary, as being in one and the same maxim a rule for the will of every rational being it could not harmonize with itself.” (CP 17, 5:19) But we can explain the problem with this maxim, once we use my account to investigate the corresponding volition.

So suppose that someone insults me, and I take my revenge. The problem with this way of acting becomes pretty perspicuous when we notice that the underlying idea is that it’s wrong

to be insulted, and you have to avenge yourself when you believe that you have been wronged. So I take my revenge on the person who insulted me. Now, what is to stop that person from feeling wronged in turn? Perhaps they think that they did me no wrong to begin with, or that my response is inappropriate or excessive in any case. Now, though, if my maxim were a universal law, I would have to judge that this person should take their revenge on me. They believe themselves to be wronged, which is exactly the case for my maxim. By my own lights, though, it would be wrong for them to take revenge on me: I have done them no wrong, I just took revenge on them for the wrong that I believe they did to me. I have to think that my vengeance was righteous, otherwise I would have to think that it was wrong for me to act on my original maxim. But if I acted rightly, then they must be acting wrongly if they take revenge on me for what I did to them.

So, when I consider this maxim of avenging insults as a universal law, I find that I have to judge both that my action ought to be avenged and that my action ought not to be avenged. Or, to put it a bit more generally, I have to judge that I ought to be wronged and that I ought not to be wronged. And the former is already a contradiction, notice: in general, it's simply incoherent for me to judge that I ought to be wronged. So if my maxim were to be a universal law, I would then have to contradict myself about what ought to happen. It follows, then, that I cannot will this maxim as a universal law after all.

6.6 Conclusion

So all three of these problem cases work out as our moral intuitions would have them, after all, at least if we understand the tennis and philosophy maxims to already respect the relevant kinds of social coordination. And the vengeance maxim turns out not to project an acceptable kind of social practice, because it ends up forcing incompatible judgments on us.

I have certainly not answered every problem for the Formula of Universal Law, not even in the neighborhoods of the problems discussed. I have answered Wood, and I have provided a framework for interrogating volition and for analyzing cases like the maxims discussed in the

previous sections. Beyond that, though, I have mostly just pushed the disagreements away from Kant's moral theory and into broader disagreements about logic itself. That was, after all, the main result of our discussions of Hegel's "empty formalism" objection and of the gerrymandering problem.

But perhaps that is only to be expected: a moral theory that wants to apply to all rational beings, and which wants to base moral analysis on the form of the principles that rational beings act on, will naturally be influenced by the nature of logic itself. So disagreements about logic will naturally redound to how the Formula of Universal Law functions, and to whether or not it will be adequate as a fundamental principle of morality. Disagreements about logic are, however, quite beyond the scope of the present study. So we can turn now to concluding the present work, by looking at the deduction of the Formula of Universal Law, as it occurs in the Third Section of the *Groundwork of the Metaphysics of Morals*.

7. The Form of Law and the Idea of Freedom: Kant's Deduction of the Formula of Universal Law

We have now offered both general and specific defenses of Kant's derivation of the Formula of Universal Law, as it occurs in the Second Section of the *Groundwork of the Metaphysics of Morals*, at any rate. If that argument is sound, then the Formula of Universal Law will turn out to be the fundamental principle of morality and its system of categorical obligations. As such, if there is such a thing as moral or categorical obligation, then the Formula of Universal Law would express the fundamental such obligation, and all genuinely categorical obligations would turn out to be an instance of that single, fundamental obligation. Which is why the Formula of Universal Law would also unify moral and categorical obligation into a single system.¹

However, the argument in the *Groundwork's* derivation does not show, nor is it intended to show, that there is any such thing as morality or categorical obligation at all. (Neither does the derivation in the *Critique of Practical Reason*, which is also to show that the Formula of Universal Law is the fundamental principle of morality, only via a different strategy for the argument.) The derivation in the *Groundwork* proceeded from an analysis of the concept of a categorical imperative, but Kant insists that that analysis cannot prove that categorical obligation is real, any more than a derivation of the fundamental principle of caloric fluid from its concept could prove that caloric is real.

So, put more technically, Kant's derivation of the Formula of Universal Law in the *Groundwork* does not even purport to show that that Formula of Universal Law is objectively valid, either for imperfectly rational beings in general, or for human beings in particular. And put less technically: Kant's argument in the derivation does not even attempt to show that there is any sense at all in which imperfectly rational beings really are obligated, required, or

¹ To be clear, as I understand Kant, we are subject to moral obligation by virtue of being subject to the moral law, which imposes these obligations. Perfectly rational beings are subject to that same law, however it does not have the character of morality for them, because they are not subject to any imperatives or obligations (G 25, 4:414).

necessitated to obey the Formula of Universal Law. For all the *Groundwork's* derivation tries to show, imperfectly rational beings, like us human beings, may make no error of any kind whatsoever if they ignore the Formula of Universal Law entirely. In fact, we could even be making a fundamental rational error, if we adhere to this Formula's demands.

So for the last part of our study here, we can now turn to the Third Section of the *Groundwork*, where Kant finally does argue that we human beings really are subject to the demands expressed in the Formula of Universal Law. (Though only in a practical respect, or from the practical point of view—more on that later.) It would then follow that morality and categorical obligation are real, and that the Formula of Universal Law, as their fundamental principle, is really valid for us. (Though only speaking practically.)

Thus, the task of the *Groundwork's* Third Section is simply to show that we really are subject to the fundamental requirement expressed in the Formula of Universal Law—that each of us really must “*act only in accordance with that maxim through which you can at the same time will that it become a universal law*” (G 31, 4:421). And, as I mentioned before, the argument for this conclusion is usually called the “deduction” of the Formula of Universal Law, because the Transcendental Deduction of the Categories from the *Critique of Pure Reason* aims at an analogous conclusion.²

And Kant's argument in the *Groundwork's* deduction has three main phases. First, building off of the concluding discussion in the Second Section, Kant argues that the fundamental principle of morality must also be the fundamental principle of freedom. The Second Section argued that morality, if it exists, requires us to act only on maxims that we can will as universal laws, and also that if autonomy exists, then the very nature of autonomy itself

² See note 5 in Chapter 5 for some commentary on the term “deduction”. Also, remember that the *Critique of Practical Reason* gives another deduction, which aims at the same conclusion, but which relies on the idea that “Consciousness of this fundamental law may be called a fact of reason” (CP 28, 5:31). Some commentators have thought that this change in argument indicates a rejection of the strategy in the *Groundwork*. I tend to think that the two arguments are complementary, but I will conclude the discussion here with some reflection on why Kant might have preferred the “fact of reason” argument, even if he never rejected the deduction in the *Groundwork*.

would require precisely the same thing of us, at least if we are to use autonomy in a way that is proper to its nature (G 47, 4:440). The Third Section then argues that analysis of the concept of freedom shows that freedom, in its proper sense, is just autonomy—is just the property of being a law to yourself. And, in turn, as in the Second Section, analysis of the concept of being a law to yourself is supposed to show that autonomy requires us to act only according to maxims that we can will as universal laws. So Kant then concludes that there can be no fundamental principle of autonomy, or of freedom in the proper sense, except the Formula of Universal Law. The fundamental principle of freedom would then coincide with the fundamental principle of morality and of categorical obligation.

In this way, the argument in the deduction starts by trying to prove that morality, freedom, and the Formula of Universal Law all impose the same fundamental requirements. The deduction would then show that the Formula of Universal Law would have to be the fundamental principle of both morality and freedom, which is why Kant concludes that “a free will and a will under moral laws are one and the same.” (G 53, 4:447)³

Next, Kant argues that rational beings “cannot act otherwise than *under the idea of freedom*” (G 53, 4:448). In other words, rational beings must think of their actions and their agency as free. Accordingly, Kant argues that all laws of freedom are really, objectively valid for all rational agents, at least in a practical respect.⁴

³ Henry Allison (1990, 201-202) has called the claim that “a free will and a will under moral laws are one and the same” Kant’s “Reciprocity Thesis”. Kant does not explain what, exactly, he means by this claim, but I take it that he means to say that a will with the property of freedom is the same as a will that is subject to the system of distinctively moral obligations. Kant believes this, I take it, because he argues that the fundamental principle of freedom would have to be the same as the fundamental principle of morality. More on this in the next section.

⁴ Strictly speaking, Kant only says that all rational beings “having a will” must act under the idea of freedom (G 54, 4:448). But that is equivalent to saying that rational agents must act under the idea of freedom, because rational agency, in the proper sense, requires a will. In Kant’s mind, rational agency requires reason itself to be capable of determining our actions all by itself, which in turn requires us to have a will. So a rational being without a will does not *act* in the proper sense. Instead, if a being does not have a will, reason can at best serve other causal faculties. So Kant’s clarification is important, but not strictly necessary, and so we may sometimes elide it. Again I will say more about what it means for a

Kant's point here is that rational agency carries a necessary self-conception along with it: to think of yourself not just as a rational *being*, but also as a rational *agent*, you must think of yourself as a free agent as well. Rational agents can be more or less aware of this commitment to their own freedom, to be sure, just as they can be more or less confused about this necessary self-conception. But, all the same, Kant thinks that if I am to think of myself as having a rational will, I must also think of myself as having a free will. Reason is committed to thinking of itself as free, Kant thinks, so rational agency is accordingly committed to thinking of itself as free agency.

However, a being that must attribute freedom to itself must also think of the fundamental principle of freedom as the fundamental principle for all of their action and volition—as their fundamental practical principle. For, on the one hand, the deduction has already argued that a free will is one that must act “only” on maxims that it can will as universal laws. And, on the other hand, we have already seen that freedom, for Kant, is the property of the will whereby it is an absolutely spontaneous, absolutely self-determining causality. A free will determines itself *absolutely*, not just relative to the desires and inclinations to which it is subject. Accordingly, the concept of freedom, along with its fundamental principle, must apply to, and be valid for, all of a free will's actions and volitions. If you think of yourself as free, then, you are committed to thinking of yourself as subject to the fundamental principle of freedom, too, and to thinking that this principle is your fundamental practical principle *tout court*.

So if a rational agent must think of themselves as free, Kant thinks it will follow that they must ultimately think of themselves as subject to the Formula of Universal Law, too. Rational agents must think of themselves as free, and the first phase of the deduction argued that free agents must act only on maxims that they can will as universal laws. So rational agents must also think of themselves as fundamentally subject to the Formula of Universal Law. Therefore, the

principle to be “objectively valid in a practical respect” in due course. But note that freedom “in a practical respect” is not the same as “practical freedom”. More on that point later, also.

necessary self-conception of a rational agent carries a commitment to the idea that the Formula of Universal Law is fundamentally valid for them.

However, the derivation in the *Groundwork's* Second Section was supposed to show that the Formula of Universal Law is also the fundamental principle of morality and of categorical obligation in general. And the Reciprocity Thesis reflects the identity of these principles. So if a rational being must think of themselves as fundamentally bound to the Formula of Universal Law, then they must also think of themselves as fundamentally subject to categorical obligation in general. In which case rational agents must view themselves as subject to the fundamental principle of morality as well. Or so this deduction would have it, at least.

Now, an important clarification: the argument in the deduction is supposed to show that the Formula of Universal Law is valid *objectively*. But Kant's argument so far wants the validity of the Formula of Universal Law to follow from a certain self-conception, which might make it seem as though the validity of the Formula of Universal Law hinges on whether or not someone happens to have this self-conception. And that would make it sound as if his argument here can only show that the Formula of Universal Law is *subjectively* valid.

However, the commitment to the validity of the Formula of Universal Law is supposed to be part and parcel of the self-conception of rational agents. So all rational agents would have to think of themselves as subject to the Formula of Universal Law. And, moreover, they would have to think of themselves as subject to this Formula *precisely because* they think of themselves as rational agents. So it would then follow that the Formula of Universal Law must be objectively valid, at least for any being that must think of themselves as a rational agent. For this self-conception is to hinge on something necessary about rationality and rational nature, and not on anything that could be idiosyncratic to a specific rational subject.

Still, Kant's argument in the deduction so far remains limited in important ways. I have already mentioned how this argument can only show that the Formula of Universal Law is valid "in a practical respect" or from "the practical point of view". For Kant's deduction only tries to

show that rational agents must think of themselves as subject to the Formula of Universal Law, which is far from saying that anyone is actually subject to the Formula of Universal Law just *simpliciter*, to say nothing at all of giving a theoretical proof of the validity of the Formula of Universal Law. The deduction is certainly not supposed to show that physics or psychology must accept that the Formula of Universal Law is objectively valid for us. In fact, the theoretical sciences could conclude that we are not free in the way that the relevant self-conception requires, without troubling Kant's argument here at all. (As we will see more clearly later.)

This limitation of the argument to the practical point of view is crucial, to be sure. But more than that, though, the argument so far is purely analytic: Kant analyzes the concept of freedom in order to show that the Formula of Universal Law is the fundamental principle of freedom, and he analyzes the concept of a rational agent in order to show that rational agents must think of themselves as free, and thus as subject to the Formula of Universal Law, too. That is supposed to show that the Formula of Universal Law is objectively valid, from the practical point of view, for all rational agents, or at least for all beings that must think of themselves as rational agents. This argument cannot show, however, that there are any rational beings of the relevant kind at all, still less that we human beings are such rational agents, or even that we must think of ourselves as such rational agents. So, likewise, the argument so far does not show that the Formula of Universal Law is valid for us, even from the practical point of view. Thus, for all that has been said so far, morality might still be a mere "high-flown fantasy" (G 8, 4:394).

Accordingly, for the third and final phase of the deduction, Kant worries, infamously, that "a kind of circle comes to light here from which, as it seems, there is no way to escape" (G 55, 4:450). I will try to explain this nature of this circle more fully in due course. For now, though, it is enough to note that Kant tries to show how consciousness of the distinction between phenomena and *noumena* resolves this worry about a circle in the deduction of the Formula of Universal Law. The *Critique of Pure Reason*, along with "common sense" reflection on our faculties, is supposed to show that we must think of ourselves as *noumena*. That self-

conception, however, requires us to think of ourselves as rational agents in the relevant sense, Kant argues. In which case the rest of the argument in the deduction is supposed to show that we must think of ourselves as subject to the Formula of Universal Law after all, at least from the practical point of view. Kant's deduction of the Formula of Universal Law is then complete.

To explicate the *Groundwork's* deduction of the Formula of Universal Law, then, we will likewise divide this final chapter into three corresponding parts: one on the idea that the Formula of Universal Law is the fundamental principle of both morality and freedom, one on the connection between rationality and acting under the idea of freedom, and one on why human beings must think of themselves as rational agents, in the relevant sense. This method will allow us to finally return to the questions about freedom and necessity first raised in Chapter Two, and to resolve them through the concept of the form of law.

7.1 The Reciprocity Thesis: Freedom, Necessity, and the Form of Law

So the first step in Kant's deduction of the Formula of Universal Law, in the *Groundwork of the Metaphysics of Morals*, at least, seeks to show that "a free will and a will under moral laws are one and the same." (G 53, 4:477) Allison (1990, 201-202) has called this claim Kant's "Reciprocity Thesis", since it holds that freedom and morality reciprocally imply one another, at least so far as rational wills are concerned: a free will is a will under moral laws, and a will under moral laws is a free will.

Now, I think that this nomenclature is essentially fine: Kant does say that "freedom and the will's own lawgiving are both autonomy and hence reciprocal concepts" (G 55, 4:450), and he does give arguments in both of these reciprocal directions. For example, besides arguing that the Formula of Universal Law is the fundamental principle of moral obligation, the *Groundwork's* Second Section also concludes that the fundamental principle of distinctively moral obligation must also be a principle of autonomy (G 40-41, 4:432-433). And then the Third Section begins by arguing that the fundamental principle of an autonomous will must be the Formula of Universal Law. And the *Critique of Practical Reason* also gives arguments in both of

these reciprocal direction, over the course of its own derivation of the Formula of Universal Law (CP 26-28, 5:28-30).

Still, whatever we want to call Kant's idea that "a free will and a will under moral laws are one and the same", he does not really explain what he means by this idea. I take it that he means to say that a will with the property of freedom is also a will that is subject to the whole system of distinctively moral obligations, and *vice versa*. As we are about to see more clearly, Kant begins the Third Section by explaining what freedom is, as a property of the will, and by arguing that the Formula of Universal Law must be the fundamental principle of a will with that property. Moreover, the entire strategy of the *Groundwork's* deduction is to argue from the need to attribute the property of freedom to the objective validity of the Formula of Universal Law. So the immediate context and the broader argumentative strategy both indicate that the Reciprocity Thesis means to say that a will with the property of freedom is the same as a will that is subject to the entire system of genuinely moral obligations.

