UNIVERSITY OF CALIFORNIA,
IRVINE

Towards a Mathematical Theory of Group Creativity and Collaboration

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical Behavioral Sciences

by

Santiago Ortolano Guisasola

Dissertation Committee:
Professor Donald Saari, Chair
Professor Louis Narens
Professor Natalia Komarova

2018

DEDICATION

To my family and friends. I love you all.

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vi**

ACKNOWLEDGMENTS **viii**

CURRICULUM VITAE **ix**

ABSTRACT OF THE DISSERTATION **x**

1 Introduction **1**

2 Potential Structure of $2 \times \ldots \times 2$ Games **5**

2.1 Introduction 5
2.2 Strategic and Behavioral Decomposition 7
2.3 2×2 Potential Structure 12
2.3.1 Introduction 12
2.3.2 Utility Functions from the Decomposition 13
2.3.3 New Basis for Utility Functions 17
2.3.4 A Basis for $\mathbb{G}_P^N$ and $(\mathbb{G}_P^N)^\perp$ 21
2.3.5 Three Classes of Potential Games 25
2.3.6 An Application: The 2×2 Congestion Game 26
2.3.7 Ordinal Potential Games 26
2.3.8 Discussion 28
2.4 $2 \times 2 \times 2$ Potential Structure 29
2.4.1 Emergent Structures in the Utility Functions 29
2.4.2 Potential Structure 32
2.4.3 Symmetric $2 \times 2 \times 2$ Games 35
2.4.4 An Application: The Legislator Game 38
2.5 $2 \times \ldots \times 2$ Potential Structure 40
2.5.1 $2 \times \ldots \times 2$ Potential Games 42
2.5.2 Generalized Matching Pennies 45
2.6 3×3 Potential Structure 47
2.6.1 A Change of Basis 48
2.6.2 Potential Game Requirements 52

2.6.3 The 3×3 Orthogonal Subspace . . . 54
2.7 Externalities . . . 55
2.7.1 2×2 Externality Structure . . . 56
2.7.2 $2 \times 2 \times 2$ Externality Structure . . . 56
2.7.3 $2 \times \ldots \times 2$ Externality Structure . . . 58
2.7.4 3×3 Externality Structure . . . 58
2.8 Identical Play . . . 59
2.9 Common Interest . . . 61

3 Innovation Diffusion on Networks **63**
3.1 Introduction . . . 63
3.2 Disagreement in Potential Games . . . 65
3.2.1 Symmetric 2×2 Potential Games . . . 66
3.3 Breaking α Symmetry . . . 84
3.4 Breaking β Symmetry . . . 91
3.5 Potential and Welfare on Networks . . . 93
3.5.1 Young's Example . . . 98
3.6 Autonomy . . . 101
3.6.1 Potential Autonomy . . . 101
3.6.2 Agency Autonomy . . . 102
3.6.3 An Alternative Parametrization . . . 104
3.7 Conclusion . . . 109

4 Collaboration on Networks **111**
4.1 Introduction . . . 111
4.2 Collective Improvisation Experiments . . . 112
4.2.1 The Nine Sessions . . . 114
4.2.2 Discussion . . . 121
4.2.3 Next Experimental Steps . . . 121
4.2.4 Analytic Tools . . . 122
4.3 Model: A Trio on Networks . . . 122
4.3.1 Wheel Network . . . 125
4.3.2 Star Network . . . 132
4.3.3 Discussion . . . 137
4.4 Model: An Orchestra and the Conductor . . . 138
4.4.1 The Conductor . . . 143
4.5 Model: Collaboration . . . 147
4.5.1 Perfect Communication . . . 150
4.6 $2 \times 2 \times 2$ Collaboration . . . 152
4.7 Collaboration on Networks . . . 162
4.7.1 2×2 . . . 162
4.7.2 $2 \times 2 \times 2$ . . . 169
4.8 Information Flow and the Trio Model . . . 179

Bibliography **181**

LIST OF FIGURES

Page

2.1 $\gamma\delta$-plane with $\alpha > 0$ . . . 36

3.1 $\alpha\gamma$-plane . . . 70
3.2 $\alpha\beta$-plane with $\gamma \geq 0$ . . . 83
3.3 $\alpha\beta$-plane with $\gamma < 0$ . . . 83
3.4 24 regions for arbitrary γ . . . 85

4.1 Session 1 . . . 115
4.2 Session 2 . . . 115
4.3 Session 3 . . . 116
4.4 Session 4 . . . 117
4.5 Session 5 . . . 118
4.6 Session 6 . . . 118
4.7 Session 7 . . . 119
4.8 Session 8 . . . 120
4.9 Session 9 . . . 120
4.10 Wheel Network . . . 126
4.11 Star Network . . . 133
4.12 Orchestra without a Conductor . . . 139
4.13 Orchestra with a Conductor . . . 143
4.14 Wheel and Star Networks . . . 170
4.15 Full Coordination . . . 178
4.16 Division of Labor . . . 178
4.17 Legislator . . . 178

LIST OF TABLES

Page

2.1 The Basis Vectors e_1 and e_2 in Normal-Form . . . 7
2.2 Arbitrary 2×2 Game $\mathcal{G}$ . . . 8
2.3 The kernel component $\mathcal{G}^K$ . . . 8
2.4 The Nash component $\mathcal{G}^N$ . . . 9
2.5 The behavioral component $\mathcal{G}^B$ . . . 10
2.6 Decomposed Game $\mathcal{G}$ . . . 14
2.7 Basis for the Space of 2×2 $\mathcal{G}^N$ Components . . . 18
2.8 A New Basis for $\mathbb{G}^N$ . . . 19
2.9 The Basis Vector n_γ . . . 23
2.10 The Basis Vector $\widetilde{n}$ . . . 24
2.11 Cycle-Inducing Inequalities . . . 27
2.12 Emergent Structure in $2 \times 2 \times 2$ Games: γ . . . 30
2.13 Emergent Structure in $2 \times 2 \times 2$ Games: δ . . . 31
2.14 Arbitrary $2 \times 2 \times 2$ Nash Component . . . 33
2.15 Symmetric $2 \times 2 \times 2$ Game . . . 35
2.16 Region 1; $|\alpha + \delta| < 2\gamma$ & $\delta > \alpha$ . . . 35
2.17 Region 2; $\alpha + \delta > 2|\gamma|$ & $\delta > \alpha$ . . . 36
2.18 Region 3; $|\alpha + \delta| < -2\gamma$ & $\delta > \alpha$ . . . 36
2.19 Region 4; $|\alpha + \delta| < -2\gamma$ & $\delta < \alpha$ . . . 37
2.20 Region 5; $\alpha + \delta > 2|\gamma|$ & $\delta < \alpha$ . . . 37
2.21 Region 6; $|\alpha + \delta| < 2\gamma$ & $\delta < \alpha$ . . . 37
2.22 Region 7; $\alpha + \delta < -2|\gamma|$ . . . 37
2.23 The Legislator Game . . . 38
2.24 Legislator Game with $b = 2$ and $c = 1$ . . . 39
2.25 Nash Component of Legislator Game . . . 39
2.26 Behavioral Component of Legislator Game . . . 39
2.27 3×3 Nash Structure . . . 47
2.28 3×3 Behavioral Structure . . . 48
2.29 3×3 Kernel . . . 48
2.30 2×2 Externality Structure . . . 56
2.31 $2 \times 2 \times 2$ Externality Structure . . . 57

3.1 Symmetric Potential Game . . . 66
3.2 Arbitrary Nash Component of a Game $\mathcal{G}$ in Normal-Form . . . 71

3.3 $F(\mathcal{G})$, and $G(\mathcal{G})$ 71
3.4 2×2 Broken α Symmetry 84
3.5 Arbitrary Nash Component $\mathcal{G}^N$ in Normal-Form 87
3.6 $F(\mathcal{G})$, $G(\mathcal{G})$, and $H(\mathcal{G})$ 88
3.7 Coordinate System 94
3.8 Young's Example of Disagreement 100
3.9 Young's Social Innovation Game 102
3.10 Newton *et al.* 's Model 103
3.11 Nash Component of Newton *et al.* 's Model 103
3.12 Behavioral Component of Newton *et al.* 's Model 103
3.13 Parametrization of 2×2 Coordination Games 105

4.1 2×2 Discussion Model 148

ACKNOWLEDGMENTS

I would like to thank everyone who has encouraged me not only to start but to keep going.