However, whatever this Reciprocity Thesis is supposed to mean, Kant is clear about why he holds it: Kant argues that the fundamental principle of a free will is also the fundamental principle of morality and of categorical obligation, and the Reciprocity Thesis is supposed to follow from that argument. For, he asks,

what, then, can freedom of the will be other than autonomy, that is, the will's property of being a law to itself? But the proposition, the will is in all its actions a law to itself, indicates only the principle, to act on no other maxim than that which can also have as object itself as a universal law. This, however, is precisely the formula of the categorical imperative and is the principle of morality; hence a free will and a will under moral laws are one and the same. (G 52-53, 4:477)⁵

⁵ Of course, the other formulae of the categorical imperative are equally principles of autonomy. So it is at least a little confusing for Kant to say here that "the proposition, the will is in all its actions a law to itself, indicates *only* the principle, to act on no other maxim than that which can also have as object itself as a universal law" (emphasis mine). However, Kant famously, if a little cryptically, insists that the primary "three ways of representing the principle of morality are at bottom only so many formulae of the very same law, and any one of them of itself unites the other two in it." (G 42, 4:436) He also insists that the Formula of Universal Law has a certain kind of priority over the other formulae—it is "first amongst equals", if you will. For he says that "one does better always to proceed in moral *appraisal* by the strict method and put at its basis the universal formula of the categorical imperative: *act in accordance with a maxim that can at the same time make itself a universal law*. If, however, one wants also to provide access for the moral law, it is very useful to bring one and the same action under the three concepts

So it is clear that the basis for the Reciprocity Thesis is to be the identity of the underlying principles of freedom and morality, in the form of the Formula of Universal Law: Kant's "hence" before his statement of the Reciprocity Thesis can leave no doubt about its basis in this threefold identity. At this stage of the discussion, Kant can take it as proved that only the Formula of Universal Law could be the fundamental principle of morality. After all, that was the headline result of the argument in the Second Section. But why should the Formula of Universal Law also be the fundamental principle of freedom?

In short, the concept of autonomy is supposed to provide the link between the Formula of Universal Law, considered as the fundamental principle of morality, and this same Formula, considered as the fundamental principle of freedom. Kant's strategy at the outset of the *Groundwork's* Third Section has three phases: (1) to articulate the concept of freedom, (2) to show that freedom in the proper sense is just autonomy, and (3) to show how the Formula of Universal Law is the fundamental principle of autonomy. The Formula of Universal Law will then be the fundamental principle of freedom, in the proper sense. Then, since the argument in the Second Section's derivation has already shown that the Formula of Universal Law is the fundamental principle of morality, this argument in the Third Section will then be able to prove that the Formula of Universal Law is the fundamental principle of both freedom and morality. Thus, "a free will and a will under moral laws are one and the same."

mentioned above and thereby, as far as possible, bring it closer to intuition." (G 44, 4:436-437) So it is fair to treat the Formula of Universal Law as the "only" principle of autonomy, since it includes the other, equivalent formulae as well, and since the Formula of Universal Law has priority in "moral appraisal" anyway. Moreover, strictly speaking, the above statements of the Formula of Universal Law add this extra idea about a maxim making "*itself a universal law*" and having "as object itself as a universal law". Some commentators have tried to put a lot of weight on these expressions, but they are really just alternate ways of stating the idea that we must ask whether we could will our maxims as universal laws. If I adopt a maxim, it then follows that I must will the end that the maxim contains. One way to will a maxim as a universal law, then, is just to take the end already contained in the maxim and add to it the maxim itself, considered as a universal law. For example, in willing the maxim of the lying promise as a universal law, I will that I should get the loan, and that my maxim of the lying promise should be a universal law. Thereby, I will both the end contained in this maxim and the maxim itself, considered as a universal law. Thus, we ask whether the maxim's end can also be itself, considered as a universal law, and this is just equivalent to asking whether the maxim can be willed at the same time as a universal law. I take it that this reflexivity is supposed to help capture the idea of autonomy, self-legislation, or self-grounding.

In that spirit, Kant begins the *Groundwork's* Third Section by explaining that “*Will* is a kind of causality of living beings insofar as they are rational, and *freedom* would be that property of such causality that it can be efficient independently of alien causes *determining* it, just as *natural necessity* is the property of the causality of all nonrational beings to be determined to activity by the influence of alien causes.” (G 52, 4:446) Kant does not explain what he means by “alien causes” here, but it is clear that he means causes alien to or outside of the will itself. (Rather than, say, external to the organism as a whole, which might also seem to be an eligible reading.) For the will is “a kind of causality”, he says, while freedom is the “property of such causality” whereby it can act “independently of alien causes *determining* it”, “it” meaning the will considered as a kind of causality. So these alien causes must be alien to or outside of the will. Kant’s comparison of freedom to natural necessity makes the same point, since he likewise describes natural necessity as “the property of the causality of nonrational beings” whereby this causality really is determined “by the influence of alien causes”.

This definition of freedom is important and substantive in itself, since it already rules out determination by desire or inclination as unfree. Those would be causes “external” to the will, so they would already count as alien causes. Again, free choice can be influenced by desire and inclination, only a free choice cannot be determined solely by a desire or an inclination, or by anything “alien” to the will.

Moreover, it is worth emphasizing that Kant begins by treating freedom as a property of the will in general, and of the will’s causality in particular. The nature of freedom, then, would stipulate the nature of free actions, and that nature would then be expressed in the fundamental principle of freedom. But free wills might vary dramatically in kind, and some kinds of free wills might be capable of misusing their freedom by deviating from this principle, analogously to how imperfectly rational beings can be capable of making rational errors. In fact, the idea of freedom as absolute self-determination could easily imply a capacity to deviate from the fundamental principle of freedom, at least in certain kinds of free agents. The concept of freedom and its

fundamental principle will then constitute a norm or standard for such “imperfectly” free agents, which will determine the proper use of their freedom, but without implying that they are categorically unfree when they misuse their freedom. For they can still maintain their capacity to use their freedom correctly, just as imperfectly rational beings can maintain their capacity for right reason, even when they make grave rational mistakes.

I mention this partly for clarity, and partly because Kant’s identification of freedom and morality has long led his commentators to wonder about how immoral actions can still be free, under his picture. That is a complicated question, which I cannot try to thoroughly answer here, although I do believe that I have just stated the core idea of the solution to the puzzle. Much more would need to be said about Kant’s views on immoral action, however. Nevertheless, it remains crucial to understand the fundamental principle of freedom not just as the principle for the proper kind of freedom, but for the proper use of that proper freedom. And it should also be kept in mind that some uses of freedom and free causality might not be proper.

In any case, though, Kant hastens to point out that this “definition of freedom” as independence from the determination of causes alien to the will “is *negative* and therefore unfruitful for insight into its essence” (G 52, 4:446). So far, this definition only explains what freedom consists in by contrasting freedom and free causality with these “alien causes” and their natural necessity. Thus, the definition so far only says what freedom is not, only what freedom is independent of, and not what freedom consists in positively, in itself.

So Kant continues:

Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity. Natural necessity was a heteronomy of efficient causes, since every effect was possible only in accordance with the law that something else determines the efficient cause to causality; what, then, can freedom of the

will be other than autonomy, that is, the will's property of being a law to itself? (G 52, 4:446-447)⁶

So Kant concludes that freedom, considered positively, is just autonomy, or “the will's property of being a law to itself”.

Now, it is important to notice that freedom in the positive sense is supposed to include freedom in the negative sense, because Kant thinks that the will could not be a law to itself if it were also subject to determination by alien causes. Kant will still insist, of course, that autonomy really is compatible with determination by alien causes, so long as we consider the will from two different points of view, or in two different relations (G 60-61, 4:456). From one point of view—the point of view of natural science—the will is always determined by natural laws and natural necessity, Kant thinks, and so the will in nature is always determined by alien causes, like desires and inclinations. Thus, if the will is also to be autonomous, and thereby free, Kant thinks that there must be another point of view, wherein the will is not a part of nature, so that there is room for it to be free both positively and negatively, compatibly with the first point of view. That will be the practical point of view, and I will say more about what it consists in and what its commitments are later on.

In any case, the important point for now is that freedom as such is to be identical with freedom in the positive sense, because that is the sense that is supposed to give us “insight into its essence”. Which is why we can say that freedom, in the proper sense, is just autonomy, according to Kant's argument here at the beginning of the Third Section of the *Groundwork*.

So Kant concludes that freedom in the proper sense is the same as autonomy—the same as “the will's property of being a law to itself”. And his argument is perspicuous enough by itself,

⁶ Kant's commentators sometimes describe certain acts of volition, as in immoral actions and in some cases where we act from inclination, as heteronomous. However, Barbara Herman has persuaded me that, strictly speaking, there can be no such thing as heteronomy for the will of a normal adult human, according to Kant. If either deduction of the Formula of Universal Law is sound, it will follow that normal adult human wills are, and always have been, autonomous, and this is not something that we can relinquish, except perhaps *in extremis*, as through severe intoxication, illness, extreme distress, and so on. Certain acts of will can still be heteronomous *by analogy*, but not *per se*. For Kant, autonomy is a property of the will, as I just emphasized. If we have it, we have it, *ceteris paribus*. We can misuse our autonomy, to be sure, but we need not relinquish our autonomy in so doing.

and far enough from our current purposes, that we can safely forego explicating it more closely for the moment. Especially when our past discussions have already been at some not-inconsiderable pains to explain why Kant thinks that “the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited”, as the key premise of this argument holds. But I will still say a word or two presently about Kant’s thinking here.

Our current question, though, is only about how Kant goes from defining proper freedom as autonomy to the Reciprocity Thesis, when the rest of this paragraph continues, as already quoted, but repeated here for convenience and continuity:

what, then, can freedom of the will be other than autonomy, that is, the will’s property of being a law to itself? But the proposition, the will is in all its actions a law to itself, indicates only the principle, to act on no other maxim than that which can also have as object itself as a universal law. This, however, is precisely the formula of the categorical imperative and is the principle of morality; hence a free will and a will under moral laws are one and the same. (G 52-53, 4:477)

Why should it be that “the proposition, the will is in all its actions a law to itself, indicates only the principle, to act on no other maxim than that which can also have as object itself as a universal law”?

There are really two points here that need to be explained: why the concept of autonomy implies that the will is a law to itself “in all its actions”, and why the idea of a will that is a law to itself implies that this will acts under maxims that it can will as universal laws. We need to put both of these points together in order to arrive at the idea that the fundamental principle of autonomy says that free agents must act *only* on maxims that they can will as universal laws.

The former point is simpler, so let’s start there: for Kant, negative freedom implies that the will is not determined by any alien causes. But the will still remains a causality or a causal power, so it must still be determined by something. After all, Kant says that the concept of causation and causality implies determination by a cause or ground, according to a law. So if a free will is not determined by anything outside of or alien to itself, then it must always be

determined by itself, and by laws that it gives to itself. So a free will must always act according to laws that it gives to itself, and only according to such laws. This, in brief, is why negative freedom implies positive freedom, and why autonomy should be freedom in the proper sense.

So why should acting on maxims that you can will as universal laws be the same as being a law to yourself? That question is actually quite trickier than it might seem, particularly given how Kant presents this question as hardly needing answering at all. Kant appears to be thinking something along these lines: the will is a faculty of causality. So if we can will a maxim as a universal law, then it follows that we could make that maxim into a law for all rational and autonomous wills, if only we had the power to do so. The Formula of Universal Law, then, looks as though it should be the fundamental principle of autonomy, insofar as it requires us to act under maxims that we could make into laws for ourselves, considered just as autonomous wills.

But that line of thought does not fully resolve the need to explain why acting on maxims that you could will as universal laws should be the same as being a law to yourself, particularly in the light of the idea that the Formula of Universal Law is also the fundamental principle of morality. Suppose that I am conscious of being subject to the Formula of Universal Law, because I am conscious of being subject to morality's categorical obligation. Still, I might experience that obligation and its necessitation on me as precisely the kind of "alien cause" that is to be incompatible with the notion of freedom, particularly when fulfilling that obligation would require me to relinquish something that I desire profoundly. Kant sometimes says that the moral law "*strikes down* self-conceit altogether" and even "humiliates it" (CP 63, 5:73), which an agent could easily experience as the effect of an "alien cause". And moral obligation might appear as an alien cause, whenever we focus on the necessity that it imposes on us, or on the sense in which we are subject to the moral law and to categorical obligation, rather than on the sense in which we are supposed to give that necessity to ourselves, too.

If the Formula of Universal Law is to be the principle of categorical obligation, then, it seems as though agents could easily experience it as an "alien cause" operating on them and

infringing on their freedom. But, moreover, it is not so easy to see why acting on maxims that can be willed as universal laws should be equivalent to being a law to yourself, because of how difficult the notion of “being a law to yourself” is in the first place. The law here is supposed to be a law for your causality, and so this law must carry necessity with it, Kant thinks. After all, the underlying premise here is that “the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, *must be posited*” (emphasis mine). But at the same time, you are supposed to give yourself that law, along with the necessity that it contains. How can that be possible? The idea of giving yourself a necessity seems incoherent. If you give it to yourself, can’t you just as easily release yourself from it? In which case, how was it ever a necessity to begin with? Something akin to the paradox of voluntarism is looming here, and if it turns out that the idea of giving a law to yourself is incoherent, then it would be unhelpful, to say the least, to say that the Formula of Universal Law is the fundamental principle of autonomy because it tells us to act on maxims that we could make into laws for ourselves, and for all autonomous agents.

So here we are, once again in the full throes of reconciling freedom, as absolute self-determination, with the necessity inherent in laws and in causality itself. How does the idea of acting on maxims that you can will as universal laws help us to reconcile these apparently competing claims of freedom and necessity? How does it help us to understand what a “lawful” freedom would look like?

To answer these questions, we need a deeper analysis of the idea of “being a law to yourself”, which the previous chapters have finally put us in a position to provide. There are really two elements here that we need to understand: the concept of a law, and the concept of giving or being a law to yourself. The former concept represents necessity, while the latter represents freedom, as absolute self-determination. So we will first analyze these parts, before trying to put them back together, into the proper Kantian synthesis of freedom and necessity, in the form of autonomy.

Let's start with the concept of a law. We want to understand what it is for a will to be a law for itself, or for the will to give laws to itself. And, more specifically, we want to know why the will's being a law for itself is equivalent to its acting on maxims that it can will as universal laws. So, really, we want to understand what maxims a will can adopt, such that it thereby counts as a law for itself, and why those maxims should be only those that can be willed as universal laws. Posing the question in those terms is apt to the will's nature as the faculty for deriving actions from principles (G 24, 4:412), and this framing will also allow us to use the concept of a maxim to connect the idea of a will that is a law to itself with the concept of a will acting on maxims that it can will as universal laws.

So our first question should be: what kind of maxims could count as a law, so that they would be suitable or apt for the will's self-legislation? In Kant's view, not every maxim could be a law, so his basic idea here is that the concept of a law imposes constraints on what counts as "being a law to yourself". You have to give yourself a law, so what counts as a law will modulate and constrain what counts as giving a law to yourself. That idea is controversial, of course: we have already seen how some of the oldest objections to Kant's moral theory have held that really any maxim whatsoever could count as a law for yourself, so that the idea of "being a law to yourself" would be normatively empty—an "empty formalism". I tried to answer some of those objections in the previous chapter, though, so we can take it for granted here that the concept of a law really can tell us something substantive about what it is to "be a law to yourself".

In fact, the overall purpose of the study so far has been to articulate Kant's concept of a law to the point where we could actually try to say something about the constraints that this concept might impose on the idea of "being a law to yourself". After all, the account of Kant's practical principles that I have been developing here is supposed to derive from and to help articulate Kant's more general thinking about laws. Specifically: my account of the basic structure of Kant's practical principles is supposed to help articulate what Kant sometimes calls "the form of law" (CP 26, 5:28-29), or "the form of a law in general" (CP 61, 5:70).

To wit: under my interpretation, the form of law (in general) consists, for Kant, in a certain kind of relation between a ground, its consequence, and whatever it is about that ground that makes it have that consequence. The ground/consequence relation in general logic is basic and fundamental here, Kant thinks. In logic, the relevant kind of relation is between a ground, a consequence, and the logic of the ground whereby it must have this consequence. Causality is to be an instance of this logical relation, where the relation is between a cause, its effect, and the action of the cause whereby it must produce its effect. And in volition, the relation is to be between a ground for action, an end, and the action that is to be the means to that end.

Thus, Kant's form of law consists in a certain kind of triadic relation, obtaining between these three elements. Strictly speaking, though, the form of law is to consist in the specific way in which these three elements relate to each other: specifically, the relation between them must be both necessary and rule-governed, in the sense that whenever the ground obtains, its consequence must follow, because of whatever it is about the ground that makes it have this consequence. Thus, this relation also exhibits a strict kind of universality as well, since there can never be a ground of this kind that does not have these consequences.

So if my account here is correct, then the form of law, for Kant, consists just in a relation of this kind, between these three kinds of elements. And, to be clear, if a relation has the form of law, that only indicates that this relation could be a law, not that it actually is a law. All actual laws have the form of law, of course, but the form of law is what both actual and possible laws have in common. How else could morally permissible maxims, as principles representing only subjective laws, also have the form of objective, universal law?