Special thanks to all members of the weekly research seminars with Don Saari, Dan Jessie, Ryan Kendall, Tomas McIntee, and June Zhao, and the members of my committee, Don Saari, Louis Narens, and Natalia Komarova.

I would like to especially thank Don Saari for guiding me through this challenge and for being a truly positive influence on my life. Words will never be enough to express my gratitude, but, thank you, Don.

Throughout my time at UCI, Louis Narens and John Sommerhauser provided an incredible amount of help in navigating paperwork, deadlines, and funding. Thank you. And thank you to all those who made my life at UCI brighter.

The musicians who made the collective improvisation experiments possible have my sincere thanks and appreciation.

The University of California Natural Reserve System hosted me for a few writing retreats. Specifically, I had the privilege to spend time at the James San Jacinto Mountains Reserve, Burns Piñon Ridge Reserve, and the Kenneth S. Norris Rancho Marino Reserve. I will be forever grateful for the opportunity to focus on my thesis in the beautiful and diverse nature of California.

The McNair Scholars program and the UC Irvine Graduate Division are largely responsible for opening the door to my doctoral studies. Thank you to all who were involved in this.

Alexander Tovbis, Barry Griffiths, and Joseph Brennan were highly influential during my undergraduate career. Thank you for your support and encouragement.

The support I received from a very special group of people was key in my success. You have my deepest gratitude.

Some friends showed enormous support during the last few months of my graduate career. I am humbled and grateful.

The inspiration for most ideas in this thesis came from my experiences in life. Thank you to all who are a part of it, most notably my closest friends, a genuinely funky group of people.

Finally, an enormous thank you to my wonderful parents.

CURRICULUM VITAE

Santiago Ortolano Guisasola

EDUCATION

Doctor of Philosophy in Mathematical Behavioral Sciences **2018**
University of California Irvine *Irvine, CA*

Bachelor of Science in Mathematics, minor in Computer Science **2012**
University of Central Florida *Orlando, FL*

ABSTRACT OF THE DISSERTATION

Towards a Mathematical Theory of Group Creativity and Collaboration

By

Santiago Ortolano Guisasola

Doctor of Philosophy in Mathematical Behavioral Sciences

University of California, Irvine, 2018

Professor Donald Saari, Chair

This dissertation is a stepping stone into an in-depth study of group creativity. The foundation for this work concerns collective musical improvisation, and the project grew from an experiment with improvising musicians on networks of controlled listening. This experiment served as a source for intuition and its many sessions records of the unfolding creative and collaborative process in real-time.

A literature review of related themes suggested potential games as a starting point for the theoretical modeling due to their embodiment of the many facets of coordination, undeniably a strong force in music and group creativity. The current status of these games proved to be insufficient for the goal of this thesis. For this reason, the beginning of this thesis, chapter 2, provides a fairly complete picture of 2-strategy potential games for any number of agents, and for 2-agent 3-strategy games.

In chapter 3, to show the power of the coordinate system developed for potential games in chapter 2, issues in the literature involving possible conflicts between measures of "good" group outcomes are elucidated. Finally, in chapter 4, the collective improvisation experiment is described followed by models of an improvising trio, an orchestra, and the collective generation of information. Admittedly, the work in chapter 4 is still in progress, and will be improved on in the years to come.

Chapter 1

Introduction

The total objective of this dissertation is to start an in-depth study of group creativity and collaboration. The first step in this challenge involves laying a concrete foundation from which to draw analogies and results. Consequently, as described in more detail in chapter 4, collective musical improvisation was selected as an anchor to the idea of group creativity. The topic of group creativity has been, although perhaps not explicitly stated, at least implicitly considered in a lot of the game theory literature. Some implicit work has been done by Young [18] [19], Newton and Sercombe [12], Bala and Goyal [1] [2], Zollman [21] [22] [20], and many others. In [18] [19] [12] the work focuses on the diffusion of innovation on networks, which relates to questions in group creativity concerning the spread of new ideas. The results of [1] [2] detail models of network formation and the transmission of information. On the other hand, [21] [22] [20] examines scientific collaboration and the collective pursuit of truth.

The initial stage of this exploration involved an experiment with musicians where the musicians improvised over several networks of controlled listening. The intention behind the experiments was to observe and draw intuition from an unfolding creative process in real-time, under varying topologies of information flow. Because this experiment was done prior to the theoretical research, it cannot be used to justify the theoretical discussion which follows, but it most surely has motivated what has been done. In this manner, it has played a crucial role in the development of this project and for this reason, it is included in this thesis and more explicitly spelled out in chapter 4.

In music, controlling the feedback of information is a common technique for achieving certain creative goals. For instance, "overdubbing" in music production is where specific parts to a song are recorded over those that have already been recorded [3]. The reasons for this may be that the musicians involved are geographically separated, unavailable at the same time, or that, purposely, control over the network of influence is desired. A notable example comes from bassist Jaco Pastorious' recording of "Crisis" [13]. Pastorious' intention, an experiment in free-jazz, was to create a song that was mostly "chaotic" but whose parts still had a common underpinning, giving rise a strong, but not complete sense of dissonance [10]. Roughly, Pastorious recorded his bass part and then engaged other musicians, one or a few at a time, to play their piece over the recorded bass lines, sometimes guiding the musician(s) with gestures and leaking in other recorded parts. In total, including Pastorious, 11 musicians participated in the generation of "Crisis." For more details see [10].

In many ways the above is alluding to the strong presence of coordination in group creativity and music, where controlling the network of influence extends to control over coordination. After reviewing the literature, it became apparent that a good starting point for the theoretical modeling of the creative effort is with potential games. The reason for this is that potential games, also referred to as common interest games, are those in which the agents are in some strong sense trying to coordinate, which is undoubtedly a factor in music and group creativity. Indeed, there have been several papers which have captured interesting aspects, such as Young [18] [19] and Newton *et. al.* [12]. But for the purposes of this research, it became clear that the current status of potential games was not sufficiently developed.

Therefore, in chapter 2, a fairly complete description of potential games is offered for any number of players with 2 strategies and and for 2 agents and 3 strategies. More specifically, a coordinate system with intuitive parameters is developed, offering a precise language to analyze potential games and coordination. As a result, it goes beyond the 1-dimensional model used by Young [19] and the 2-dimensional model by Newton *et. al.* [12], where for 2-agent games the 7-dimensional situation can be covered in the framework. As such, there are several new results. The many variants of coordination for n-agents are exposed through the coordinate system, detailing situations of pure coordination, anti-coordination, and the many possibilities in between. Moreover, this framework pays special attention to the structure of externalities in games, which has been shown in [8] to slip through most solution concepts in game theory.

How does one measure whether an effort in a game for group creativity is successful? There are several ways, and two of them are described in this dissertation. One of them is the utilitarian social welfare function, which adds up the individual benefits, and is often used a measure for what is good for the group. The second is the coordination involved in potential games. In the literature, some preliminary results in this direction have shown they can be in conflict but there is not a strong understanding of why this is so. Consequently, in chapter 3, the conflict between social welfare and the potential function is spelled out completely. The externality structure of games is shown to be a major culprit behind this conflict.

Finally, chapter 4, which admittedly is work still in progress, and probably will be for years to come, pulls together some of these elements and looks at what happens on networks, which is the area in which the experiments were conducted. One of the goals of modeling group creativity is to offer an alternative perspective to the literature on scientific collaboration as a social enterprise. There, the focus is on the pursuit of an underlying "truth", where the networks represent information flow, specifically the sharing of results and theories. Group creativity, on the other hand, generally has no "truth" to be uncovered, which is especially true in musical improvisation. At the same time, as is the case of scientific collaboration, the flow of information is instrumental in the process. In chapter 4, the collective improvisation experiments are described briefly before some seeds are planted for future work. These are the beginnings of a model of an improvising trio, an orchestra, and the collective generation of information in a collaboration. Roughly, the improvising trio comprises a multi-dimensional game that takes into account the rhythmic, harmonic, and melodic elements of music, and the role of different instruments in pronouncing these elements. The model for the orchestra makes an argument for the conductor's role in assisting the coordination of interpretations. Lastly, the model of collaboration uses the structure of game theory to model the generation and transmission of information.

Chapter 2

Potential Structure of $2 \times \ldots \times 2$ Games

2.1 Introduction

Potential games grew from Rosenthal's 1973 congestion game [14], a model of agents choosing a road to a nearby town. Rosenthal showed that congestion games always have pure strategy Nash equilibria, and his method was generalized in 1996 by Shapley and Monderer in their work on potential games [11].

The key characteristic of a potential game is its potential function, a global payoff function that contains information about the game's unilateral incentive structure. The potential function is nice because it aggregates an entire game into a single function. It is no surprise that it is behind most research into potential games. Potential games, in addition, capture a strong essence of coordination, which is reflected in their guaranteed existence of pure strategy Nash equilibria. These games are often called "common interest games."

A recent decomposition of games [8], which is described in the beginning of this chapter, showed that a game can be uniquely decomposed into a trivial component, a component that contains the structure of unilateral deviations, and a component that describes the externality structure of the game. This externality structure was shown to be ignored by most solution concepts that focus on unilateral deviations in game theory.

Because the potential function is tied to a potential game's unilateral structure, an initial intuition, which is proved to be the case, is that the potential function ignores the externality component. In games where "common interest" is of major concern, the externality structure should play an important role.

To clarify the structure of potential games, this chapter describes a coordinate system for 2-strategy potential games with any number of agents, and for 3-strategy 2-agent potential games following a similar methodology as in [8]. We begin by reviewing the decomposition, before uncovering the structure of $2 \times \ldots \times 2$ potential games, building up from the 2×2 and $2 \times 2 \times 2$ structures. After all of this, the structure of 3×3 potential games is unveiled. Special attention is paid to the externality structure of potential games. Lastly, a special case of the externality structure gives rise to identical play games, which are those where all agents receive the same payoff in a given strategy profile. This is given attention due to its relevance in group creativity and the collective generation of information, the focus of the following chapters.

The analysis begins with 2×2 games. We decompose the potential structure, and prove several structural results. These include identifying the class of games that are orthogonal to potential games, and relating the structure of potential games to the structure of identical play games.

Then, we look into $2 \times 2 \times 2$ games and once certain structures emerge, we extend the results to $2 \times \ldots \times 2$ games. Here we prove that the pure Nash strategy structure of a symmetric game is given by the dominant potential component. We discuss how this result generalizes.

2.2 Strategic and Behavioral Decomposition

In this subsection we review the essentials of a recent decomposition of games [8]. We discuss only 2×2 games and note that the results extend to any finite number of agents and strategies. For a full exposition see [8].

Let us define a game $\mathcal{G}$.

Definition 2.1. A game $\mathcal{G}$ consists of a set of agents $\mathcal{N} = \{1, 2\}$ where each agent $i \in \mathcal{N}$ has a set of strategies $S_i = \{\sigma_{i1}, \sigma_{i2}\}$. and a payoff function $\pi_i : S_1 \times S_2 \to \mathbb{R}$. We write this as $\langle \mathcal{N}, S_1, S_2, \pi_1, \pi_2 \rangle$

For simplicity, refer to strategies σ_{11} and σ_{21} as strategy A, and to σ_{12} and σ_{22} as strategy B. The space of all games $\mathcal{G}$ forms a vector space, which we denote by $\mathbb{G}$. In the case of 2×2 games we have that $\dim(\mathbb{G}) = 8$. The canonical basis for $\mathbb{G}$ is $e_1 = (1, 0, 0, 0, 0, 0, 0, 0)$, $e_2 = (0, 1, 0, 0, 0, 0, 0, 0), \ldots, e_8 = (0, 0, 0, 0, 0, 0, 0, 1)$. We show the normal-form representation of e_1 and e_2 in Table 2.1.

$e_1 =$

	A		B	
A	1	0	0	0
B	0	0	0	0

$e_2 =$

	A		B	
A	0	1	0	0
B	0	0	0	0

Table 2.1: The Basis Vectors e_1 and e_2 in Normal-Form

For for an arbitrary game $\mathcal{G} \in \mathbb{G}$ with payoffs given in Table 2.2, we can write $\mathcal{G} = (a_1, a_2, b_1, c_2, c_1, b_2, d_1, d_2)$. Alternatively,

$$\mathcal{G} = a_1 e_1 + a_2 e_2 + b_1 e_3 + c_2 e_4 + c_1 e_5 + b_2 e_6 + d_1 e_7 + d_2 e_8$$

	A		B	
A	a_1	a_2	b_1	c_2
B	c_1	b_2	d_1	d_2

Table 2.2: Arbitrary 2×2 Game $\mathcal{G}$

The decomposition offered in [8] amounts to a change of basis that reflects meaningful symmetries inherent to games. This basis is then categorized into three components, the Nash component $\mathcal{G}^N$, the behavioral component $\mathcal{G}^B$, and the kernel $\mathcal{G}^K$. In other words, for any game $\mathcal{G}$, we can write $\mathcal{G} = \mathcal{G}^N + \mathcal{G}^B + \mathcal{G}^K$.

The kernel contains the average of all payoffs of each agent, which we denote by κ_1 and κ_2 for agents 1 and 2 respectively. There is no contribution to the payoff structure of $\mathcal{G}$ beyond a common value, for each agent, in all strategy profiles. We can express the basis of the kernel using the canonical basis, where $e_1^K = e_1 + e_2 + e_3 + e_4$, and $e_2^K = e_5 + e_6 + e_7 + e_8$. Then, any kernel component $\mathcal{G}^K$ can be written $\mathcal{G}^K = \kappa_1 e_1^K + \kappa_2 e_2^K$. It is immediate that $\dim \mathbb{G}^K = 2$, where $\mathbb{G}^K$ is the kernel subspace of the vector space of games $\mathbb{G}$. We write the kernel in normal-form in Table 2.3.

$\mathcal{G}^K =$

	A		B	
A	$\kappa_1,$	κ_2	$\kappa_1,$	κ_2
B	$\kappa_1,$	κ_2	$\kappa_1,$	κ_2

Table 2.3: The kernel component $\mathcal{G}^K$

The kernel component simply scales the game. Any method of comparing payoffs will eliminate the common kernel quantity. Because of this, throughout this dissertation, we mostly ignore the kernel except for important cases.[1]

[1]In theory, the effect of the kernel component when analyzing games is non-existent or negligent. Like we stated, this is because any method of comparing payoffs will get rid of the common kernel quantity. In

Beyond this common value, any differences between unilateral payoffs is captured by the Nash component, $\mathcal{G}^N$. The authors in [8] show that many solution concepts, like the Nash equilibrium, quantal response equilibrium, and best response dynamic, only focus on this this component of a game. Indeed, the $\mathcal{G}^N$ component contains only, and all of the information necessary to compute these solution concepts.