Hence, a maxim has the form of universal law whenever its ground, end, and action stand in a relation to one another that could be universal, necessary, and rule-governed, in the way required by the form of law. In contrast, genuine causal laws and the actual laws of logic are always objective, because the ground in the law always guarantees its consequence. In volition, though, laws can be either subjective or objective, because some grounds might be idiosyncratic

to a particular subject. My grounds must involve rules for my actions across time and relevantly similar circumstances, since grounds must be rule-governed in that way, according to Kant. That is why it makes sense to say that maxims represent subjective “laws”. And some practical grounds will generalize across agents, but some will not: sometimes the same something might be a ground for me, but not a ground for you. Under Kant’s concepts, something is a ground only because it stands in the right kind of relation to the right kind of consequence. So if we are both free, that something might be a ground for me, but not for you, or the same ground might have different consequences for both of us. Hence the distinction that Kant draws between maxims and practical laws.

When it comes to Kant’s concept of autonomy, then, the crucial point is really about the form of objective law: the form of laws that are universal, in the sense of applying to all autonomous agents. (Kant often speaks of the “form of law” when by my lights he more specifically means the “form of objective law”.) An autonomous will is to be a law to itself, or to give laws to itself. However, this must mean that it is to give laws to itself *precisely as an autonomous will*, in abstraction from any of its subjective idiosyncrasies. (Or, maybe better, in abstraction from and in addition to its subjective idiosyncrasies.) Accordingly, the laws that an autonomous will gives to itself must be capable of being laws for any autonomous will. (And only capable of that, because no single autonomous will can actually make a law for any other autonomous will, except in the sense in which an autonomous will gives itself the fundamental principle of autonomy and its consequences.)

Therefore, the form of objective law is supposed to provide a constraint on how an autonomous will can be a law to itself. For any law suitable for this self-legislation must exhibit this form, or else the will’s autonomy will be misused. And we previously examined some examples of principles that are supposed to be incapable of having this form, such as the maxim of the lying promise and the maxim of theft, from the previous chapter. So an autonomous will cannot act properly, according to its nature as autonomous, if it follows maxims of that kind.

So here, then, is one necessity that an autonomous will is to be subject to, just by dint of its own nature as autonomous. Adhering to the form of objective law is only a necessary condition for being a proper law to yourself, however, not a sufficient condition. This form is to make a law suitable for autonomous lawgiving, but it does not yet imply that the act of giving such a law to yourself would be autonomous or self-determined. For that, we need to understand what it is to give a law to yourself.

To be a law to itself, the will must both give and receive the law: the will must be both the author of the law, and subject to the law at the same time. So the nature of this will must be relevant to what laws it can give to itself, and to what it means for the will to be a law to itself. In the first place, the will is practical reason itself, Kant thinks. As such, any law that the will could give or receive must conform to the prior principles of reason, primarily the principle of non-contradiction. This point should already be familiar from the last two chapters, and it means that non-contradiction is another constraint on autonomy. Therefore, adhering to the form of objective law is not enough for autonomy: we must also be able to will our principles as objective laws, because we must be able to accept them as such without having to contradict ourselves. We have already seen how these ideas are equivalent, for Kant.

So, when we put these last two points together, we can already see how the Formula of Universal Law turns out to be a necessary condition for autonomy: to give a *law* to yourself, your principle must have the form of objective law, and to give a law *to yourself as a will*, you must be able to accept that principle as a universal law without contradicting yourself. So the Formula of Universal Law turns out to be fundamental to autonomy, as Kant thinks that we can see just by analyzing the concept of a will that is a law to itself.

Still, though, we have only given constraints that the will must meet in order to be a proper law to itself, and so we have only given necessary conditions for autonomy proper. Following the Formula of Universal Law is also supposed to be sufficient for proper autonomy, but we have not explained that. Not yet. For we have only explained how acting on a maxim that

we can will as a universal law does not violate the concept of a will that is a law to itself. We still have not explained the positive sense in which acting on a maxim that we can will as a universal counts as positively autonomous, as positively self-determined.

To be clear, though, Kant thinks that we sometimes act autonomously in a positive sense just when we respect these necessary conditions on autonomy. We can reject a maxim that we cannot will as a universal law, and we can also act a certain way because we cannot will its opposite as a universal law, as when we tell the truth because we cannot will a lie as universal law. Both of these ways of acting already count as autonomous in a positive sense, because these positive acts of will are already directly required by the constraints on autonomous volition. These actions are directly required by the nature of autonomy and its fundamental principle, we might say. Such actions count as positively “self” determined, then, because of how they are positively required by our natures as autonomous and rational beings—by our self, as we are rational and autonomous.

But it is harder to see how volition that is not positively required by the Formula of Universal Law could count as positively autonomous. If the Formula’s two constraints on autonomy merely permit us to act a certain way, but do not require it, then how would we give a law to ourselves, if we act as we are permitted to do? There has to be an underlying necessity there, if there is to be a law to give to ourselves. But where would that necessity come from? And how would it could as self-legislated?

Well,

There is, however, *one* end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely *could* have but that we can safely presuppose they all actually *do have* by a natural necessity, and that purpose is *happiness*. (G 26, 4:415)

I mentioned before, in Chapter Two, that it is puzzling how a free, autonomous will could have an end “by a natural necessity”. I take it that Kant means “by a necessity inherent to their nature”, though, rather than “by a necessity of the natural world or of the laws of nature”.

(Kant's picture of the will, I take it, is rather like Saint Anselm's, in *On the Fall of the Devil* (*De Casu Diaboli*) (2007).) Our wills are just constituted with the end of happiness "hardwired" or "built in", so to speak, which is why Kant connects this end to our nature "insofar as imperatives apply to" us, "namely as dependent beings", rather than to our human nature.

Be that as it may, though, the crucial point for our current purposes is that happiness is supposed to be a source of necessity in our volition. It is to be an end that we must have, and so it will always structure our agency, at least in part. As we have seen already, an end is a state of affairs that we are committed to realizing, so an end that we must have implies a necessary practical commitment, and a corresponding necessity to act towards achieving that end. Therefore, to say that an action is permissible from the point of view of the Formula of Universal Law is not to say that it cannot be necessitated at all. For the necessity derived from happiness remains, Kant thinks.

This is not to say that happiness has to be taken in the narrow, egoistic, personal utility maximizing, rationally self-interested "economists" sense of happiness, however. Our happiness can, and presumably will, vary wildly with our inclinations and desires, which we can perfectly well suppose can be genuinely sympathetic, altruistic, or otherwise socially- or other-oriented. Still, though, however we conceive of it, our happiness requires us to act a certain way, Kant thinks, because of how we cannot help but have happiness as one of our ends.

Happiness may not be able to command anything very specific, because of how maddeningly difficult it can be to find anything that is "indispensably necessary" to our happiness, as a genuine hypothetical necessity would require (G 28, 4:417). Moreover, Kant will surely insist that the demands of happiness are always only conditional, since the Formula of Universal Law takes priority over them, by dint of its status as a principle of reason, morality, and autonomy. So, in the terms introduced in Chapter Two, the practical necessity in happiness will characteristically be highly general and highly conditional. Accordingly, that necessity will typically be quite weak and quite highly attenuated, and, as a result, often quite difficult to

perceive as well. Still, though, Kant thinks that there always remains a residual, structural necessity to our own happiness, which we cannot help but be committed to, and which we cannot help but care about.⁷ A person cast into a suitably byzantine maze might have no idea at all of how to escape. There might also be multiple equally good paths to the exit, for all they know. Or perhaps there is really no way out at all. For all that, though, it can still be necessary for them to escape, and that necessity will carry through, in some way and to at least some degree, into the means that they adopt towards that end.

In this way, the practical necessity deriving from happiness, as a naturally necessary end, bears a certain resemblance, for Kant, to the moral necessity deriving from our own perfection and the happiness of others. We are categorically obligated to adopt those ends—they are “**Ends That Are Also Duties**” (MM 150, 6:385), as Kant puts it. However, despite the strict, unconditional necessity in adopting these ends, Kant still thinks that “the law cannot specify precisely in what way one is to act and how much one is to do by the action for an end that is also a duty” (MM 153, 6:390). That does not mean, though, that we are under no necessity at all to pursue our own perfection or the happiness of others. Quite to the contrary, actually.

Much, much more would need to be said about Kant’s obligatory ends in order to explain what their corresponding practical necessity involves, how that necessity can be limited and even discretionary, to an extent, as well as to explain how that practical necessity differs from the necessities in happiness and in strict or “narrow” moral obligations. Still, though, Kant’s doctrine of “ends that are also duties” helpfully illustrates how he thinks that general, non-specific, “wide” practical necessities can derive from an end that it is necessary for us to have, in some sense of “necessary”.

Moreover, happiness represents an end of a very intimate sort: no matter what our happiness contains for us, it is fundamentally *ours*. It derives from our particular, idiosyncratic,

⁷ The idea that we cannot help but have happiness as one of our ends, so that we cannot help but care about it, plays a large role in Kant’s discussion of stoicism and in the antinomy of practical reason (CP 92ff, 5:110ff).

subjective natures, in all the messy, knotted, ragged glory that reason and autonomy are supposed to abstract away from, at least when we consider them just in themselves. And happiness represents our deepest personal projects and aspirations, along with our most profound hopes for success, satisfaction, and fulfillment in our lives.

For Kant, then, an action that is a matter of indifference for the Formula of Universal Law, and hence for the necessary conditions of autonomy, can still count as positively self-determined, can still count as the product of genuine *self*-legislation, because of how such actions can still derive from or connect back to the necessity in happiness. As long as our pursuit of happiness is constrained by the fundamental principle of autonomy, Kant can allow happiness to be a source of laws for ourselves, of genuinely self-given necessities, because happiness is an end that we have by a necessity of our natures, and because that end represents and derives from our particular selves—the particular self that is rational and autonomous.⁸

Now, let's tie the above points together, into Kant's account of autonomy. We will then be able to explain why he holds that the Formula of Universal Law is the fundamental principle

⁸ Barbara Herman asked me about a case where a rational being lacks the idea of happiness, and instead just adopts ends and acts accordingly on the basis of instinct. Without happiness as an end to tie together, unify, and structure their instincts, my account so far might make it seem as though this being's instinctual actions do not connect back to any practical necessity, in which case my account so far might also make it seem as though their actions cannot be positively self-determined, and thus perhaps even outside of the scope of evaluation under the Formula of Universal Law. This is a challenging and interesting case. I do not feel certain that Kant would accept the case as described, however. The case asks us to suppose that this rational being acts on instincts unsystematically, without the idea of happiness as the sum of the satisfaction of these instincts. (Which is close to one of Kant's several concepts of happiness (G 12, 4:399).) However, if this being's rationality is to involve pure practical reason, then Kant might not allow us to suppose that they lack this idea of happiness, since pure practical reason requires all of our ends to be united into the concept of the highest good (CP 90-91, 5:108) It would then seem as though this being would need this concept of happiness as one of the components of this concept of the highest good, or else we cannot suppose that they possess pure practical reason. (And then the question here would lose its bite for me, since pure practical reason is equivalent to autonomy, for Kant (CP 3, 5:3-4).) Be that as it may, though, if we understand this being's instincts to be implanted in or deriving from their nature, then there is no barrier to understanding their instincts as a source of practical necessity, analogously to how happiness is supposed to be an end for us by a "natural necessity". We could then apply the account given above to this case, too, *mutatis mutandis*: their instincts cause them to adopt and pursue ends, as we supposed *ex hypothesi*, so these instincts would be grounds for them, by a natural necessity, instead of an end that they have by a natural necessity. In general, though, these instincts must derive practical necessity from somewhere, otherwise they cannot be grounds for a rational being, and that source must somehow connect back to this being's nature, otherwise they cannot be grounds for an autonomous rational being. That is what matters, ultimately, not whether this source is a unified end.

of autonomy, and hence the fundamental principle of freedom in the proper sense, as the Reciprocity Thesis requires.

As we have been explaining it so far, Kant's picture of autonomy has four elements or moments, which are to be combined together and structured into the fundamental principle of autonomy—into the Formula of Universal Law. The first moment is negative freedom, which holds that the will cannot be determined by any “alien causes”. As a result, the will's volition is to be the result of its own free choices. Autonomy itself—positive freedom—is then supposed to follow from negative freedom, and autonomy is also to include the moment of negative freedom within its concept. Moreover, the concept of autonomy is to contain the second and third moments within itself, as well. Autonomy, Kant thinks, consists in the will's property of being a law to itself. Thus, the second moment is the form of objective law, as the form of the *laws* that the will is to give to itself, while the third moment is reason—the will is the faculty of practical reason, so the principle of non-contradiction will be the fundamental form of the will's law-giving. And the fourth, and final, moment is happiness, which the will is to have as one of its ends, “by a natural necessity”.

Now, the Formula of Universal Law represents the synthesis of these four moments into a single principle. This Formula says “*act only in accordance with that maxim through which you can at the same time will that it become a universal law.*” (G 31, 4:421) The maxim represents the moment of happiness: the maxim is to be the subjective principle, which the agent brings to the Formula of Universal Law for evaluation. As a practical principle, their maxim contains their end, the action that they propose as a means thereto, and the ground that they propose to act from, which, as subjective, will commonly derive from their desires and inclinations—from their happiness. (The maxim could also represent something that the agent takes to be morally necessary, but, barring that, the maxim will represent something in the pursuit of happiness.)

Next, the idea of “willing” the maxim as a universal law represents the moment of reason, since Kant thinks that the will is practical reason, and that non-contradiction is the fundamental standard for what we can and cannot will. The concept of a universal law, of course, contains the form of objective law, which is necessary for self-legislation. Furthermore, the command to “act” according to maxims that we can will as universal laws contains the moment of negative freedom, since this command presupposes that it is always up to us to obey the Formula of Universal Law. And, last, the Formula of Universal Law represents the unity of these moments, synthesized into a single whole, by subordinating all of these moments to the demands of what we can will as a universal law. Act “only” according to maxims that we can will as universal laws, it commands, and thereby it reveals to us the full, unified structure of the moments that it contains.

We can see, then, how the Formula of Universal Law is to be the fundamental principle of autonomy, since it explains how to structure the practical necessities in the form of law, reason, and happiness into a coherent and autonomous whole. In following the Formula of Universal Law, our actions will always conform to the necessary conditions of autonomy, and they will always derive from a necessity inherent in an aspect of ourselves as well: be it our nature as law-governed beings, as rational beings, or as dependent beings, who cannot help but care about our happiness. As a result, all of our actions under this Formula will count as both necessitated and as absolutely self-determined, and Kant takes himself to have resolved the competing demands of freedom and necessity in the Formula of Universal Law.

For example, I suggested earlier that a person might experience the constraints of categorical obligation as an “alien cause” imposing on their freedom. And, indeed, conflicts between the necessities deriving from the different facets of our selves are quite common. These conflicts are a problem for our ability to act coherently, whole-heartedly, and autonomously. But the Formula of Universal Law seeks to resolve these conflicts by reminding us that autonomy and reason are parts of ourselves as well, so that the demands of the form of law and of reason

itself are also demands “of the self”. Moreover, the Formula of Universal Law tells us how these competing necessities ought to be structured, when it points out that our natures as rational and as autonomous are more fundamental than our particular subjectivities, both rationally and to our natures. Accordingly, this Formula says that the demands of reason and autonomy have priority over the claims of happiness. And what else should we have expected of the fundamental principle of autonomy, except that it should tell us to prioritize our own autonomy?

Moreover, the fact that these necessities can conflict with another, and the fact that we do not always structure them according to the Formula of Universal Law, can help us explain how we can deviate from the Formula of Universal Law without giving up our freedom, even though this Formula is the fundamental principle of freedom. If we are free—if we have the property of freedom—we are capable of prioritizing the demands of happiness over the demands of reason and autonomy, just because, as free, we are absolutely self-determining. But even if we transgress the demands of freedom, we do not thereby become unfree just *eo ipso*. For we can still remain subject to the demands that we transgressed, and, as free, we can still remain capable of adhering to those demands. In this way, freedom constitutes a norm for our agency, even if we consider freedom just its own terms, apart from issues about morality. Violating the norms of freedom then counts only as a misuse of freedom, rather than an abrogation of it, because we can perfectly well retain our full freedom throughout.

In a way, Kant’s picture of autonomy here is profoundly striking and unexpected, since it explains how absolutely self-determined actions actually derive from a thoroughgoing, but deeply structured kind of practical necessity. Proper freedom, in Kant’s view, is ultimately, and simply, a matter of how the necessities deriving from the competing parts of ourselves are structured relative to one another. Proper freedom is a matter of the *form* of these necessities in our volition, and of which have priority over the others: our wills are free when the necessities of autonomy and reason have priority over the necessities of happiness. Or, at least, this is what the proper use of our freedom looks like, according to Kant.