This has important implications. For example, suppose we have two seemingly different games $\mathcal{G}_1$ and $\mathcal{G}_2$ such that $\mathcal{G}_1^N = \mathcal{G}_2^N$. These games will have the same set of Nash equilibria and will behave exactly the same under all related, unilateral-focused, solution concepts. In other words, varying the $\mathcal{G}^K$ and $\mathcal{G}^B$ (soon to be defined) components will have no effect on these solution concepts.[2]

The Nash component $\mathcal{G}^N$ is given in Table 2.4, where the change of basis here is given by $e_1^N = e_1 - e_5$, $e_2^N = e_3 - e_7$, $e_3^N = e_2 - e_4$, and $e_4^N = e_6 - e_8$. In this case we have $\dim \mathbb{G}^N = 4$, where $\mathbb{G}^N$ is the Nash subspace of $\mathbb{G}$, and a Nash component $\mathcal{G}^N$ can be written as $\mathcal{G}^N = \alpha_{11}e_1^N + \alpha_{12}e_2^N + \alpha_{21}e_3^N + \alpha_{22}e_4^N$.

$\mathcal{G}^N =$	A	B
A	α_{11} α_{21}	α_{12} $-\alpha_{21}$
B	$-\alpha_{11}$ α_{22}	$-\alpha_{12}$ $-\alpha_{22}$

Table 2.4: The Nash component $\mathcal{G}^N$

In $\mathcal{G}^N$, an agent's payoffs, for a fixed strategy of the opponent, are centered around zero. This way, the positive value always represents the direction of unilateral incentives, while

reality, however, the effect of the kernel component has not been explored. One might imagine that extremely large $\mathcal{G}^K$ compared with $\mathcal{G}^N$ and $\mathcal{G}^B$ will wash out the effects of $\mathcal{G}^N$ and $\mathcal{G}^B$. Imagine playing a game with another person to win one of \$1,000,000.10, \$1,000,000.06, \$1,000,000.00, and \$1,000,000.02 dependent on both of your decisions. Here the kernel is \$1,000,000. Would you try to be strategic or cooperative in this scenario? Probably not. Nonetheless, we will ignore the kernel component throughout this dissertation, unless explicitly brought up.

[2] In [7], the authors have experimental subjects play a variety of games with constant $\mathcal{G}^N$ structure but varying $\mathcal{G}^B$. The theory predicts constant behavior, which is far from what was observed. An implication of this is that the $\mathcal{G}^B$ structure plays an important role in game theory that must be better understood.

preserving the difference. For example, if $\alpha_{11} > 0$ in Table 2.4, then agent 1's best response to agent 2 playing A is to play A, and their change in payoff is $2\alpha_{11}$. It follows that any strategy profile with positive payoffs for all agents is a Nash equilibrium. We will state this as a theorem but leave the formal proof to [8].

Theorem 2.2. *A strategy profile is a Nash equilibrium of a game $\mathcal{G}$, where each player has only two strategies, if and only if all payoffs in this strategy profile are positive in $\mathcal{G}^N$*

Together, the kernel and the Nash component are 6-dimensional, where the space of all 2×2 games is 8-dimensional. One might wonder, what else could there be? After all, a game's unilateral deviation structure is captured in $\mathcal{G}^N$, and any additional scaling is contained in $\mathcal{G}^K$. The remaining 2 dimensions form the behavioral component, $\mathcal{G}^B$, that, because of its orthogonality to $\mathcal{G}^N$, cannot contain any unilateral information, and, because of its orthogonality to $\mathcal{G}^K$, is not trivial. Before discussing the behavioral component further, it is useful to visualize it in normal-form, as displayed in Table 2.5, where the change of basis is given by $e_1^B = e_1 - e_3 + e_5 - e_7$ and $e_2^B = e_2 + e_4 - e_6 - e_8$.

$\mathcal{G}^B =$

	A	B
A	$\beta_1, \quad \beta_2$	$-\beta_1, \ \beta_2$
B	$\beta_1, \ -\beta_2$	$-\beta_1, -\beta_2$

Table 2.5: The behavioral component $\mathcal{G}^B$

It is clear that there is no information on unilateral deviations in the payoffs of $\mathcal{G}^B$. However, this is not a trivial structure, but, because it had not been introduced to the game theory literature, it is ignored by most modern solution concepts. By including this component, as done in this thesis, new results and explanations are derived. We see that when column player chooses strategy A, row player gets a payoff of β_1 invariant over their own strategy. On the other hand, if column player chooses B, row agent gets the payoff of $-\beta_1$. Similarly we have row player, in a way, *giving* column player either β_2 or $-\beta_2$ as a consequence of choosing A or B. Because of this, one interpretation of $\mathcal{G}^B$ is that it describes the structure of *externalities* in a game.

The interplay of the $\mathcal{G}^N$ and $\mathcal{G}^B$ components is a fascinating feature of game theory that has not been properly explored simply because modern solution concepts analyze only $\mathcal{G}^N$. Its presence, however, pervades. It is the tension between where unilateral incentives lead and what they produce in terms of externalities that is behind, among others, the enigmatic prisoner's dilemma and stag hunt, giving rise to the fragile and elusive notion of cooperation.

It is obvious, with this decomposition, that if agents are only playing inside of $\mathcal{G}^N$, then they will never get to taste the fruits of $\mathcal{G}^B$. Most, if not all, techniques of achieving and sustaining cooperation must involve somehow bridging the information from $\mathcal{G}^B$ and $\mathcal{G}^N$. In [8] the authors show this to be the case in the grim-trigger and tit-for-tat strategies of repeated games, where the values of $\mathcal{G}^B$ play a crucial role.

In what follows, we use the decomposition to analyze the structure of potential games, or "common interest" games. We begin with 2×2 games before moving on to $2 \times 2 \times 2$. Once certain structures emerge, we extend the results to n-agent $2 \times \ldots \times 2$ games. Throughout this exposition, we keep in mind the externalities, and show that a specific alignment of the $\mathcal{G}^N$ and $\mathcal{G}^B$ constitutes identical play games.

To summarize this section, the Nash decomposition induces the following change of basis $e_1^N = e_1 - e_5$, $e_2^N = e_3 - e_7$, $e_3^N = e_2 - e_4$, $e_4^N = e_6 - e_8$, $e_1^B = e_1 - e_3 + e_5 - e_7$, $e_2^B = e_2 + e_4 - e_6 - e_8$, $e_1^K = e_1 + e_3 + e_5 + e_7$, and $e_2^K = e_2 + e_4 + e_6 + e_8$. A game $\mathcal{G}$, then, decomposed into $\mathcal{G}^N$, $\mathcal{G}^B$, and $\mathcal{G}^K$, with payoff values given in Tables 2.4, 2.5, 2.3, can be written as the vector $(\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}, \beta_1, \beta_2, \kappa_1, \kappa_2)$. Alternatively, we can write,

$$\mathcal{G} = \alpha_{11} e_1^N + \alpha_{12} e_2^N + \alpha_{21} e_3^N + \alpha_{22} e_4^N + \beta_1 e_1^B + \beta_2 e_2^B + \kappa_1 e_1^K + \kappa_2 e_2^K.$$

2.3 2×2 Potential Structure

2.3.1 Introduction

For a game $\mathcal{G}$ to belong to the class of exact potential games, $\mathcal{G}$ must admit a potential function. A potential function is a global function that keeps track of all agents' payoffs under unilateral deviations. In other words, for a given a strategy profile σ, the payoff change of an agent for their unilateral deviation from σ is the same as the change in the potential function. Let us make this precise.