In another way, though, the general form of Kant's account here is clearly prefigured in the way that he poses the problem that his account is supposed to solve. Kant wants to explain how freedom can be grounded or lawful, so that free choices will not be arbitrary or lawless, but instead necessitated. Posing the problem that way, however, requires necessity to be the fundamental concept in the account of how freedom and necessity are to be unified with each other. If freedom were the fundamental concept, then there would be no space for any kind of ultimate necessity. (Or so Kant thinks, at least.) Our choices would always bottom out only in an ungrounded free choice, so any synthesis of freedom and necessity would be totally illusory. So perhaps we really ought to have expected an answer of this general form from the very outset of the investigation: if freedom is to be ultimately grounded and ultimately lawful, then freedom will have to turn out to be some kind of necessity, according to Kant's version of the problem.

In any case, though, we have now made good on the promissory note issued in Chapter Two, where we were able to pose Kant's problem of freedom and necessity, and say some things about how a principle could be valid or necessary for a free agent, but where we could not explain how to make that necessity compatible with freedom, as absolute self-determination. There, we wanted to know how the ultimate ground or cause of an action could be both the agent and something that the agent takes to require them to act a certain way. This chapter's account of the Formula of Universal Law as the fundamental principle of autonomy answers that question, because it explains how the necessity in the agent's ground can both derive from their self and conform to the demands of autonomy itself.

Thus, the Formula of Universal Law unifies the agent and their ground into the ultimate cause of their actions, and in a way that finally makes good on the solution suggested in Chapter Two. There, I said that Kant's view was going to be that free actions all had a certain form in common, by virtue of which they count as both free and grounded. Now we can see how the Formula of Universal Law stipulates and requires this form, when it tells us to act on maxims that we can will as universal laws. For our maxims will express what we believe our happiness to

require of us, at least when we do not take ourselves to be under a moral obligation. In turn, the Formula of Universal Law takes this maxim and asks whether it has the form of law, and the form of self-legislation. Following the Formula of Universal Law, then, takes the necessity in our pursuit of happiness and ensures that it has the form of autonomy. Maxims with this form, then, and the actions pursuant to them as well, will all count as intrinsically autonomous, and intrinsically free. (Even though maxims lacking this form need not thereby count as unfree.)

At the level of our maxims, then, the Formula of Universal Law works by checking for and requiring a certain form: it demands that the grounds, actions, and ends contained in our maxims relate to each other in a way such that we could will our maxims as universal laws. At the level of our pursuit of happiness in general, however, the Formula of Universal Law works by taking that entire domain of our agency and structuring it, shaping it, *forming* it according to the demands of autonomy. In that way, the Formula of Universal Law gives our overarching pursuit of happiness the form of autonomy, by giving it the form of universal self-legislation.⁹

So this is how the Formula of Universal Law turns out to be the fundamental principle of autonomy, for Kant, and hence the fundamental principle of freedom, in the proper sense. As the fundamental principle of freedom, the Formula of Universal Law will then turn out to be the fundamental practical principle for free agents. After all, if freedom involves absolute self-determination, then there can be no principle that would be more fundamental for a free agent, than the fundamental principle of freedom. Any other principle would represent, by definition, an “alien cause” determining the will, which could therefore not be a free will, after all.

Thus, the Formula of Universal Law stipulates how a will that is free, in the proper sense of freedom, ought always to use that freedom. We might wonder, though, why the “ought” here is the same categorical “ought” that the fundamental principle of morality is supposed to capture, express, and systematize. But a fundamental practical principle must be categorically binding; that just follows from its status as fundamental. Freedom is the fundamental property

⁹ Thanks to Barbara Herman for pressing me to be clearer here.

or feature of a free agent; it characterizes their agency in general, and its concept and principle constitute norms for all of their action and volition. Thus, there can be no further conditions that could limit these norms or take priority over them, except perhaps a more general categorical imperative. As such, for free agents, the norms of freedom must themselves be absolute, unconditional, and categorical in character, too.

So because the norms involved in the concept of freedom and in its fundamental principle are to be fundamental for free agency and free agents, these norms of freedom must also be categorical in character. A free will, then, is categorically obligated to the Formula of Universal law, since this is just the fundamental principle of their freedom. And—as it happens—Kant’s derivation in the Second Section was supposed to show that the Formula of Universal Law is also the fundamental principle of categorical obligation in general. Which means that the categorical obligation in the fundamental principle of freedom coincides with, and is actually identical to, categorical obligation just in general. Therefore, a free agent will also be subject to the fundamental principle of morality as well, because their fundamental principle makes them categorically obligated to morality in general.

Thus, when we put these pieces together, we get Kant’s Reciprocity Thesis: “a free will and a will under moral laws are one and the same”. This completes the first step in the Third Section’s deduction of the Formula of Universal Law, so we can now move on to the “idea of freedom” argument.

7.2 The Practical Point of View: Acting Under the Idea of Freedom

If Kant’s Reciprocity Thesis is correct, then, we find that the concepts of morality, freedom, and autonomy all imply one another, insofar as the Formula of Universal Law turns out to be the fundamental principle of each. Freedom and autonomy are to be equivalent concepts, the concept of categorical obligation is to imply that the Formula of Universal Law is the fundamental principle of morality, as per the derivation argument, and the concept of a will that is a law to itself is to imply that the Formula of Universal Law is the fundamental principle

of autonomy, as we just illustrated. These connections then create space for Kant to use the concept of freedom to argue that our wills must actually be subject to the Formula of Universal Law, and therefore subject to moral obligation. For if he can show that there is some sense in which our wills are free, then it will follow that there is a corresponding sense in which the Formula of Universal Law must be valid for our wills.

This is where the “Idea of Freedom” argument comes into the picture. Kant begins the argument by observing that

It is not enough that we ascribe freedom to our will on whatever ground, if we do not have sufficient ground for attributing it also to all rational beings. For, since morality serves as a law for us only as rational beings, it must also hold for all rational beings; and since it must be derived from the property of freedom, freedom must also be proved as a property of all rational beings; and it is not enough to demonstrate it from certain supposed experiences of human nature (though this is also absolutely impossible and it can be demonstrated only *a priori*), but it must be proved as belonging to the activity of all beings whatever that are rational and endowed with a will. (G 53, 4:447-448)¹⁰

Here, Kant is setting the parameters for the following argument: the Formula of Universal Law is to be the fundamental principle of morality, but Kant also thinks that morality “must also hold for all rational beings”. So the deduction needs to show that the Formula of Universal Law is valid for all rational beings—valid objectively. And if the deduction is to achieve that goal through the concept of freedom, then its argument has to be able to attribute the right kind of freedom to all rational beings whatsoever, just as such.

Along these lines, Kant continues:

¹⁰ It is a little hard to tell how strongly to take this last remark about proving freedom through experience. In the First Critique’s Canon of Pure Reason, Kant claimed that we experience ourselves as practically free, and this experience seemed to be enough to show there that we are subject to the moral law, at least in a practical respect (CR 675-676, A801-804/B829-832). But the experience described in the Canon does not seem to be an experience “of human nature”, but of our “activity” just as a being that is “rational and endowed with a will”, precisely as the *Groundwork* requires. For, as Kant puts it in the Canon, “We thus cognize practical freedom through experience, as one of the natural causes, namely a causality of reason in the determination of the will” (CR 676, A803/B821). See also my remarks in the conclusion of this chapter about sensation and *a priori* consciousness. Kant does, however, give up on the idea that morality “must be derived from the property of freedom”, at least by the Second Critique, since he gives a different deduction there, as I have mentioned already. That does not mean that Kant also gave up on the deduction in the *Groundwork*, only that he stopped viewing the *Groundwork*’s deduction as the only possible deduction for the Formula of Universal Law.

I say now: every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect, that is, all laws that are inseparably bound up with freedom hold for him just as if his will had been validly pronounced free also in itself and in theoretical philosophy. (G 53 4:448)

This is the crucial claim of the deduction, and it also articulates the basic strategy thereof. The idea of freedom is the basic concept here. For Kant, ideas are a kind of concept, and concepts attribute properties and kinds (CR 398-399, A319-320/B375-377). (The concept of redness attributes the property of redness, the concept of animals attributes the kind “animal”, etc.) So the idea of freedom attributes the property of freedom, and then this “idea of freedom” argument is ultimately making a point about beings that cannot act without attributing the property of freedom to themselves.

So, granted that the Formula of Universal Law is the fundamental principle of freedom, the idea of freedom is also committed to attributing the Formula of Universal Law, as a valid principle for whatever the idea of freedom is applied to. So if all rational beings must attribute freedom to themselves, then even that will be enough to show that there is some sense in which the Formula of Universal Law is valid for all rational beings. And then there will be some corresponding sense in which the Formula of Universal Law is objectively valid. If, then, rational beings must think of themselves as *fundamentally* free, it will follow that the Formula of Universal Law must be fundamentally, categorically valid for them, in some sense.

But before we move on to try to explain why Kant thinks that all rational beings must act only “*under the idea of freedom*”, we should stop to explain two things.¹¹ First, what does Kant mean in saying that “every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect”? What is this “practical respect”, which is supposed to articulate the sense in which the deduction is to show that we are free, and therefore also subject to the Formula of Universal Law? And, second, granted that the Formula of Universal Law is objectively valid in this practical respect, how would that show that this

¹¹ Thanks to Barbara Herman for pressing me to be clearer here.

Formula “holds for rational beings just as if their wills had been validly pronounced free also in itself and in theoretical philosophy”?

These are complicated subjects, which I can only touch on here. It is clear, though, that “freedom in a practical respect” cannot be the same as the practical freedom that Kant discusses elsewhere (CR 675-676, A801-804/B829-832; CP 79, 5:94). Practical freedom is the same as the *Groundwork’s* negative freedom, so the concept of practical freedom is inadequate to the strategy of the deduction, which requires the full concept of autonomy, as we have already seen. Accordingly, the idea of freedom must really refer to and attribute what Kant also calls “transcendental” freedom.

Thus, practical freedom is really a kind of freedom, and not a respect in which we are free or a point of view from which we represent or cognize our freedom. In fact, Kant is clear that the idea of freedom is to be an idea under which we act, and not one that we apply to ourselves “in itself” or as “in theoretical philosophy”. So we apply the idea of freedom to ourselves only in relation to our actions and in relation to our agency more generally; the idea of freedom is to be part of what is involved in representing ourselves as acting and as the agents of our actions. In short, if we can only act under the idea of freedom, Kant’s idea is that we cannot represent ourselves as acting or as agents, unless we also represent ourselves as free. So the “idea of freedom” argument really asserts that rational agents must think of themselves as transcendently free, but only insofar as their actions and agency is concerned—rational agents must think of themselves as transcendently free, but only “in a practical respect”.

Kant’s background idea here, I take it, is that representing, cognizing subjects can adopt different “standpoints” or “points of view” corresponding to the different kinds of representations or cognitions of which they are capable. Accordingly, these cognitive “points of view” will be partially constituted by these different kinds of cognitions themselves, and partially constituted by the perspective that the cognizing subject takes with regard to those different cognitions. Kant recognizes many kinds of cognitions, but how many, exactly, and how these

kinds are to be individuated are questions that we cannot answer here.¹² For present purposes, though, only two kinds of cognition are immediately relevant: theoretical and practical.¹³

According to Kant, theoretical cognition and its corresponding theoretical “standpoint” or “point of view” is concerned only with representing the nature of objects accurately. Thus, paradigmatic theoretical domains include physics and mathematics, and the theoretical point of view will be constituted in part by these cognitions themselves, and in part by the nature of the point of view or perspective that we adopt, as cognizing subjects, when we concern ourselves with investigating and discovering the nature of objects.

In turn, practical cognition and its point of view is also concerned with the proper representation of objects. However, the practical point of view must also consider these objects in relation to our volition and to our agency more broadly.¹⁴ (And “object” here includes ourselves and our own natures as agents. That will be crucial to Kant’s argument in the *Groundwork’s* Third Section.)

In the terms that we have been developing here, to consider an object in relation to our will is to consider it as a kind of end for us to achieve, a kind of action as a means to an end, or as a kind of ground for our volition. For example, I could consider something as a possible end for me to achieve, I could adopt something as the actual means to an end that I have already chosen, or I could recognize that something requires me to act a certain way.

Practical cognition, conceived in these terms, will overlap with and require theoretical cognition: for example, means/ends relations are predominantly matters for theoretical investigation and cognition (G 26, 4:415). However, the practical point of view also has an

¹² My sense, however, is that Kant thinks that cognitions are type-individuated by the sense in which they can be shown to be valid of their objects.

¹³ Transcendental cognition is invoked in the third phase of the argument, which we will discuss in the next section. However, as we will see there, Kant’s argument there does not depend on such cognition.

¹⁴ The *Critique of Practical Reason* does say that practical cognition is “cognition having to do only with determining grounds of the will” (CP 17, 5:20). However, given what a ground is, cognition of grounds for the will must also comprehend cognition of ends, means, and the relations between them also.

ineliminable, irreducible perspectival character to it, different from that of the theoretical point of view. For the practical point of view always represents objects in connection to the perspective of some agent, which may or may not actually be ourselves. As I said, the practical point of view is supposed to be the standpoint or the point of view of a cognizing *subject*, and not just of their cognitions, and here that subject is concerned with how an object relates to an agent's will. As we can already begin to see, this point about the subject's perspective is crucial to Kant's argument in the *Groundwork's* deduction of the Formula of Universal Law, because that argument hinges on the idea that we must attribute freedom to ourselves, from this practical point of view.

Accordingly, the practical point of view must take into account an entire range of relations and considerations that the theoretical point of view is not directly concerned with, even though the theoretical point of view may have conclusions and implications that need to be taken up into the practical point of view. For example, the conclusions about freedom reached from the theoretical point of view will be highly important to the practical point of view.

In any case, though, the current upshot: Kant's underlying thought here is that we might be committed to representing ourselves in a certain way from the practical point of view, even if we cannot determine whether or not we actually are that way, from the theoretical point of view. So we might have to represent ourselves as agents, even if the concept of agency carries presuppositions that we cannot determine the truth of, theoretically speaking. For example, suppose that the concept of agency presupposes that an immaterial substance or soul is capable of causal interaction with material bodies, and *vice versa*. We might not be able to make sense of how those kinds of interaction are possible, though, and our best available theoretical sciences, like physics and chemistry, might not be able to prove that anything immaterial even exists, let alone to explain how something immaterial could possibly interact with material bodies.

So we might not be able to determine whether or not this presupposition of agency is true, but we might still be committed to representing ourselves as agents, and hence as immaterial substances affecting the material world, nevertheless. In which case, Kant wants to say that these representations would still be valid, because if we are committed to representing ourselves a certain way, then we would make a rational mistake if we did not “live up to” that commitment. But if we cannot prove whether or not we are correct to represent ourselves that way, apart from that commitment, then those representations can only be valid within their own domain: if I am committed to representing my agency a certain way, but without being able to show that my agency really is that way, then that representation is valid, but only insofar as my agency is concerned. Only within the domain or point of view of my agency, in other words. Only “in a practical respect” or “from the practical point of view”. These representations will still have to govern my practical deliberations, primarily, but also my consciousness of myself as an agent and throughout my actions. But I would be mistaken if I tried to apply these representations beyond the domain of my agency.

That is to be Kant’s point about freedom: we have to represent ourselves as free agents, but we cannot determine whether or not we actually are free (G 59-60, 4:455). It would follow, then, that we would have to represent ourselves as subject to the fundamental principle of freedom, even though we cannot tell whether or not we are subject to this principle, beyond the scope of how we have to represent ourselves throughout the domain of our agency. That representation is fully valid, then, but only within its proper domain: only within the practical point of view.

Still, how should it be that if someone is “really free in a practical respect”, then “all laws that are inseparably bound up with freedom hold for them just as if their will had been validly pronounced free also in itself and in theoretical philosophy”? If the Formula of Universal Law is valid for a will in a practical respect, how is that “just as if” this Formula were shown to be valid for that same will “in itself and in theoretical philosophy”?

To start with, if theoretical philosophy could prove that the will is free, and hence that the Formula of Universal Law is objectively valid for the will in that sense, also, then it would follow trivially that the Formula of Universal Law is objectively valid for the will in a practical respect, too. We would still have to represent ourselves and our agency under the idea of freedom, and the Formula of Universal Law would still have to govern our practical deliberations, our actions, and our consciousness and self-consciousness as agents. So the practical consequences would be the same either way.

More than that, though, the character of the validity of the Formula of Universal Law would be the same in both cases: this Formula would be valid objectively and categorically in both cases, only the scope of this validity is different. In one case, this validity is limited to the practical point of view, while the validity in the other case also extends to the theoretical point of view.

These explanations are still fairly rough, but they are the best that we can do without really digging into the details of Kant's thinking about the differences between practical and theoretical cognition, along with their corresponding points of view. It is worth emphasizing, though, that Kant does not allow for the practical point of view to contradict the theoretical point of view, or to inject claims or cognitions into the theoretical point of view (G 60, 4:455-456). That is part of why the idea of freedom and the validity of the Formula of Universal Law are restricted to the practical point of view, even though that point of view is ultimately going to find itself required to posit them.

In any case, let's move on: why must all rational beings act only under the idea of freedom? On this score, Kant immediately goes on to say that

Now I assert that to every rational being having a will we must necessarily lend the idea of freedom also, under which alone he acts. For in such a being we think of a reason that is practical, that is, has causality with respect to its objects. Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgment not to his reason but to an impulse. Reason must regard itself as the author of its principles independently

of alien influences; consequently, as practical reason or as the will of a rational being it must be regarded of itself as free, that is, the will of such a being cannot be a will of his own except under the idea of freedom, and such a will must in a practical respect thus be attributed to every rational being. (G 53-54, 4:448)¹⁵

Here, Kant reverts back to an analysis of what reason is in general in order to explain why “we must necessarily lend the idea of freedom” “to every rational being having a will”: “one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgment not to his reason but to an impulse.”

“Consciously” is the crucial word here: reason’s judgments might be determined by “alien influences” *unconsciously* or *covertly*, but reason must always be conscious of itself “as the author of its principles independently of alien influences”. For if we were ever conscious that our rational judgments were being determined “from another quarter”—from outside of our reason itself—then we would never say that reason was the source of this judgment, but rather the external impulse instead. For we would be conscious that reason was not the author of this judgment, but the external impulse, whose determination we are actually conscious of, *ex hypothesi*. So, Kant concludes, we have to think of reason as the source or author of its judgments, or at least of the judgments that we can properly call “rational”. Otherwise we could be conscious of having our judgment determined by an external impulse and still view reason as the author of this judgment.

Of course, reason can certainly still base properly rational judgments on considerations from another “quarter”. Only these considerations cannot determine reason’s judgments by themselves, without an intervening act on the part of reason itself. Otherwise, a rational subject

¹⁵ Some heuristic hay has been made over Kant’s diction when he says that we must “lend” (*leihen*) the idea of freedom to every rational being with a will. But I think “lending” the idea of freedom means nothing more than applying this concept to all rational beings, even though the First Critique is supposed to show that we cannot produce a theoretical proof that would demonstrate that such beings really are free. The whole point of the argument in the *Groundwork*’s Third Section is to justify applying the concept of freedom to ourselves, despite the impossibility of a theoretical proof that the concept of freedom is objectively valid for us. So, to keep to the metaphor, the point is to show that we can be justified in “lending” to all rational beings a property that we cannot truly “give” to them.

could not view their reason as the source or author of the judgment; instead, they “would then attribute the determination” of their judgment to an impulse.

Now, Kant’s explanation here might appear to violate the stricture he adopted previously, when he insisted that “it is not enough to demonstrate” freedom “from certain supposed experiences of human nature”. For isn’t he now adverting to how we experience our judgments, in order to say that we can only be conscious of reason as the source of its judgments? Perhaps, but experience (*Erfahrung*) is a technical concept for Kant, which properly refers to a certain kind of cognition with an element of passivity or receptivity, as when we cognize an object through how our senses are affected (CR 255, B147; 155, A19-20/B34). A judgment, however, is something that we do; it is an activity of reason, and so our cognition of it can instead be based in our immediate consciousness of this activity (G 56, 4:451), and not in something properly called an “experience”, which can be conscious, yes, but is always mediated through how our faculties are affected. Accordingly, it is crucial that Kant only says that we cannot imagine a reason that would “consciously” (*Bewusstsein*) receive direction from outside.

But even if Kant is referring to an experience of our rational judgments, or to a judgment or attribution about them based in an experience, this experience would still not be an experience “of human nature”, but rather of the activity of our reason just in general. So this assessment that reason must be viewed as the source or author of its judgments would still be based only on “the activity of all beings whatever that are rational and endowed with a will”, as Kant previously required of the deduction of the Formula of Universal Law.

So, to return to the argument, remember that the will is a faculty of reason and of rational judgments, because it is the faculty for deriving actions from the representation of laws (G 24, 4:412). To derive an action from the representation of a law, we must apply the concepts involved in the representation of the law to a particular action in the empirical world. That is just to judge an action according to the representation of the law. So if reason in general has to be thought of as the source or author of its judgments, and the will is a faculty of practical reason

and practical judgment, then it will follow that we will have to think of the will as the source or author of its practical judgments: of how it derives actions from the representations of laws. And the will is supposed to be a causal faculty as well: specifically, a faculty for causing actions according to its practical judgments. So, by parity of reasoning, we must consider the will as the source or author of its acts of volition, and of the “external” actions that result from these as well. This is all just part of what it is to think of the will as a *practical* reason.

But to think of the will as the source or author of its judgments, volitions, and actions is just to think of it as autonomous. Broadly speaking, these are all actions of the will, which are to be derived from the representations of laws. So if the will is also to be the source of these actions, we will have to view the will as giving these laws to itself. Otherwise, we would have to think of the laws as the source of these actions, and we could only think of the will as executing these laws. We will therefore have to regard the will as autonomous, and therefore as free in the proper sense.

If Kant’s analysis here is correct, then, every rational agent must be thought of as free, or as autonomous, for this is just part of their concepts as rational agents. We must, in short, “lend” them the idea of freedom, under which alone they can act. To put the point a little differently, if we do not view the will as the source or author of its actions, then we do not view the will itself as acting. Instead, we view it as merely the instrument of whatever external impulses determine it to action, or of whatever laws are valid for it. Thus, the will can only act under the idea of freedom. For if we do not think of the will as free, then it cannot be the will that acts, but something else, something external to the will.

To think of the will as free, though, is also to think of the will as subject to the Formula of Universal Law, just because this Formula is the fundamental principle of freedom, as the last section illustrated. From here, though, the Reciprocity Thesis then shows immediately that we cannot think of a will except as subject to moral obligation: a will must be thought of as free, but a “free will and a will under moral laws are one and the same” (G 53, 4:447). And if we cannot

think of a will except as subject to the Formula of Universal Law, then it follows that the Formula of Universal Law is objectively valid for the will, albeit only in a practical respect. For this is how the agency of the will must be represented, and how the will must be conscious of its agency, even though we cannot determine whether its agency is actually as it must be represented to be.

7.3 A Circle in the Deduction?

At this point, Kant concludes that

We have finally traced the determinate concept of morality back to the idea of freedom; but we could not even prove the latter as something real in ourselves and human nature; we saw only that we must presuppose it if we want to think of a being as rational and endowed with consciousness of his causality with respect to actions, that is, with a will, and so we find that on just the same grounds we must assign to every being endowed with reason and will this property of determining himself to action under the idea of freedom. (G 54, 4:448-449)

Yet Kant does not think this is a totally satisfying deduction for the Formula of Universal Law. Not yet, anyway. Interestingly, in fact, Kant presents the section containing the “idea of freedom” argument as “further preparation” necessary for “the deduction of the concept of freedom from pure practical reason, and with it the possibility of a categorical imperative as well” (G 53, 4:447). So there is still an important gap remaining in the argument.

Specifically: all that has been shown so far is that rational agents must think of themselves as free, and hence as subject to the Formula of Universal Law. From that, though, it only follows that *if* “we want to think of a being as rational and endowed with consciousness of his causality with respect to actions, that is, with a will,” *then* we must think of that being as subject to the Formula of Universal Law. But that is at best only a general truth about rational agents: Kant has not shown that there are any such agents, still less that we human beings are rational agents of this kind. Kant has not even shown that we human beings must think of ourselves as this kind of rational agent.

For example: for all that has been said so far, what if it occurred to some moral skeptic to doubt that we humans have to think of ourselves as rational agents of the right kind? This

skeptic might be prepared to grant the account of rational agency so far, and so they might agree that rational agents have to think of themselves as free, and therefore as subject to the Formula of Universal Law as well. A skeptic of this kind might even be ready to concede that many humans are in fact “endowed with consciousness of [their] causality with respect to actions, that is, with a will”.¹⁶ Furthermore, this skeptic could also accept that many more of us are able to acquire such consciousness, with the benefit of a little bit of direction, at least. This skeptic might dispute that this consciousness is ever accurate, however. But they could just as well grant that this consciousness can be accurate, and they could even concede that the accuracy of this consciousness is irrelevant: for they could grant that this consciousness is enough by itself to make the Formula of Universal Law binding on those who possess it, or can possess it, albeit binding only in a practical respect.

For all that, though, this moral skeptic might still doubt that *every* human being is a rational being of this kind, or that we all *must* think of ourselves as rational beings of this kind, or that we all *must* be able to possess this kind of consciousness of our own wills. In that case, for all Kant has shown so far, there could still be human beings that do not have to think of themselves under the idea of freedom. Kant’s deduction, though, is supposed to show that the Formula of Universal Law is objectively valid, and so valid for all human beings. This being so, his argument has not yet reached its conclusion.

Kant does sometimes assert that “All human beings think of themselves as having free will.” (G 59, 4:455) As a result, defending our right to maintain this natural self-conception is a large component of the project of the deduction (G 63, 4:459).¹⁷ And if we grant to Kant that all human beings really do think of themselves as free, then the argument in the deduction so far

¹⁶ Kant often speaks of this kind of consciousness throughout the penultimate part of the *Groundwork’s* Third Section, which is called “Of the Extreme Boundary of all Practical Philosophy”. But there again, this consciousness is insufficient for the deduction unless Kant can show that every human is at least capable of it, which is why the discussion there also emphasizes the considerations that we are about to discuss. See especially (G 61, 4:457), which neatly summarizes the lacking argument.

¹⁷ Thanks to Barbara Herman for suggesting I say more here.

would, in fact, be sufficient for its conclusion. Or close to sufficient, in any case: we would really need to grant that all human beings *must* think of themselves as free, in order for the Reciprocity Thesis to show that freedom and morality are objectively valid for us. The mere fact that we all actually happen to think of ourselves as free is not sufficient for the deduction of the Formula of Universal Law, because that happenstance would not guarantee that no exceptions are possible to this self-conception as free, nor would it preclude our idiosyncratic subjectivities from happening to converge on that self-conception.

But if we grant to Kant that all human beings must think of themselves as free, then the Reciprocity Thesis would already be enough to guarantee that we must think of ourselves as subject to the Formula of Universal Law, considered as the fundamental principle of morality. “If, therefore, freedom of the will is presupposed, morality together with its principle follows from it by mere analysis of its concept” (G 53, 4:447), as Kant puts it towards the start of the *Groundwork’s* Third Section. In that case, then, this Formula would be objectively valid for us, albeit still only “in a practical respect” or “from the practical point of view”.

However, if it occurs to someone to doubt that all human beings must think of themselves as free, or as rational beings of the right kind, the argument so far will have nothing to say to them. Once that doubt is finally answered, then “nothing is left but *defense*, that is, to repel the objections of those who pretend to have seen deeper into the essence of things” (G 63, 4:459). Until that point, though, the deduction requires more argument, but Kant also seeks to provide it.

Thus, even with the Reciprocity Thesis and the “idea of freedom” argument in hand, Kant still thinks that he needs to answer questions like “why, then, ought I to subject myself to this principle and do so simply as a rational being, thus also subjecting to it all other beings endowed with reason?” (G 54, 4:449) And this question is especially pressing, he thinks, since we are imperfectly rational beings, as our moral skeptic is sure to emphasize. As Kant puts it, when I ask after why “ought I to subject myself” to the Formula of Universal Law,

this “ought” is strictly speaking a “will” that holds for every rational being under the condition that reason in him is practical without hindrance; but for beings like us—who are also affected by sensibility, by incentives of a different kind, and in whose case that which reason by itself would do is not always done—that necessity of action is called only an “ought,” and the subjective necessity is distinguished from the objective. (G 54-55, 4:449)

In short, the argument so far applies to all rational beings in general, so it applies to us, but only insofar as we are rational. Because we are imperfectly rational, then, we need to do a little more work to determine how far the argument so far has taken us, so far as us humans are concerned. And until we can answer this question about the Formula of Universal Law satisfactorily, “we should have got no further with respect to its validity and the practical necessity of subjecting oneself to it” (G 55, 4:449). For “we cannot yet see” that we ought “to regard ourselves as free in acting and so to hold ourselves yet subject to certain laws”; “we cannot yet see how this is possible, and hence *on what grounds the moral law is binding.*” (G 55, 4:450)

So a moral skeptic might still want to know how Kant can show that every human being has to think of themselves as a rational being of the kind that he has described. In Kant’s mind, the underlying issue with the deduction so far, which prevents it from already answering such a skeptic is that

It must be freely admitted that a kind of circle comes to light here from which, as it seems, there is no way to escape. We take ourselves as free in the order of efficient causes in order to think ourselves under moral laws in the order of ends; and we afterwards think ourselves as subject to these laws because we have ascribed to ourselves freedom of the will; for, freedom and the will’s own lawgiving are both autonomy and hence reciprocal concepts and for this very reason one cannot be used to explain the other or to furnish a ground for it but can at most be used only for the logical purpose of reducing apparently different representations of the same object to one single concept (as different fractions of equal value are reduced to their lowest expression). (G 55-56, 4:450)

The problem here is by no means perspicuous, especially because it is all too easy to think that the “kind of circle” that Kant has in mind here is a circular *argument*. Kant reinforces this impression a little later, when, after resolving this problem, he explicitly mentions the “*petitio principii*” fallacy (G 58, 4:453), which is just another name for begging the question or arguing in a circle. But the ultimate problem is not really about a circle in the argument that constitutes

the deduction, but instead in the concepts that it relies upon: “for, freedom and the will’s own lawgiving are both autonomy and *hence reciprocal concepts* and for this very reason one cannot be used to explain the other or to furnish a ground for it but can at most be used only for the logical purpose of reducing apparently different representations of the same object to one single concept” (Emphasis mine).

Put this way, we can see that Kant is cycling back to challenge the deduction at its very foundation: the Reciprocity Thesis. The reciprocal implications between freedom, autonomy, and morality were supposed to allow us to deduct the Formula of Universal Law from the concept of freedom. But now the worry is that this reciprocity actually makes such a deduction impossible.

Because freedom and morality are reciprocal concepts, we need to prove the objective validity of one of them in order to use it to show the objective validity of the other. Otherwise we can only use the concept of the one to clarify the logic of the other. But the argument so far has not met this burden: all we have done in the deduction so far is analyze the concept of a rational being in order to show that practical rationality presupposes the idea of freedom and the objective validity of the Formula of Universal Law. We have not actually shown that any of these concepts are objectively valid for us, even in a merely practical respect. And this is just as our moral skeptic would allege: consciousness of having a will might show that the concept of freedom is valid for whichever individuals have that consciousness. And anyone who has that consciousness might therefore be subject to the Formula of Universal Law, too, just as the Reciprocity Thesis would have it. But this does not yet show that the concept of freedom is *objectively* valid, and so it does not show that the Formula of Universal Law is objectively valid either.

Hence, the ultimate problem is that the deduction so far is purely *analytic*; that is why it cannot yet answer our moral skeptic. But Kant was always clear that analysis would not be

enough for a deduction of the Formula of Universal Law: even before beginning the main argument of the deduction, he insists that if

freedom of the will is presupposed, morality together with its principle follows from it by mere analysis of its concept. But the principle of morality—that an absolutely good will is that whose maxim can always contain itself regarded as a universal law—is nevertheless always a synthetic proposition; for, by analysis of the concept of an absolutely good will that property of its maxim cannot be discovered. (G 53, 4:447)¹⁸

So how can we show such a synthetic proposition?