Definition 2.3. Consider a game $\mathcal{G} = \langle N, S_1, S_2, \pi_1, \pi_2 \rangle$, where $N = \{1, 2\}$ is the set of agents, and, for $i \in N$, $S_i = \{A, B\}$ is agent i's strategy set, and $\pi_i : S_1 \times S_2 \to \mathbb{R}$ is i's utility function. The game $\mathcal{G}$ is an *exact potential game* if there exists a function $P : S_1 \times S_2 \to \mathbb{R}$, called a potential function, such that,

1. $P(A, \sigma_2) - P(B, \sigma_2) = \pi_1(A, \sigma_2) - \pi_1(B, \sigma_2)$
2. $P(\sigma_1, A) - P(\sigma_1, B) = \pi_2(\sigma_1, A) - \pi_2(\sigma_1, B)$

where σ_1 and σ_2 are arbitrary strategies for agents 1 and 2, respectively.

In the literature this definition is sometimes extended to consider *weighted potential games* and *ordinal potential games*. A weighed potential game has a vector of positive weights that scales the differences in unilateral deviations for each agent. An ordinal potential game, discussed later, means that the potential function only preserves information pertaining to the sign of the payoff difference. This dissertation will not consider weighted potential games; later we will describe the necessary adjustments to our results so that they extend to ordinal potential games.

The potential function is often praised for its properties in detailing the Nash structure of the potential game. For instance, the potential function is maximized at the Nash equilibria. Any unilateral deviation from a Nash equilibrium will give the deviating agent a smaller payoff; hence a negative difference in payoffs. By definition of the potential function, the change in the potential function from the Nash equilibrium profile must also be negative. Therefore the potential function must be at a maximum since any deviation by any agent causes the potential function to decrease in value.

Mapping the agents' strategies to numerical values, we construct payoff functions for the agents. These payoff functions are used throughout this thesis. In this chapter, we use them to construct the general potential function for a given potential game.

2.3.2 Utility Functions from the Decomposition

The discussion so far on potential games and the $\mathcal{G}^N, \mathcal{G}^B$, and $\mathcal{G}^K$ components described at the beginning of the chapter, hints at strong connection between the decomposition and the potential function. This is because the important quality of a potential function is its reflection of *unilateral* changes in the game. The potential function must be connected, then, to the $\mathcal{G}^N$ component of a potential game $\mathcal{G}$. This is true, and this fact will permit deriving several new conclusions. First, let us use the decomposition to build payoff functions for the decomposed game $\mathcal{G}$ shown in Table 2.6.

Define the function $t : S_1 \times S_2 \to \{-1, +1\} \times \{-1, +1\}$ such that for $\sigma \in S$, $t(\sigma) = t(\sigma_1, \sigma_2)$ $= (t(\sigma_1), t(\sigma_2))$, and $t(A) = +1$ and $t(B) = -1$. In other words, the strategy A is mapped to the value of $+1$ for both agents, and the strategy B is mapped to -1. As in Table 2.6, we will now refer to any strategy pair (σ_1, σ_2) by the pair $(t_1, t_2) = (t(\sigma_1), t(\sigma_2))$. With this, we can summarize the complete payoff structure of the game using the payoff functions defined in Theorem 2.4.

	+1		−1	
+1	κ_1	κ_2	κ_1	κ_2
−1	κ_1	κ_2	κ_1	κ_2

Kernel

	+1		−1	
+1	α_{11}	α_{21}	α_{12}	$-\alpha_{21}$
−1	$-\alpha_{11}$	α_{22}	$-\alpha_{12}$	$-\alpha_{22}$

Nash

	+1		−1	
+1	β_1	β_2	$-\beta_1$	β_2
−1	β_1	$-\beta_2$	$-\beta_1$	$-\beta_2$

Behavioral

Table 2.6: Decomposed Game $\mathcal{G}$

Theorem 2.4. *A game $\mathcal{G} = \mathcal{G}^N + \mathcal{G}^B + \mathcal{G}^K$, as written in normal-form with separated components in Table 2.6, can be represented by payoff functions*

$$\pi_i(t_1, t_2) = \frac{\alpha_{i1} + \alpha_{i2}}{2} t_i + \frac{\alpha_{i1} - \alpha_{i2}}{2} t_i t_{\neg i} + \beta_i t_{\neg i} + \kappa_i \tag{2.1}$$

where $\neg i \in \{1, 2\} \backslash \{i\}$, for each agent $i = 1, 2$.

Proof. The four corners of the game matrix represented by strategy profiles (A, A), (A, B), (B, A), and (B, B) get mapped, respectively, to $(+1, +1), (+1, -1), (-1, +1)$, and $(-1, -1)$. At each of the points the value of the function for each agent is exactly the value found in $\mathcal{G}$. □

These functions are almost unique. The uniqueness of the coefficients comes from the uniqueness of the decomposition of $\mathcal{G}$. However, notice that replacing the t_i with t_i^m for odd m describes the same payoff structure. By restricting the exponent of the t_i to 1 for $i = 1, 2$, then the functions are unique. To see this, simply take a polynomial in t_1 and t_2 up to power

1, and show that its coefficients are exactly those given in Theorem 2.4. Throughout this dissertation we always assume that the t_i variables have a maximum exponent of 1.

Let us now associate components of this function with the components of the game $\mathcal{G}$.

Definition 2.5. Define the Nash component of $\pi_i(t_1, t_2)$ to be

$$\pi_i^N(t_1, t_2) = \frac{\alpha_{i1} + \alpha_{i2}}{2} t_i + \frac{\alpha_{i1} - \alpha_{i2}}{2} t_i t_{\neg i}.$$

Similarly define the behavioral component of $\pi_i(t_1, t_2)$ to be

$$\pi_i^B(t_1, t_2) = \beta_i t_{\neg i}.$$

Finally, define the kernel component of $\pi_i(t_1, t_2)$ to be

$$\pi_i^K(t_1, t_2) = \kappa_i.$$

Let us now build a potential function using the agents' utility functions. We first state a theorem without proof and refer the reader to [8].

Theorem 2.6. *$\mathcal{G}$ is an exact potential game if and only if $\alpha_{11} - \alpha_{12} = \alpha_{21} - \alpha_{22}$.*

Let us call this difference d.

Theorem 2.7. *If G is an exact potential game, then an appropriate potential function is*

$$P(t_1, t_2) = \frac{\alpha_{11} + \alpha_{12}}{2} t_1 + \frac{\alpha_{21} + \alpha_{22}}{2} t_2 + \frac{d}{2} t_1 t_2 + c$$

where $d = \alpha_{11} - \alpha_{12} = \alpha_{21} - \alpha_{22}$ and $c \in \mathbb{R}$ is an arbitrary constant.

Proof. Going through the calculations shows that $P(A, t_2) - P(B, t_2) = \pi_1(A, t_2) - \pi_1(B, t_2)$ and that $P(t_1, A) - P(t_1, B) = \pi_2(t_1, A) - \pi_2(t_1, B)$. □

This gives the immediate corollary,

Corollary 2.7.1. *If $\mathcal{G}$ is an exact potential game, then its potential function depends only on π_i^N for each $i \in N$.*

Proof. The proof of this is immediate from Theorem 2.7. The parameters of the potential function of $\mathcal{G}$ depend only on the payoff values in $\mathcal{G}^N$. □

An implication of this is that varying the values of β_1 and β_2 (and κ_1 and κ_2) has no effect on the potential function. Although the potential function captures the Nash structure of a game, it paints an incomplete picture of the game as a whole. The remainder of this picture lies in the $\mathcal{G}^B$ component, on which the potential function makes no restrictions. Because of this, we focus on the Nash structure $\mathcal{G}^N$ first, before discussing the $\mathcal{G}^B$ structure. At this point, it is worth noting that the space of potential games is seven dimensional. This is because of the eight dimensions for the space of games, only the constraint (Thm. 2.6) is imposed on the Nash structure. Already, this fact suggests how the decomposition permits more general conclusions. In his influential paper [19], Young for example, concentrates on a one-dimensional subspace. As such, generalizations can be, and are, offered in what follows.