¹⁸ This remark is obviously essential to Kant's whole project in the *Groundwork*, but it might also seem really rather baffling at first, particularly given the method explicitly adopted in the Second Section. There, remember, the first task was to "follow and present distinctly the practical faculty of reason, from its general rules of determination to the point where the concept of duty arises from it." (G 24, 4:412) From this concept of duty, we then get the concept of a categorical imperative, and from there we are supposed to be able to derive the Formula of Universal Law, as the fundamental categorical imperative. But all imperatives, categorical or otherwise, "say that to do or to omit something would be good" (G 24, 4:413), so that a specifically categorical imperative represents "an action as *in itself* good" (G 25, 4:414). It then follows that obeying the Formula of Universal Law makes the will absolutely good. So it looks for all the world as though the Second Section does give us an analysis of an absolutely good will wherein the Formula of Universal Law is its principle. For if we analyze the concept of an absolutely good will, we find that its principle has to be a categorical imperative, or that its principle must at least support a categorical imperative. After all, only categorical imperatives can represent actions as good in themselves, not as good only relative to an end, and therefore as good absolutely. And the derivation reaches the Formula of Universal Law through an analysis of the concept of a categorical imperative. So why isn't this "principle of morality" really analytic? I think that it is analytic, but only in a way. Analysis of the concept of a *perfectly* rational will does yield "the principle of morality": for perfectly rational beings, an absolutely good will is one under the Formula of Universal Law, and we can show this by analysis of the relevant concepts, along the lines of the Second Section. But, strictly speaking, this principle is not an imperative for them. It is an imperative for imperfectly rational beings like us, but no analysis can show that the Formula of Universal Law is an imperative. That "the above principle of autonomy is the sole principle of morals can well be shown by mere analysis of the concepts of morality" (G 47, 4:440)—so an absolutely good will must be one under the Formula of Universal Law. And this analysis even shows that the fundamental principle of morality "must be a categorical imperative" (G 47, 4:440). But this point is only about what the rule would have to be for us, if it is valid for us at all. "That this practical rule [actually] is an imperative, that is, that the will of every rational being is [actually] bound to it as a condition, cannot be proved by mere analysis of the concepts to be found in it, because it is a synthetic proposition" (G 47, 4:440). For imperfectly rational beings like us, Kant explains the syntheticity of "this categorical imperative or law of morality" (G 30, 4:420) by saying that "I connect the deed with the will, without a presupposed condition from any inclination, a priori and hence necessarily (though only objectively, i.e. under the idea of a reason having complete control over all subjective motives). This is, therefore, a practical proposition that does not derive the volition of an action analytically from another volition already presupposed (for we have no such perfect will), but connects it immediately with the concept of the will of a rational being that is not contained in it." (G 30n, 4:420n) In short, we cannot show that we are obligated to follow the moral law by analyzing our inclinations or by analyzing anything that we already will, for that method could not show that the moral law is objectively valid for us. Couldn't someone lack that inclination or volition? And analyzing "the idea of a reason having complete control over all subjective motives" is no help, either, since that analysis cannot show that we are in possession of such a faculty of reason. So the Formula of Universal Law must be synthetic for us.

Thus, the problem here is parallel to the problem that Kant confronts in the Transcendental Deduction of the Categories: how can we show that one of the concepts in question is objectively valid? More specifically, how can we show—and show *a priori*, as per the parameters Kant sets at the beginning of the deduction—that one of these concepts is objectively valid for us imperfectly rational agents? If we can do this for even one concept in this “circle”, then the analytic connections with the others will give us the deduction we are seeking. But it is hard to see right away how the argument should proceed.

The strategy of the deduction, though, is to begin with the concept of freedom. So to make this strategy work, Kant needs to be able to give us a sense in which the concept of freedom is objectively valid for us. The difficulty here looks rather dire, however, given how the *Critique of Pure Reason* is supposed to show that we cannot give a theoretical proof of our freedom (CR 116, Bxxviii). Of course, the First Critique is also supposed to show that we cannot prove that we are not free either, theoretically speaking. But be that as it may, how, then, can Kant use the concept of freedom to try to show that we are subject to the moral law—that the concept of morality and its fundamental principle is objectively valid for us—when we cannot show that we are free, from the point of view of theoretical cognition?

Since theoretical cognition has nothing to say about whether or not we are free, it would seem as though the practical point of view would be on its own, so far as the deduction of the Formula of Universal Law is concerned. In other words, it would seem as though the only way to show the objective validity of the Formula of Universal Law would be to find something in our thinking about ourselves as agents, or in our consciousness of ourselves as agents, which would prove that we must think of ourselves “under the idea of freedom”. But that approach runs us straight into the teeth of our moral skeptic. So how can Kant’s argument work its way out of this position?

Basically, Kant’s strategy is to show that the deduction does not need to rely solely on the practical point of view. Instead, he argues that there is a feature of common human

consciousness, which implies that we must think of ourselves as free, albeit only “in a practical respect”. Accordingly, the concept of freedom must be objectively valid for us, albeit only from a practical point of view. And then the Reciprocity Thesis will guarantee that the Formula of Universal Law is also objectively valid for us, although again only from this practical point of view.

Here is how Kant’s argument goes:

No subtle reflection is required to make the following remark, and one may assume that the commonest understanding can make it, though in its own way, by an obscure discrimination of judgment which it calls feeling: that all representations which come to us involuntarily (as do those of the senses) enable us to cognize objects only as they affect us and we remain ignorant of what they may be in themselves so that, as regards representations of this kind, even with the most strenuous attentiveness and distinctness that the understanding can ever bring to them we can achieve only cognition of *appearances*, never of *things in themselves*. (G 56, 4:451)¹⁹

Kant’s point here is that if we reflect a little on our own minds, we can become conscious of how certain representations, like those “of the senses”, “come to us involuntarily”. Immediately after the passage just quoted, Kant describes how we can “notice” a difference “between representations given us from somewhere else and in which we are passive, and those that we produce simply from ourselves and in which we show our activity” (G 56, 4:451). Consciousness of this difference between our activity and our passivity is then supposed to allow

¹⁹ To my mind, remarks like this one are ultimately among the best evidence for the metaphysical two worlds reading of Kant’s transcendental idealism that I tend to favor. Here, Kant is saying that when we reflect on our cognition, and especially on our theoretical cognition, we can become conscious that all of our empirical representations come to us involuntarily, though our own activity is still necessary to constitute their form. Abstruse philosophy is not supposed to be necessary to notice this, only consciousness of how our faculties are affected in every representation that originates in our senses. From this, we are supposed to become conscious of how we could never show that theoretical cognition is valid for objects apart from how they affect us. Thus, reflection from this point of view—the transcendental point of view—does reveal to us an epistemic limitation, as interpretations like Allison’s (2004) rightly emphasize. And this transcendental reflection is also supposed to help us to illuminate the conditions for our theoretical knowledge. But the fact that transcendental reflection involves and directs our consciousness shows that arguments, like those in the First Critique’s Transcendental Aesthetic, cannot be only abductive, inference-to-the-only-possible-explanation kind of arguments, as readings like Allison’s tend to end up supposing. Allison contends that those arguments are best understood as only limning the conditions of our knowledge. But that reading overlooks the fact that these arguments, though they are to be deductively certain, are also to be backed up by consciousness of how our minds work. For this consciousness carries ontological commitment, as when my consciousness of how I am affected is also consciousness that there is something affecting me, which I then know is there, though I cannot ever cognize what it is in itself.

us to become conscious that we can “cognize objects only as they affect us”, so that “we remain ignorant of what they may be in themselves”.²⁰ This consciousness may only come to the common understanding through “an obscure discrimination of judgment which it calls feeling”. But the important point is that we are still led thereby to consciousness of the distinction between “*appearances*” and “*things in themselves*”.

The need to draw this distinction between appearances and things in themselves is one of the headline conclusions of the *Critique of Pure Reason*, of course. But Kant also emphasizes that “No subtle reflection is required to make” it (G 56, 4:450). In fact, he goes on to say that

A reflective human being must come to a conclusion of this kind about all the things that present themselves to him; presumably it is also to be found even in the most common understanding, which, as is well known, is very much inclined to expect behind the objects of the senses something else invisible and active of itself—but it spoils this again by quickly making the invisible something sensible in turn, that is, wanting to make it an object of intuition, so that it does not thereby become any the wiser. (G 57, 4:451-452)

And, in fact, if the deduction of the Formula of Universal Law is going to rely on consciousness of the difference between our activity and passivity, then it will be crucial for this consciousness to be available to the common understanding. Otherwise, it would be quite hard to see how this consciousness could ever show the Formula of Universal Law to be objectively valid. In fact, relying on the arguments in the *Critique of Pure Reason* to produce this consciousness might be rather counterproductive. Instead, since every human being relies on their senses, Kant can think that we are all, on some level, already conscious of this difference between our activity and passivity. As such, if this consciousness commits us to thinking of ourselves in a certain way, it will likewise be capable of showing how that commitment is objectively necessary for us.

Now, we cannot try to determine whether or not Kant is right to say that consciousness of the difference between our activity and passivity leads us to consciousness of the difference

²⁰ This distinction between our activity and our passivity is also Kant’s famous distinction between spontaneity and receptivity. Both of the latter concepts show up in the *Groundwork*’s deduction (G 56-57, 4:451-452), but the vocabulary of activity and passivity is enough for our purposes, especially when spontaneity is also captured in the idea of absolute self-determination, as I explained in Chapter Two. Kant confirms that connection by relating spontaneity and “self-activity” (G 57, 4:452).

between appearances and things in themselves. Everyone will have to reflect on their own minds in order to determine for themselves whether or not they can find that distinction for themselves. (Or consult the arguments in the *Critique of Pure Reason*, for another way into the distinction.) Even psychology and cognitive science may have nothing to say here, in Kant's view. For the kind of reflection that he describes and which he asks of us here is supposed to require us to take a different standpoint towards ourselves and towards our minds than the standpoint of psychology and cognitive science.²¹ So we will simply have to concede this consciousness to Kant, if we are to understand the rest of his deduction for the Formula of Universal Law.²²

For once we are led to make the general distinction between appearances and things in themselves, Kant thinks that we then must apply that distinction to ourselves as well.

Even as to himself, the human being cannot claim to cognize what he is in himself through the cognizance he has by inner sensation. For, since he does not as it were create himself and does not get his concept a priori but empirically, it is natural that he can obtain information even about himself only through inner sense and so only through the appearance of his nature and the way in which his consciousness is affected—although beyond this constitution of his own subject, made up of nothing but appearances, he must necessarily assume something else lying at their basis, namely his ego as it may be constituted in itself; and thus as regards mere perception and receptivity to sensations he must count himself as belonging to the *world of sense*, but with regard to what there may be of pure activity in him (what reaches consciousness immediately and not through affection of the senses) he must count himself as belonging to the *intellectual world*, of which however he has no further cognizance. (G 56, 4:451)

Now, the distinction that Kant makes here between different ideas of ourselves might be a little unintuitive, and his notion of “inner sense” and “inner sensation” is quite technical (CR 157, A22-23/B37; 257-259, B152-156). But Kant's point here does not hinge on any of these complexities, any more than it did previously. Instead, Kant's point is essentially the same as

²¹ Kant is asking us to take the transcendental point of view, which “is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible *a priori*” (CR 149, B25), or, as the First Edition has it, is “occupied not so much with objects but rather with our *a priori* concepts of objects in general” (CR 133, A11/12). By contrast, psychology and cognitive science take the theoretical point of view.

²² But notice how the “fact of reason” deduction in the *Critique of Practical Reason* makes the same kind of move: “Consciousness of this fundamental law may be called a fact of reason” (CP 28, 5:31).

before: notice how you experience pain “involuntarily”, or how feelings, emotions, and moods arise in your mind “of their own accord”, in the same kind of way that your perceptions of external objects arise. We are no more conscious of ourselves as producing our feelings than we are of ourselves as producing anything in the world outside of ourselves. So here again, Kant thinks, we can be conscious of our passivity, and thus of “the appearance of our nature and the way in which our consciousness is affected”.

By contrast, we can also be conscious of “what there may be of pure activity in” us, which “reaches consciousness immediately and not through affection of the senses”. So, as before, we “must necessarily assume something else lying at” the basis of how we appear to ourselves, namely “our ego as it may be constituted in itself”. In short, we must distinguish ourselves as appearances from ourselves as things in themselves, just as we did before. And, also as before, this distinction is supposed to be already present in, and available, to, common understanding.

With that in mind, Kant now wants to go on to show how consciousness of ourselves as *noumena*, as things in themselves, commits us to thinking of ourselves as free. Hence, we will be committed to thinking of ourselves as subject to the Formula of Universal Law in turn, even if this consciousness does not prove that we really are free. For, Kant thinks, consciousness of the difference between our phenomenal and noumenal selves has implications for how we have to think of ourselves from the practical point of view.

As Kant puts it,

Because of this a rational being must regard himself as *intelligence* (hence not from the side of his lower powers) as belonging not to the world of sense but to the world of understanding; hence he has two standpoints from which he can regard himself and cognize laws for the use of his powers and consequently for all his actions; *first*, insofar as he belongs to the world of sense, under laws of nature (heteronomy); *second*, as belonging to the intelligible world, under laws which, being independent of nature, are not empirical but grounded merely in reason. (G 57, 4:452)²³

²³ “Cognizing laws for the use of our powers” should not be read purely descriptively or theoretically, under the model of cognizing laws of physics, for example, at least or especially not when we consider ourselves “as belonging to the intelligible world”. From that point of view, at least, cognizing laws for the use of our powers is inseparable from cognizing incentives, and from cognizing the moral law as itself a

In thinking of ourselves as *noumena*, and as belonging to the world of understanding, we place ourselves outside of nature and its laws, Kant says. Therefore, we must think of ourselves as free in the negative sense. But we are still conscious of our activity from this point of view, so we must still think of ourselves as subject to laws nevertheless. After all, the concept of causality still carries with it the concept of laws. So we must think of our activity as *noumena* as autonomous—free in the positive sense—and therefore as subject to all of the laws of freedom. And, of course, the Formula of Universal Law expresses the demands of the fundamental law of freedom.

However, even granted that we must think of ourselves as *noumena*, and even granted that we must therefore also think of ourselves as free and autonomous, why must we think of ourselves as autonomous with regard to our actions, and not just with regard to our judgment and understanding? Kant's examples so far—sense perception and “inner sense”—all pertain to our theoretical cognition, either of objects outside of us or of our own nature. So isn't there still room for our moral skeptic to agree that we must think of ourselves as the autonomous authors

sufficient incentive for our wills (CP 62ff, 5:71ff). Also, I tried to explain at least the general contours of how I understand Kant's talk of “standpoints” earlier, but much has been made of this language in his commentary. There is some tendency to read “standpoint” in a deflationary sense, as in “only a standpoint” that we are constrained to take for one reason or another. Read this way, claims that are valid from one standpoint need not be taken as carrying ontological or theoretical commitment beyond that standpoint, and perhaps not even within that very standpoint, either. For example, the transcendental standpoint distinguishes appearances from things in themselves, but if this is “only a standpoint”, then perhaps this distinction need not carry any ontological commitment, either within this point of view or beyond it. Perhaps, for example, the transcendental standpoint merely posits *noumena* because it requires them to explain something, while remaining neutral about whether or not they really exist, and neutral about which standpoints need to recognize *noumena*. Certainly some of the standpoints that Kant discusses are like this. The practical standpoint, for example, has some theoretical commitments that cannot be carried over into the theoretical standpoint, and the theoretical standpoint leaves some things open that can be filled in from the practical standpoint, albeit only within that standpoint. Transcendental freedom, for example. And the whole noumenal standpoint involved in the idea of freedom is merely a standpoint for practical reason (G 62, 4:458), since we cannot determine whether this standpoint is objectively valid from any other point of view, except perhaps in the broadest outlines. At the same time, however, some parts of these standpoints are clearly not “just standpoints”. The other standpoints cannot, for example, contradict the theoretical standpoint, at least not without creating an irresolvable conflict or antinomy within reason itself. Likewise for the transcendental standpoint, I should think. So it would be too quick to say, for example, that the noumenal standpoint is “just a standpoint”, because the status of the claims made from that point of view is not settled by this terminology.

of our theoretical judgments, while still insisting that our actions are totally determined by our desires and inclinations, and thus by natural causation and laws of nature more generally?²⁴

Not from the practical point of view, according to Kant. Consider again those desires and inclinations, which we are to suppose wholly determine our actions. We can be conscious of our passivity with regard to those desires and inclinations, in precisely the same way that we can be conscious of our passivity with regard to anything else that we sense or feel. As a result, Kant will think that we must be able to be conscious of ourselves as *noumena*, existing apart from and “behind” our desires and inclinations, so to speak, in the same way that we must think of ourselves as *noumena* “behind” our sensations, our sense perceptions, and any other modification of our “inner sense”. That means, though, that we must think of ourselves as free from the determination of our desires and inclinations, and hence, by parity of previous reasoning, as autonomous with regard to our actions.²⁵

Thus, it turns out that we must think of the Formula of Universal Law as our fundamental practical principle. And so Kant concludes that

As a rational being, and thus as a being belonging to the intelligible world, the human being can never think of the causality of his own will otherwise than under the idea of freedom; for, independence from the determining causes of the

²⁴ Thanks to Barbara Herman for reminding me of this important question.

²⁵ As Barbara Herman also reminded me, there is an important footnote in the *Religion*, where Kant asserts the possibility of a rational being who “might still need certain incentives, coming to him from the objects of inclination, to determine his power of choice.” (R 51n, 6:26n) This footnote also finds Kant going on to say that if the moral law were not “given to us” as an incentive “from within, no amount of subtle reasoning on our part would produce it or win our power of choice over to it.” For “this law is the only law that makes us conscious of the independence of our power of choice from determination by all other incentives (of our freedom)”. My interpretation of the “idea of freedom” deduction certainly does not try to lead us to the moral law within us, at least not in the sense of this footnote. Finding yourself committed to the objective validity of the Formula of Universal Law is not the same as becoming conscious of the moral law as an incentive for you, at least not in quite this footnote’s sense. For the *Groundwork*’s deduction has no immediate consciousness of being subject to moral or categorical obligation, as there is in the Second Critique’s “fact of reason” deduction. However, as I have reconstructed the argument in the Third Section of the *Groundwork*, it does not seek to make us *conscious* of our freedom, either. Instead, it seeks to make us conscious only of the difference between ourselves as phenomena and ourselves as *noumena*, which difference is then supposed to require us to *think* of ourselves as free, from the practical point of view. I have also emphasized how this argument is not supposed to depend on any “subtle reasoning”, but only on considerations available to common human consciousness and understanding. In any case, my concluding remarks on the possible differences between Kant’s two deductions of the Formula of Universal Law will have some relevance here, too.

world of sense (which reason must always ascribe to itself) is freedom. With the idea of freedom the concept of *autonomy* is now inseparably combined, and with the concept of autonomy the universal principle of morality, which in idea is the ground of all actions of *rational beings*, just as the law of nature is the ground of all appearances. (G 57, 4:452-453)

So, to summarize: to finish the deduction, Kant needed to show that the concept of freedom is objectively valid for us, at least in some sense, for only then can the Reciprocity Thesis show that the Formula of Universal Law is objectively valid for us as well. And his argument relies on the distinction between our phenomenal and noumenal selves, which Kant claims is implicit even in common understanding and in common human consciousness. Kant's argument does not use moral or practical considerations to show that we must think of ourselves as *noumena*. Instead, common human consciousness of the difference between our activity and our passivity is supposed to make us conscious of ourselves as *noumena*, which is supposed to commit us to thinking of ourselves as beings who are both independent of natural law and positively autonomous. This consciousness does not prove that we really are autonomous, but it is supposed to show that we are still committed to our own autonomy nevertheless. So this consciousness commits us to a certain self-conception, which must then govern our thinking about our agency. We must act under the idea of freedom, but only in a practical respect, because we still cannot prove that we are actually free.

Now, to clarify the argument a bit: at the beginning of his answer to the circle problem, Kant explains that we are going to try "to inquire whether we do not take a different standpoint when by means of freedom we think of ourselves as causes efficient a priori than when we represent ourselves in terms of our actions as effects that we see before our eyes." (G 56, 4:451) This is, of course, part of the argument, and Kant often speaks this way in the last parts of the Third Section. But there must still be more to the argument in the deduction. For asking after the standpoint involved in the idea of freedom is just more analysis of the concept of freedom, which cannot resolve the analytic circle by itself. Instead, we need an independent reason to

think that we have to take the standpoint involved in the idea of freedom, and we have already illustrated how Kant's actual strategy for the argument tries to meet that burden.

Next, the argument in the deduction is supposed to be an *a priori* proof that the Formula of Universal Law is objectively valid—*a priori* was one of the parameters Kant set for himself at the outset of the deduction, after all, and this parameter was supposed to be necessary in order to guarantee that the argument would apply to all rational beings, or at least to all human beings (G 53, 4:447-448). But we have shown that Kant's argument relies on consciousness of the difference between our activity and our passivity. Doesn't that consciousness derive from experience, or at least from the senses? Don't we need to sense something in order to become conscious of our passivity in sensation? So under my telling, hasn't Kant really given an *a posteriori* argument, and thereby violated his own strictures for the deduction?

Briefly: no, because the validity of this consciousness does not depend to any degree on anything that derives from our senses. Because it is a consciousness of ourselves and of our own minds, Kant thinks that whether or not this consciousness is accurate does not depend at all on anything that we sense. We have to sense something in order to have this consciousness, to be sure. But we are still immediately and directly conscious of our passivity in what we sense, as is evident from how the specific content of what we sense is totally irrelevant to that consciousness of our passivity. Because of the nature of our sensibility, we can be conscious of our receptivity in literally any sensation whatsoever. Whether that consciousness is *a posteriori* or *a priori* has to do with how that consciousness is shown to be valid, though. So whether or not that consciousness *is possible* without sensation is not to the point, so long as sensation is irrelevant to how that consciousness *is shown to be veridical*. And sensation is irrelevant to that, so the consciousness at the basis of the deduction is really valid *a priori*, as Kant always required.

Now, why isn't Kant's argument in the deduction actually a theoretical proof of our freedom? It might seem so: common human understanding and consciousness is supposed to be committed to distinguishing our phenomenal and noumenal selves. So if there must be laws for

the activity of our noumenal selves as well, then why doesn't this consciousness commit us to our autonomy even from the theoretical point of view?

Basically, because we cannot show that the laws for our noumenal selves have any effect in nature. (Beyond their role in making experience possible in the first place, as the *Critique of Pure Reason* argues.) Neither common human consciousness nor any other point of view can give us any insight at all into our intrinsic noumenal constitution, Kant thinks (G 56, 4:450-451). As such, we still have no basis for proving that our noumenal selves can really cause anything in nature. And the deduction of the Formula of Universal Law does not change this basic situation. Common human consciousness and understanding may commit us to thinking that our noumenal selves have such causal powers, but they do not contribute at all to proving that we do. Instead, they only remind us that we must consider ourselves as *noumena*, and so consider ourselves as outside of the nexus of natural causation, and hence as autonomous rational agents, under laws of freedom. The autonomy of reason in itself or in the world of understanding is quite different from its autonomy in nature, and to deny that we can give a theoretical proof of freedom is precisely to deny that we can show either freedom in nature or freedom outside of nature, but with effects in nature. But in order to consider ourselves as agents from this noumenal point of view, we must think of ourselves as autonomous, and so as subject to the Formula of Universal Law as well.

But now we are on the point of another puzzle, which will not require us to add to the preceding deduction so much as to draw out something implicit in its distinction between our phenomenal and our noumenal selves. To escape the analytic circle between the concepts of freedom and morality, Kant seeks to show that there is some sense in which the concept of freedom is objectively valid for us. The argument so far shows that if we are to think of ourselves as agents from the noumenal point of view, then we must think of ourselves as free, so that we must also think of ourselves as subject to the fundamental principle of freedom and morality. And the deduction so far has tried to show that we must think of ourselves this way. But if we

cannot know that we are the causes of effects in nature, then why must we think of ourselves as free agents *in nature*, and not just in the world of the understanding? Again, the deduction requires this, or else it cannot show that the Formula of Universal Law is really valid for all rational beings, at least as it pertains to our actions in the empirical world.

Here, the relation between our phenomenal and noumenal selves is crucial. In the course of introducing this distinction in the *Groundwork's* Third Section, Kant twice describes the noumenal world as the “basis” for the phenomenal world, once in making the general distinction between phenomena and *noumena*, and once again in applying this distinction to our selves (G 56, 4:451). And the point here should be familiar: appearances cannot exist on their own. Instead, an appearance must be an appearance *of something*. By the very concept of an appearance, they cannot be an ultimate reality—Kant sometimes treats that idea as an analytic truth, or as something approaching an analytic truth, anyway (CR 115, Bxxvi-xxvii; G 56, 4:451). So the existence and nature of appearances must ultimately be determined by or grounded in something that exists in its own right—a thing in itself. Accordingly, “*the world of understanding contains the ground of the world of sense and so too of its laws*” (G 58, 4:453).

Because of this general relation between phenomena and *noumena*, once we are forced to distinguish our phenomenal and noumenal selves, as common human consciousness requires, we then become compelled to view our noumenal self and its activity as the ultimate basis, ground, or source for our phenomenal actions. For we are conscious that both of these selves are ours—are us, so to speak—and so we must regard our noumenal self as the ultimate ground for our phenomenal self. However, we cannot *know* whether our noumenal selves are really the *ultimate* ground for our phenomenal self and for its actions in nature. Consciousness of our own activity allows us to know that we are a contributing ground to at least some parts of our phenomenal selves, like the form of our experience, for example. So the noumenal point of view that we take on ourselves is not “just” a standpoint. However, the ultimate ground of our

phenomenal selves remains beyond our ken: it is important that Kant only describes how we must regard ourselves through these sections of the deduction (G 56-59, 4:450-455).

However, regarding our noumenal selves as the ultimate ground of our phenomenal selves in turn compels us to view even our phenomenal actions as ultimately free from natural causation and as subject to the laws of freedom. For we must regard our noumenal self as fully autonomous. We still cannot know whether our noumenal self is the ultimate cause of our phenomenal actions at all, still less whether it is an autonomous cause. But this argument, if sound, still proves that we must regard ourselves as subject to the Formula of Universal Law from the practical point of view, and so Kant's deduction is complete.

So Kant thinks that the distinction between the world of sense and the world of understanding, and the corresponding distinction in our selves as members of both worlds, demonstrates that we must think of ourselves as the kind of rational being described in the earlier "idea of freedom" argument: the kind of rational being who must view themselves as the source or author of their own practical judgments and actions, and so who can act only under the idea of freedom. Likewise, this distinction even demonstrates that we must regard ourselves as the kind of rational being that Kant describes at the outset of the derivation of the Formula of Universal Law in the Second Section. In so doing, the argument escapes the analytic circle by showing how the concept of freedom finally must find a purchase on our self-conception as agents.

And despite the fact that common human understanding is committed to the idea that the noumenal is the ground of the phenomenal, the deduction of the Formula of Universal Law is still no theoretical argument for the freedom of our wills. For the deduction still only describes our necessary self-conception as it pertains to our actions—as it pertains to their ultimate source or ground, specifically—and so the conclusion of the deduction remains solely within the practical point of view. The deduction does show that we must regard ourselves as free both practically and transcendentally: free not just with respect to the consequences of our wills, but

free with respect to its grounds as well. For, after all, the deduction shows that we must think of ourselves as both free from determination by natural causes and positively autonomous. But we still only attribute these kinds of freedom to ourselves from the practical point of view, at least as they pertain to our actions in nature. For nothing in this argument proves or commits us to the idea that our noumenal selves actually do cause our phenomenal actions. For all that has been said in the deduction, our phenomenal actions could still ultimately be caused only by the nexus of natural causes, or even by things in themselves external to our wills.

7.4 Conclusion

Kant's summary of the deduction of the Formula of Universal Law confirms our account so far:

A rational being counts himself, as intelligence, as belonging to the world of understanding, and only as an efficient cause belonging to this does he call his causality a *will*. On the other side he is also conscious of himself as a part of the world of sense, in which his actions are found as mere appearances of that causality; but their possibility from that causality of which we are not cognizant cannot be seen; instead, those actions as belonging to the world of sense must be regarded as determined by other appearances, namely desires and inclinations. All my actions as only a member of the world of understanding would therefore conform perfectly with the principle of autonomy of the pure will; as only a part of the world of sense they would have to be taken to conform wholly to the natural law of desires and inclinations, hence to the heteronomy of nature. (The former would rest on the supreme principle of morality, the latter on that of happiness.) But because *the world of understanding contains the ground of the world of sense and so too of its laws*, and is therefore immediately lawgiving with respect to my will (which belongs wholly to the world of understanding) and must accordingly also be thought of as such, it follows that I shall cognize myself as intelligence, though on the other side as a being belonging to the world of sense, as nevertheless subject to the law of the world of understanding, that is, of reason, which contains in the idea of freedom the law of the world of understanding, and thus cognize myself as subject to the autonomy of the will; consequently the laws of the world of understanding must be regarded as imperatives for me, and actions in conformity with these as duties. (G 58, 4:453-454)

Here, Kant walks us back through all of the moments of the deduction described above: the distinction between the worlds of sense and the world of understanding, our membership in both, the fact that the latter world grounds the former, so that the laws of the world of understanding are also laws for nature, the idea of freedom as the idea of the law for the world

of understanding, the reciprocity of freedom and morality, and the ultimate need for us to view ourselves under the idea of freedom, and therefore as subject to the Formula of Universal Law as our fundamental practical principle. Kant even emphasizes the point that we have identified as fundamental to the argument: “*the world of understanding contains the ground of the world of sense and so too of its laws*”.²⁶

Our moral skeptic, then, wanted to know why we must think of ourselves under the idea of freedom, or, what amounts to the same thing, why we must think of ourselves as rational, in the sense at issue in the “acting under the idea of freedom” phase of the deduction. And Kant’s answer is, in short: because you must think of yourself as a member of the world of understanding, and as such you must think of yourself and of your reason as the autonomous ground of your actions in the empirical world. And then, because of the Reciprocity Thesis, it follows then that you must also think of yourself as subject to the Formula of Universal law, if only from the practical point of view. And, in a way, the whole point of this argument is to help us to become conscious of our necessary self-conception as rational agents, so as to demonstrate how everyone must be capable of this consciousness.

With that in mind, all that remains for us here is some concluding clarifications. First, then, let’s return to our interpretation of the “circle”.

I argued that the problem here is best understood as an issue about the objective validity of concepts. Up to that point in the discussion, Kant’s argument has proceeded analytically and has demonstrated how freedom, autonomy, and morality reciprocally imply one another. But

²⁶ Kant emphasizes the same point several times in the concluding parts of the Third Section. Compare: “it follows of itself that we must admit and assume behind appearances something else that is not appearance” (G 56, 4:451), “This must yield a distinction, although a crude one, between a *world of sense* and the *world of understanding*, the first of which can be very different according to the difference of sensibility in various observers of the world while the second, which is its basis, always remains the same” (G 56, 4:451), “beyond this constitution of his own subject, made up of nothing but appearances, he must necessarily assume something else lying at their basis, namely his ego as it may be constituted in itself” (G 56, 4:451), “things in themselves (though hidden) must lie behind appearances as their ground” (G 63, 4:459), and “*what belongs to mere appearance is necessarily subordinated by reason to the constitution of the thing in itself*” (G 64, 4:461).

that is not to show that any of these concepts really apply to us in any sense; it does not show that there is any sense in which any of these concepts are objectively valid for us.

But the “circle” here is sometimes understood to be a circular argument, and we noted how Kant gives us some evidence for that reading as well. For, in wrapping up his answer to the circle problem, he says that

The suspicion that we raised above is now removed, the suspicion that a hidden circle was contained in our inference from freedom to autonomy and from the latter to the moral law—namely that we perhaps took as a ground the idea of freedom only for the sake of the moral law, so that we could afterwards infer the latter in turn from freedom, and that we were thus unable to furnish any ground at all for the moral law but could put it forward only as a *petitio principii* disposed souls would gladly grant us, but never as a demonstrable proposition. (G 57-58, 4:453)

Here, Kant seems to say outright that the problem is a circular “inference”, and he explicitly cites the fallacy of circular reasoning: also known as the “*petitio principii*” fallacy. So why isn’t that a better reading of the circle problem?

To begin with, Kant’s official explanations of the problem do not make it look like a circularity in argument, at least not ultimately. He says that “we perhaps took as a ground the idea of freedom only for the sake of the moral law, so that we could afterwards infer the latter in turn from freedom, and that we were thus unable to furnish any ground at all for the moral law”. And he earlier described how “We take ourselves as free in the order of efficient causes in order to think ourselves under moral laws in the order of ends; and we afterwards think ourselves as subject to these laws because we have ascribed to ourselves freedom of the will” (G 55, 4:450). But in neither case does Kant describe an argument that reduces to the question-begging “the moral law is objectively valid for us because the moral law is objectively valid for us”. For there is nothing illegitimate or circular about taking the idea of freedom as a ground “only for the sake of the moral law”, at least so long as we can adequately ground the idea of freedom in turn. Likewise, there is nothing wrong or circular about thinking of “ourselves as subject to” moral

laws “because we have ascribed to ourselves freedom of the will”. Not so long as we have a sufficient justification for this ascription, anyway.

In both cases, then, the ultimate problem is that we ascribe freedom to ourselves in order to derive morality from it, but without giving an adequate basis for the ascription of freedom. Therefore, instead of a circularity in argument, the fundamental issue in both cases must be that the deduction of the Formula of Universal Law has not been adequately grounded, because the ascription of the idea of freedom has not been adequately grounded. And Kant’s solution to the circle problem better matches this statement of the problem anyway, since he resolves it by explaining why we must think of ourselves under the idea of freedom.

Thus, the ultimate problem underlying the suspicion of a circle is the fact that Kant has not yet gone beyond the analytic connections between freedom, autonomy, and morality, so as to provide a ground for the concept of freedom.²⁷ Without that further argument, though, it is true that the deduction could at best be circular: for then the best that we could do would be to covertly presuppose the validity of the Formula of Universal Law and try to mask this subterfuge under the idea of freedom; a mistake that would be easy to make, given how freedom implies morality and morality implies freedom. So the best argument that we could give would, in short, really be “a *petitio principii* disposed souls would gladly grant us”. But this is not the fundamental problem that needs to be solved. Instead, the real circle is in the analysis of the relevant concepts, though, as we have just illustrated, that circle in the analysis of concepts can project a corresponding circle in inference. As is only to be expected if the three concepts reciprocally imply one another.