We now state another corollary to Theorem 2.7.

Corollary 2.7.2. *If $\mathcal{G}^N = \tilde{\mathcal{G}}^N$ for two exact potential games $\mathcal{G}$ and $\tilde{\mathcal{G}}$, then, restricting the exponents of t_i for $i = 1, 2$ to 1, their potential functions differ only by a constant.*

Proof. Suppose $\mathcal{G}^N = \tilde{\mathcal{G}}^N$ for two exact potential games $\mathcal{G}$ and $\tilde{\mathcal{G}}$. Then the payoff values are the same in $\mathcal{G}^N$ and $\tilde{\mathcal{G}}^N$. This means $\pi_1^N(t_1, t_2) = \tilde{\pi_1}^N(t_1, t_2)$ and $\pi_2^N(t_1, t_2) = \tilde{\pi_2}^N(t_1, t_2)$. Since an exact potential game's potential function depends only on the Nash component of the payoff functions, and because the exponents of t_i for $i = 1, 2$ are bounded by 1, we have that $P(t_1, t_2) = \tilde{P}(t_1, t_2)$. Therefore, $\mathcal{G}$ and $\tilde{\mathcal{G}}$ have the same potential function. □

As stated in the previous subsection, two seemingly different games with the same underlying $\mathcal{G}^N$ component result in the same predictions offered by several solution concepts, such as the Nash equilibrium. We now see that the same is true for potential games and their corresponding potential functions. If two seemingly different exact potential games have the same $\mathcal{G}^N$ structure, then, up to a constant, they have the same potential function.

The restriction potential games place on a game's Nash structure, described in Theorem 2.6, means a degree of freedom is lost, and the dimensionality of the Nash structure is reduced from 4 to 3. We explore this further by changing the basis of $\mathcal{G}^N$ to reflect the structure of the potential function. In doing so, we are able to partition the space $\mathbb{G}^N$ into the 3-dimensional Nash subspace of potential games, and its orthogonal complement, $\mathbb{G}_P^N$ and $(\mathbb{G}_P^N)^\perp$, respectively.

2.3.3 New Basis for Utility Functions[3]

As shown in Corollary 2.7.1, the potential function of a game $\mathcal{G}$ depends only on its Nash component $\mathcal{G}^N$. We also mentioned, in Theorem 2.6, that $\alpha_{11} - \alpha_{12} = \alpha_{21} - \alpha_{22}$. This means that one of the values can be written as a linear combination of the other three independent values. In other words, there is a Potential Nash subspace of $\mathbb{G}^N$, which we will denote by $\mathbb{G}_P^N$, such that $\dim \mathbb{G}_P^N = 3$. Moreover, $\mathbb{G}^N = \mathbb{G}_P^N \oplus (\mathbb{G}_P^N)^\perp$ where $\dim(\mathbb{G}_P^N)^\perp = 1$. Let us find the basis for $\mathbb{G}_P^N$ and for $(\mathbb{G}_P^N)^\perp$.

The canonical basis for $\mathbb{G}^N$ is given in Table 2.7.

This is the basis developed by Jessie and Saari in [8]. Notice how we can also represent this basis using vectors, namely, $e_1^N = (1,0,0,0)$, $e_3^N = (0,1,0,0)$, $e_2^N = (0,0,1,0)$, and $e_4^N = (0,0,0,1)$. Here, for the coordinate (x_1, x_2, x_3, x_4) we have that x_1 is agent 1's payoff for the strategy profile $(+1,+1)$, x_2 is agent 1's payoff for the strategy profile $(+1,-1)$, x_3 is

[3]Heavily influenced by discussions in the weekly research seminars with Don Saari.

Table 2.7: Basis for the Space of 2×2 $\mathcal{G}^N$ Components

$e_1^N =$

	+1	−1
+1	1, 0	0, 0
−1	−1, 0	0, 0

$e_3^N =$

	+1	−1
+1	0, 0	1, 0
−1	0, 0	−1, 0

$e_2^N =$

	+1	−1
+1	0, 1	0, −1
−1	0, 0	0, 0

$e_4^N =$

	+1	−1
+1	0, 0	0, 0
−1	0, 1	0, −1

agent 2's payoff for strategy profile $(+1, +1)$, and finally x_4 is agent 2's payoff the strategy profile $(-1, +1)$. This is enough to generate all payoffs in the Nash component of a 2×2 game $\mathcal{G}$ since all remaining entries are simply negatives of their unilateral counterparts. For example, agent 1's payoff for strategy profile $(-1, +1)$ is simply the negative of their payoff in $(+1, +1)$. This is due to the symmetry structure inherent to the space $\mathbb{G}^N$.

Now, for any $\mathcal{G}^N \in \mathbb{G}^N$, we can find unique coefficients $\alpha_{11}, \alpha_{12}, \alpha_{21}$, and α_{22}, so that $\mathcal{G}^N = \alpha_{11}e_1^N + \alpha_{12}e_3^N + \alpha_{21}e_2^N + \alpha_{22}e_4^N$.

Since the functional forms for π_1 and π_2 are polynomials in t_1 and t_2, a change of basis can be performed so that the constants in front of t_1, t_2, and $t_1 t_2$ are elements of the new basis. In other words, for agent i we would like to find α_i, not to be confused with α_{ij}, and $\gamma_{i\neg i}$ such that $\alpha_i = \frac{\alpha_{i1}+\alpha_{i2}}{2}$ and $\gamma_{i\neg i} = \frac{\alpha_{i1}-\alpha_{i2}}{2}$.

Solving for α_{i1} in both equations gives $\alpha_{i1} = 2\alpha_i - \alpha_{i2}$ and $\alpha_{i1} = 2\gamma_{i\neg i} + \alpha_{i2}$. Adding them and diving by 2 gives $\alpha_{i1} = \alpha_i + \gamma_{i\neg i}$. Substituting this α_{i1} and solving for α_{i2} in either of the previous two equations for α_{i1} gives $\alpha_{i2} = \alpha_i - \gamma_{i\neg i}$. We now substitute all $\alpha_{11}, \alpha_{12}, \alpha_{21}$, and α_{22} in $\mathcal{G}^N = \alpha_{11}e_1^N + \alpha_{12}e_3^N + \alpha_{21}e_2^N + \alpha_{22}e_4^N$, giving

$$\mathcal{G}^N = (\alpha_1 + \gamma_{12})e_1^N + (\alpha_1 - \gamma_{12})e_2^N + (\alpha_2 + \gamma_{21})e_3^N + (\alpha_2 - \gamma_{21})e_4^N.$$

Rearranging,

$$\mathcal{G}^N = \alpha_1(e_1^N + e_3^N) + \gamma_{12}(e_1^N - e_3^N) + \alpha_2(e_2^N + e_4^N) + \gamma_{21}(e_2^N - e_4^N).$$

This suggests defining the vectors $n_1 = e_1^N + e_3^N = (1, 0, 1, 0)$, $n_3 = e_1^N - e_3^N = (1, 0, -1, 0)$, $n_2 = e_2^N + e_4^N = (0, 1, 0, 1)$, and $n_4 = e_2^N - e_4^N = (0, 1, 0, -1)$. It is easy to verify that n_1, n_2, n_3, and n_4 is an orthogonal set of vectors. Furthermore, since there are four of them, these vectors form an orthogonal basis for $\mathbb{R}^4$, the space of $\mathbb{G}^N$. We can now write

$$\mathcal{G}^N = \alpha_1 n_1 + \gamma_{12} n_3 + \alpha_2 n_2 + \gamma_{21} n_4.$$

These vectors are shown in normal-form in Table 2.8.