²⁷ As an aside, Allison, citing Reinhold (1975) and Prauss (1983), mentions how the analytic connections between freedom, autonomy, and morality help create the sense that immoral actions cannot be free, according to Kant (Allison 1990, 134). Putting aside their proposed solutions to the problem and the varying degrees to which they are intended to preserve Kant’s thought, our discussion here ought to bring out how complicated the attribution of freedom to us is in the *Groundwork*, and the corresponding complexity in how we are subject to the moral law. In any case, though, I have suggested that emphasizing freedom as a property of the will solves their core problem.

Next, we do not have space here to fully discuss the differences between the “idea of freedom” deduction in the *Groundwork* and the “fact of reason” deduction in the *Critique of Practical Reason*, still less to evaluate the common suggestion that Kant rejected the former in favor of the latter. My own view, following a suggestion of Barbara Herman’s, is that the arguments are complementary in a structurally important way. After all, the Reciprocity Thesis holds that “a free will and a will under moral laws are one and the same”, so it is important that we get one deduction through the idea of a free will and one deduction through the idea of a will that is conscious of being under categorical or moral obligation (CP 28-29, 5:31).

But be that as it may, my interpretation here gives us some insight into why Kant may have concluded that the “idea of freedom” deduction was less than fully satisfying, if indeed he did end up thinking this. For our reading has argued that the *Groundwork*’s deduction fundamentally depends on conclusions from the *Critique of Pure Reason*, although put on a different footing: the deduction relies on the distinction between objects as appearances and objects as things in themselves, even though this distinction is to be found in common human consciousness and understanding, rather than in rigorous transcendental reflection. The *Groundwork*’s deduction of the Formula of Universal Law needs this distinction in order to explain why we must think of our practical selves and our agency under the idea of freedom. But in relying on our consciousness of the distinction between appearances and things in themselves, Kant’s deduction in the *Groundwork* relies on resources outside of the practical point of view itself. The conclusion of the deduction is restricted to the practical point of view anyway, to be sure, but the deduction itself is not solely practical, nevertheless.

This limitation need not cast any doubt on the security of the *Groundwork*’s deduction: the *Critique of Pure Reason* is supposed to back up common human consciousness with arguments that are supposed to be apodictically certain, as certain as the theorems of arithmetic, geometry, and fundamental physics, the possibility of which the First Critique sets out to explain (CR 102, Axv). Nevertheless, the deduction in the *Groundwork* might make it

appear as though practical reason cannot demonstrate the validity of its fundamental principle solely through its own authority, which in turn makes practical reason appear subordinate to the outside sources that prove the validity of its fundamental principle.

This result, if it is one, need not call the whole unity of reason into question, since reason could still be a unity under those outside sources. (Transcendental reason, presumably.) But Kant may actually think that reason, as an active faculty, needs must be a unity under practical reason. Regardless, though, one of the deepest themes of his critical practical philosophy, and one of the most important conclusions of the *Critique of Practical Reason* is the priority of practical reason in the constitution of reason and its unity (CP 100-102, 5:119-121). This priority cannot be demonstrated by the *Groundwork's* deduction.

So we can easily see why Kant would have wanted another deduction for the *Critique of Practical Reason*, even if he still thought the *Groundwork's* deduction was both perfectly sound and structurally necessary. For, given the distinctly critical project of the Second Critique, it would seem natural for its deduction to require the harder project: to give a purely practical deduction of the Formula of Universal Law, and thereby to create space for the priority of practical reason. And our immediate consciousness of categorical obligations for our actions, invoked in the “fact of reason” deduction, is surely part and parcel of the practical point of view.

For another, albeit related, thing: Kant never gave up on the idea that we can recognize our freedom and our morality only from the practical point of view. But the deduction in the *Groundwork* is still limited in a way that the Second Critique's deduction is not. The *Groundwork* only attempts to show that we must think of ourselves as free from the practical point of view, so that we must also think of ourselves as subject to the Formula of Universal Law from that point of view. It then follows that we really are subject to the Formula of Universal Law from the practical point of view, because if we have to represent ourselves and our agency in a certain way, then it follows that we really are that way, if only insofar as our agency is concerned.

But this argument is still rather indirect, since it depends on a self-conception that is necessary for us. The “fact of reason” deduction, however, can show that we are subject to the Formula of Universal Law directly, since this deduction holds that we just are, as a “fact of reason”, immediately conscious of being subject to categorical obligation (CP 28, 5:31). The “fact of reason” deduction still limits the validity of the Formula of Universal Law to the practical point of view, to be sure. However, this deduction points to something that we can be conscious of immediately, rather than something that we can be conscious of only through a complex consciousness of the difference between our activity and our passivity, and of the commitments entailed by the consciousness of that difference. Accordingly, the kind of reflection that the *Critique of Practical Reason* calls for is less abstruse, and less subject to the controversies connected to the *Critique of Pure Reason* and to Kant’s transcendental idealism.

Moreover, the “fact of reason” deduction makes it easier to understand the phenomenon of respect for the moral law, and easier to locate that respect within our “practical” consciousness, so to speak. It is relatively easy to understand how immediate consciousness of categorical necessitation could produce a feeling worthy of being called “respect”, and which could always be sufficient to motivate us to fulfill our categorical obligations (G 14n, 4:401n; CP 63-64, 5:73). And it would be relatively easy to search one’s own consciousness for a feeling matching that description. That is not to say that any of this would actually be easy, or that Kant’s position here would be philosophically persuasive. But it is rather harder to see why the need to think of yourself as a rational agent, under the idea of freedom, would have to produce the same feeling, or even the same consciousness of being under categorical obligation. Perhaps it would; that is a complicated question. Nevertheless, though, it is at least harder to see how the need to think of yourself as subject to the Formula of Universal Law, and the corresponding need to recognize that Formula as valid for you in a practical respect, would have to engage your feeling and your sensibility, in the way that the phenomenon of respect is supposed to entail, according to Kant.

So here again, we can see why Kant might have desired another deduction, even if he never questioned the certainty of this one in the *Groundwork*. Last, our interpretation of the *Groundwork*'s deduction of the Formula of Universal Law once again reveals the centrality of the form of law to Kant's critical ethics. We saw how the deduction hinges on the idea that "*the world of understanding contains the ground of the world of sense and so too of its laws*" (G 58, 4:453-454). Specifically, though, the world of understanding contains the ground of the world of sense through the laws of the former, which in this case are the laws of freedom, and, more specifically still, the Formula of Universal Law, as the fundamental laws of freedom. But, moreover, we can also see how important it is for morality and autonomy to be a matter of the form of our principles and of our volition. Earlier, we explained how Kant's idea is that free actions, autonomous actions, are all free by virtue of having a certain form in common, and we went on to explain how that form consists in how the component parts of an action's underlying principle are related to one another. And if the derivation and the Reciprocity Thesis are correct, it then turns out that the form of free and moral actions is the same: the form that allows their underlying maxims to be willed as universal laws. If Kant's arguments are correct, then, the "idea of freedom" ultimately amounts only to this: the idea of a will whose principles, and whose corresponding volition and subsequent actions, always exhibit this form. And to think of ourselves under the idea of freedom, or to be conscious of ourselves as rational agents, as rational beings "endowed with a will", is to be conscious of this form as setting a standard for us—an objectively valid standard, which also constitutes an absolutely good will. With that, we can bring our study here to a close.

Bibliography

Works by Kant

Deutsche Akademie der Wissenschaften zu Berlin, ed. *Kants gesammelte Schriften*. Berlin: Walter de Gruyter, 1900ff. 29 vols. Print.

Kant, Immanuel. *Critique of the Power of Judgment*. 2nd ed. Ed. Paul Guyer. Trans. Paul Guyer and Eric Matthews. New York: Cambridge UP, 2000. Print.

---. *Critique of Practical Reason*. Ed., Trans. Mary Gregor. New York: Cambridge UP, 1997.

---. *Critique of Pure Reason*. Ed., Trans. Paul Guyer and Allen W. Wood. New York: Cambridge UP, 1998. Print.

---. "First Introduction to the *Critique of the Power of Judgment*." *Critique of the Power of Judgment*. Ed. Paul Guyer. Trans. Paul Guyer and Eric Matthews. New York: Cambridge UP, 2000. 1-51. Print.

---. *Groundwork of the Metaphysics of Morals*. Ed., Trans. Mary Gregor. New York: Cambridge UP, 1997. Print.

---. *Immanuel Kant's Logic: A Manual for Lectures*. Ed. Gottlob Benjamin Jäsche. *Immanuel Kant: Lectures on Logic*. Ed., Trans. J. Michael Young. New York: Cambridge UP, 1992. 519-640. Print.

---. *Metaphysical Foundations of Natural Science*. Ed., Trans. Michael Friedman. New York: Cambridge University Press, 2004. Print.

---. *Metaphysics of Morals*. Ed., Trans. Mary Gregor. New York: Cambridge UP, 1996. Print.

---. *Prolegomena to Any Future Metaphysics*. Revised Ed. Ed., Trans. Gary Hatfield. New York: Cambridge UP, 2004. Print.

---. *Religion Within the Boundaries of Mere Reason*. Ed., Trans. Allen Wood and George di Giovanni. New York: Cambridge UP, 1998. Print.

Works by Hegel

Hegel, G. W. F. *Elements of the Philosophy of Right*. Ed. Allen W. Wood. Trans. H. B. Nisbet. New York: Cambridge UP, 1991. Print.

---. *Encyclopedia of the Philosophical Sciences in Basic Outline, Part I: Science of Logic*. Ed., Trans. Klaus Brinkmann and Daniel O. Dahlstrom. New York: Cambridge UP, 2010. Print.

---. *Phenomenology of Spirit*. Trans. A. V. Miller. New York: Oxford UP, 1977. Print.

Works by Other Authors

Allison, Henry E. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. New York: Cambridge UP, 1996. Print.

---. *Kant's Theory of Freedom*. New York: Cambridge UP, 1990. Print.

---. *Kant's Transcendental Idealism*. Revised Ed. New Haven: Yale UP, 2004. Print.

Anscombe, G. E. M. *Intention*. 2nd Ed. Oxford: Blackwell, 1963. Print.

Anselm. *On the Fall of the Devil*. *Anselm: Basic Writings*. Ed., Trans. Thomas Williams. Indianapolis: Hackett, 2007. 167-212. Print.

Aristotle. *Nicomachean Ethics*. Trans. W. D. Ross, revised by J. O. Urmson. *The Complete Works of Aristotle: The Revised Oxford Translation*, Vol. 2. Ed. Jonathan Barnes. Princeton: Princeton UP, 1984. 1729-1867. Print.

---. *Physics*. Trans. R. P. Hardie and R. K. Gaye. *The Complete Works of Aristotle: The Revised Oxford Translation*, Vol. 1. Ed. Jonathan Barnes. Princeton: Princeton UP, 1984. 315-446. Print.

Aune, Bruce. *Kant's Theory of Morals*. Princeton: Princeton UP, 1979. Print.

- Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: The University of Chicago Press, 1960. Print.
- . "Five Concepts of Freedom in Kant." *Stephan Körner—Philosophical Analysis and Reconstruction*. Ed. J. T. J. Szrednick. Dordrecht: Martinus Nijhoff, 1987. 35-51. Print.
- Bentham, Jeremy. *The Principles of Morals and Legislation*. Amherst, NY: Prometheus Books, 1988. Print.
- Bittner, Rüdiger. "Maximen." *Akten des 4. Internationalen Kant-Kongresses, Mainz 1974*. Ed. Gerhard Funke. New York: Walter de Gruyter, 1974. 485-498. Print.
- Burge, Tyler. *Origins of Objectivity*. New York: Oxford UP, 2010. Print.
- Cramer, Konrad. "Hypothetische Imperative?" *Rehabilitierung der praktischen Philosophie*, Vol 1. Ed. Manfred Reidel. Freiburg: Rombach, 1972. 159-212. Print.
- Engstrom, Stephen. "The Concept of the Highest Good in Kant's Moral Theory." *Philosophy and Phenomenological Research* 52.4 (1992): 747-780. Print.
- . *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge: Harvard UP, 2009. Print.
- , and Jennifer Whiting, eds. *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*. New York: Cambridge UP, 1996. Print.
- Fichte, J. G. *The Science of Ethics*. Trans. A. E. Kroeger. London: Kegan Paul, Trench, Truebner, 1897. Print.
- Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68.1 (1971): 5-20. Print.
- Frege, Gottlob. *Philosophical and Mathematical Correspondence*. Ed. Gottfried Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel, Albert Veraart, and Brian McGuinness. Trans. Hans Kaal. Chicago: The University of Chicago Press, 1980. Print.
- Friedman, Michael. "Causal Laws and the Foundations of Natural Science." *The Cambridge Companion to Kant*. Ed. Paul Guyer. New York: Cambridge UP, 1992. 161-199. Print.

- . *Kant's Construction of Nature*. New York: Cambridge UP, 2013. Print.
- . *Kant and the Exact Sciences*. Cambridge: Harvard UP, 1992. Print.
- Guyer, Paul. "Kant's Conception of Empirical Law." *Kant's System of Nature and Freedom: Selected Essays*. New York: Oxford UP, 2005. 38-55. Print.
- . "Kant's Ether Deduction and the Possibility of Experience." *Kant's System of Nature and Freedom: Selected Essays*. New York: Oxford UP, 2005. 74-85. Print.
- , ed. *The Cambridge Companion to Kant*. New York: Cambridge UP, 1992. Print.
- Herman, Barbara. "Mutual Aid and Respect for Persons." *The Practice of Moral Judgment*. Cambridge: Harvard UP, 1993. 45-72. Print.
- . *The Practice of Moral Judgment*. Cambridge: Harvard UP, 1993. Print.
- Höffe, Otfried. "Kants kategorischer Imperativ als Kriterium des Sittlichen." *Zeitschrift für philosophische Forschung* 31.3 (1977): 354-384. Print.
- Hume, David. *A Treatise of Human Nature*. 2nd Ed. Ed. L. A. Selby-Bigge. New York: Oxford UP, 1978. Print.
- Kerstein, Samuel J. *Kant's Search for the Supreme Principle of Morality*. New York: Cambridge UP, 2002. Print.
- Korsgaard, Christine M. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. New York: Oxford UP, 2008. Print.
- . *Creating the Kingdom of Ends*. New York: Cambridge UP, 1996. Print.
- . *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford UP, 2009. Print.
- . *The Sources of Normativity*. New York: Cambridge UP, 1996. Print.
- Langton, Rae. *Kantian Humility: Our Ignorance of Things in Themselves*. New York: Oxford UP, 1998. Print.
- Leibniz, G. W. "The Principles of Philosophy, or, the Monadology." *G. W. Leibniz: Philosophical Essays*. Ed., Trans. Roger Ariew and Daniel Garber. Indianapolis, Hackett: 1989. 213-225. Print.

- Newton, Isaac. *The Principia: Mathematical Principles of Natural Philosophy*. Trans. I. Bernard Cohen and Anne Whitman. Berkeley: University of California Press, 1999. Print.
- O'Neill, Onora. *Acting on Principle: An Essay on Kantian Ethics*. 2nd Ed. New York: Cambridge UP, 2013. Print.
- . *Constructions of Reason: Explorations of Kant's Practical Philosophy*. New York: Cambridge UP, 1989. Print.
- Pluhar, Werner S., trans. *Critique of Pure Reason*. By Immanuel Kant. Indianapolis: Hackett, 1996. Print.
- Prauss, Gerold. *Kant über Freiheit als Autonomie*. Frankfurt am Main: Vittorio Klostermann, 1983. Print.
- Rawls, John. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77.9 (1980): 515-572. Print.
- . *Lectures on the History of Moral Philosophy*. Ed. Barbara Herman. Cambridge: Harvard UP, 2000. Print.
- Reath, Andrews. *Agency and Autonomy in Kant's Moral Theory: Selected Essays*. New York: Oxford UP, 2006. Print.
- Reinhold, Carl Leonhard. "Einige Bemerkungen über die in der Einleitung zu den 'Metaphysischen Anfangsgründen der Rechtslehre' von I. Kant aufgestellten Begriffe von der Freiheit des Willens." *Materialen zu Kants "Kritik der praktischen Vernunft"*. Ed. Rüdiger Bittner and Konrad Cramer. Frankfurt am Main: Suhrkamp Verlag, 1975. 310-324. Print.
- Rousseau, Jean-Jacques. *Of the Social Contract. Rousseau: The Social Contract and Other Later Political Writings*. Ed., Trans. Victor Gourevitch. New York: Cambridge UP, 1997. 39-152. Print.
- Sophocles. *Antigone. The Three Theban Plays: Antigone, Oedipus the King, Oedipus at Colonus*. Trans. Robert Fagles. New York: Viking Penguin, 1982. 55-128. Print.

Strawson, P. F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1-25. Print.

Tittel, Gottlob August. *Über Herr Kants Moralreform*. Frankfurt: Pfahler, 1786. Print.

Watkins, Eric. *Kant and the Metaphysics of Causality*. New York: Cambridge UP, 2005. Print.

Wood, Allen W. *Hegel's Ethical Thought*. New York: Cambridge UP, 1990. Print.