Table 2.8: A New Basis for $\mathbb{G}^N$

$n_1 = e_1^N + e_3^N =$

	+1	−1
+1	1, 0	1, 0
−1	−1, 0	−1, 0

$n_3 = e_1^N - e_3^N =$

	+1	−1
+1	1, 0	−1, 0
−1	−1, 0	1, 0

$n_2 = e_2^N + e_4^N =$

	+1	−1
+1	0, 1	0, −1
−1	0, 1	0, −1

$n_4 = e_2^N - e_4^N =$

	+1	−1
+1	0, 1	0, −1
−1	0, −1	0, 1

The derivation of the basis started with taking $\alpha_i = \frac{\alpha_{i1}+\alpha_{i2}}{2}$. Hence, the value α_i is agent i's average payoff for choosing strategy $+1$. Consequently, the value $-\alpha_i$ is agent i's average payoff for choosing strategy -1. We see that this value is obtained by the agent independent of what is played by the others. For this reason, α_i is interpreted to be agent i's inherent preference of $+1$ over -1. We will use the terms *individual preference* and *personal preference* to refer to the α_i parameters.

On the other hand, we see that if γ_{12} is positive, it describes how much agent 1 gains from coordinating with agent 2. Conversely, if γ_{12} is negative, it describes how much agent 1 gains from not coordinating with agent 2. The same argument is true for γ_{21} with agent 1 and 2 swapped. In other words, interpret $\gamma_{i\neg i}$ to be the benefit, or detriment, of agent i coordinating with agent $\neg i$. We refer to this as the *coordinative pressure* of the game. We will also use the terminology *pressure to conform* to refer to these coordinates.

To reiterate, this new basis describes the entire space of $\mathbb{G}^N$ using two vectors for each player. One of these vectors, α_i, describes agent i's inherent preference over A and B, where α_1 scales n_1, and that α_2 scales n_2. The other vector, $\gamma_{i\neg i}$, describes if, and how much, agent i wants to coordinate with agent $\neg i$, where $\neg i$ identifies the agent who is not agent i. Here we have that γ_{12} scales n_3 and that γ_{21} scales n_4.

$$\mathcal{G}^N = \alpha_1 n_1^* + \gamma_{12} n_2^* + \alpha_2 n_3^* + \gamma_{21} n_4^*.$$

Moreover, we can now represent the functional forms given in Theorem 2.4 using the new basis. We have $\pi_i^N(t_1, t_2) = \alpha_i t_i + \gamma_{i\neg i} t_i t_{\neg i}$, where, again, $\neg i$ identifies the agent who is not agent i. We remind the reader that before we had $\pi_i^N(t_1, t_2) = \frac{\alpha_{i1}+\alpha_{i2}}{2} t_i + \frac{\alpha_{i1}-\alpha_{i2}}{2} t_i t_{\neg i}$.

Lastly, we rewrite Theorem 2.6 in terms of this basis.

Theorem 2.8. *$\mathcal{G}$ is an exact potential game if and only if $\gamma_{12} = \gamma_{21}$.*

Proof. The result follows immediately from Theorem 2.6 together with the fact that $\gamma_{12} = \alpha_{11} - \alpha_{12}$ and $\gamma_{21} = \alpha_{21} - \alpha_{22}$. □

Theorem 2.8 shows that the requirement for a game $\mathcal{G}$ to be an exact potential game becomes $\gamma_{12} = \gamma_{21}$. We denote this common value by γ and henceforth any use of γ implies $\gamma = \gamma_{12} = \gamma_{21}$. There are no restrictions placed on the individual preferences, and, of course, no restrictions placed on the externalities and the kernel component (since they are not picked up by the potential function in the first place). We see, then, that for a game $\mathcal{G}$ to be an exact potential game, the agents both need only have the same pressure to conform, $\gamma > 0$, or to not conform, $\gamma < 0$.

This allows us to rewrite the potential function given in Theorem 2.7.

Theorem 2.9. *A potential function for $\mathcal{G}$ is given by*

$$P(t_1, t_2) = \alpha_1 t_1 + \alpha_2 t_2 + \gamma t_1 t_2 + c$$

where $\gamma = \gamma_{12} = \gamma_{21}$ and $c \in \mathbb{R}$ is an arbitrary constant.

We are now ready to discuss the bases for $\mathbb{G}_P^N$ and $(\mathbb{G}_P^N)^\perp$.

2.3.4 A Basis for $\mathbb{G}_P^N$ and $(\mathbb{G}_P^N)^\perp$

We know that a game $\mathcal{G}$ can be written as $\mathcal{G} = \mathcal{G}^N + \mathcal{G}^B + \mathcal{G}^K$ where $\dim \mathbb{G}^N = 4$, $\dim \mathbb{G}^B = 2$, and $\dim \mathbb{G}^K = 2$. In other words, a game $\mathcal{G}$ has 8 degrees of freedom; 4 for the Nash component, 2 for the behavioral component, and 2 for the kernel. For this game $\mathcal{G}$ to be an exact potential game, we need $\gamma_{12} = \gamma_{21}$. That is, two independent parameters collapse into one, or, the dimensionality of the Nash subspace of potential games is reduced by 1.

More precisely, denote the potential subspace of $\mathbb{G}$ by $\mathbb{G}_P$, and its Nash subspace by $\mathbb{G}_P^N$. Then $\dim \mathbb{G}_P = 7$ and $\dim \mathbb{G}_P^N = 3$. Furthermore, for an exact potential game $\mathcal{G} \in \mathbb{G}_P$ we can write its Nash component $\mathcal{G}^N \in \mathbb{G}_P^N$ as $\mathcal{G}^N = \alpha_1 n_1 + \alpha_2 n_2 + \gamma(n_3 + n_4)$. Letting $n_\gamma = n_3 + n_4$, this becomes $\mathcal{G}^N = \alpha_1 n_1 + \alpha_2 n_2 + \gamma n_\gamma$. It is clear that n_γ is orthogonal to both n_1 and n_2 since n_γ is composed of a linear combination of n_3 and n_4, both of which are orthogonal to n_1 and n_2. Then, since $\dim \mathbb{G}_P^N = 3$, it is immediate that $\mathbb{G}_P^N = \text{span}\{n_1, n_2, n_\gamma\}$.

Since $\dim \mathbb{G}^N = 4$, if we find a vector $\widetilde{n}$ such that $\{n_1, n_2, n_\gamma, \widetilde{n}\}$ is a linearly independent set, then we have that $\mathbb{G}^N = \text{span}\{n_1, n_2, n_\gamma, \widetilde{n}\}$. The implication of this is that we can partition the Nash space of games into a potential subspace, $\mathbb{G}_P^N$, and a, yet to be interpreted, "anti-potential" subspace, denoted by $(\mathbb{G}_P^N)^\perp$, *i.e.*, $\mathbb{G}^N = \mathbb{G}_P^N \oplus (\mathbb{G}_P^N)^\perp$. Finding this vector $\widetilde{n}$ will illuminate the structure of the anti-potential space.

Because $n_\gamma = n_3 + n_4$, we can simply take $\widetilde{n} = n_3 - n_4$. The dot product of n_γ and $\widetilde{n}$ is clearly zero. In addition, since $\widetilde{n}$ is a linear combination of n_3 and n_4, this vector is orthogonal to both n_1 and n_2. Hence we now have

$$\mathbb{G}_P^N = \text{span}\{n_1, n_2, n_\gamma\} \tag{2.2}$$

$$(\mathbb{G}_P^N)^\perp = \text{span}\{\widetilde{n}\} \tag{2.3}$$

$$\mathbb{G}^N = \mathbb{G}_P^N \oplus (\mathbb{G}_P^N)^\perp = \text{span}\{n_1, n_2, n_\gamma, \widetilde{n}\}. \tag{2.4}$$

The point of view offered by linear algebra allows us to rephrase the conditions required for a game to be an exact potential game in terms of orthogonality. We capture this in the below Theorem 2.10.

Theorem 2.10. *Let $\mathcal{G}$ be a game. $\mathcal{G}$ is an exact potential game if and only if for every $\mathcal{G}_P^\perp \in (\mathbb{G}_P^N)^\perp$ we have that $\mathcal{G} \perp \mathcal{G}_P^\perp$.*

Proof. Let $\mathcal{G}$ be a game. Consider its Nash component $\mathcal{G}^N = \alpha_1 n_1 + \alpha_2 n_2 + \gamma_{12} n_3 + \gamma_{21} n_4$.

Suppose that $\mathcal{G}$ is an exact potential game. Then $\gamma_{12} = \gamma_{21}$, which we denote by γ. Hence $\mathcal{G}^N = \alpha_1 n_1 + \alpha_3 n_2 + \gamma n_\gamma$. Let $\mathcal{G}_P^\perp \in (\mathbb{G}_P^N)^\perp$. Then we can write $\mathcal{G}_P^\perp = \bar{\gamma}\widetilde{n}$ for some $\bar{\gamma} \in \mathbb{R}$. One can easily verify that $\mathcal{G} \cdot \mathcal{G}_P^\perp = 0$. Hence, $\mathcal{G} \perp \mathcal{G}_P^\perp$.

Now, suppose that for every $\mathcal{G}_P^\perp \in (\mathbb{G}_P^N)^\perp$ we have $\mathcal{G} \perp \mathcal{G}_P^\perp$. Write $\mathcal{G}^N = \alpha_1 n_1 + \alpha_2 n_2 + \gamma n_\gamma + \bar{\gamma}\widetilde{n}$. Furthermore, write $\mathcal{G}_P^\perp = \bar{\gamma}'\widetilde{n}$. Since $\mathcal{G} \perp \mathcal{G}_P^\perp$, we have that $\mathcal{G} \cdot \mathcal{G}_P^\perp = 0$. This means that $\bar{\gamma} \cdot \bar{\gamma}' = 0$. Since $\bar{\gamma}'$ is arbitrary, we must have $\bar{\gamma} = 0$ for this equality to hold in general. This implies $\mathcal{G}$ is an exact potential game. □

In order to fully understand the implications of Theorem 2.10, let us further explore the vectors n_γ and $\widetilde{n}$. Consider the basis vector n_γ given in Table 2.9.

		+1	−1
	+1	1, 1	−1, −1
$n_\gamma =$	−1	−1, −1	1, 1

Table 2.9: The Basis Vector n_γ

It is immediate that the vector n_γ is a fully symmetric coordination game. From here we can think of the Nash component of an exact potential game in a new way. We know we can write the Nash component of an exact potential game $\mathcal{G}$ as $\mathcal{G}^N = \alpha_1 n_1 + \alpha_2 n_2 + \gamma n_\gamma$. The quantities α_1 and α_2 tell us how much each agent invariably prefers +1 or −1. The value γ tells us the equal value both agents receive from coordinating (or not coordinating if $\gamma < 0$).

Let us now explore the class of games that are orthogonal to exact potential games. For these games the Nash component can be written as $\mathcal{G}^N = b \cdot \widetilde{n}$, for some $b \in \mathbb{R}$, where $\widetilde{n}$ is shown in Table 2.10.

Immediately we see that the basis vector $\widetilde{n}$ has the form of a fully symmetric matching pennies game. We state two theorems regarding this.

		+1	−1
	+1	1, −1	−1, 1
$\widetilde{n} =$	−1	−1, 1	1, −1

Table 2.10: The Basis Vector $\widetilde{n}$

Theorem 2.11. *The class of symmetric* 2×2 *matching pennies games is orthogonal to potential games.*

Proof. The proof follows immediately from the fact that the class of symmetric 2×2 matching pennies games is spanned by $\widetilde{n}$, which is orthogonal to all vectors spanning the class of potential games, namely n_1, n_2, and n_γ. □

Theorem 2.12. *Any game* $\mathcal{G} \in \mathbb{G}$ *with respect to the Nash structure can written uniquely as a sum of a potential game and a matching pennies games.*

Proof. This is an immediate consequence of $\mathbb{G}^N = \text{span}\{n_1, n_2, n_\gamma, \widetilde{n}\} = \text{span}\{n_1, n_2, n_\gamma\} \oplus \text{span}\{\widetilde{n}\}$. □

These results are directly linked to Candogan *et. al.*'s decomposition of games into a potential component, harmonic component, and nonstrategic component [5]. In our language these components are $\mathcal{G}_P^N$, $(\mathcal{G}_P^N)^\perp$, and $\mathcal{G}^B + \mathcal{G}^K$, respectively. Candogan et al describe that the potential component represents the common interest in a game, and the harmonic component of a game represents the conflict in interest. The second statement is clear from the structure of $(\mathcal{G}_P^N)^\perp$. The first statement deserves a deeper discussion, which we save for later in the chapter, after we have extended the theory to n-agent games.

2.3.5 Three Classes of Potential Games

The possible relationships between the parameters of the Nash component of potential games, α_1, α_2, and γ, give rise to three different classes of potential games. The first class is when the magnitudes of both α_1 and α_2 are greater than the magnitude of γ. In this class of potential games, the effect of γ is overpowered by α_1 and α_2 (notice that this class includes the case where $\gamma = 0$). Here, the agents simply play their individual preferences. We call these *independent* potential games.

Another class of potential games is where the magnitude of γ is between the magnitude of α_1 and α_2. In this case, the agent with the larger magnitude individual preference, α_1 or α_2, plays their preference. The other agent, receiving greater influence from γ, either wants to coordinate ($\gamma > 0$) or anti-coordinate ($\gamma < 0$) with the first agent, who is playing their preference. In other words, the first agent can play their preference and the second agent, despite what their preference might be, should follow (or go against) the first agent's choice. We call these *quasi-independent* potential games.

The third class of potential games is where the magnitude of γ is greater than the magnitudes of both α_1 and α_2. In this case both agents receive most of their influence from γ and hence want to coordinate with each other when $\gamma > 0$, and anti-coordinate when $\gamma < 0$. When $\gamma > 0$, this class of potential games may manifest as a Bach or Stravinsky game if the preferences are in opposing directions.[4] If the individual preferences are in the same direction, then the game is a coordination game with a superior equilibrium reflecting these preferences. When $\gamma < 0$ the game is an anti-coordination game. If the preferences are opposing, they will simply anti-coordinate on their preferences. If the agents have the same preferences, however, then only one will play their preference. These potential games are referred to as *dependent* potential games.

[4] In the literature this game is commonly referred to as a "Battle of the Sexes" game. We choose to use the name "Bach or Stravinsky" in an effort to disassociate the game from gender.

2.3.6 An Application: The 2×2 Congestion Game

The origin of potential games comes from Rosenthal's 1973 study of congestion games [14]. The basic idea behind a congestion game is that there are two roads connecting two cities. One of these roads is preferred by all agents until it is too crowded. Rosenthal showed that this game always has a pure strategy Nash equilibrium. Using the language of what has been developed so far it is easy to demonstrate Rosenthal's assumptions in terms of the individual preference parameters, α_i for $i = 1, 2$, and the common interest parameter γ.

The agents both have the same individual preference, which means that $\operatorname{sgn} \alpha_1 = \operatorname{sgn} \alpha_2$. The agents also both have the common interest of not taking the same road. This means that $\gamma < 0$. Let us analyze the congestion game for the classes of potential games defined in the previous section.

In the case of an independent potential game, despite not wanting to coordinate, the agents' preferences for the road dominates the pressure to not conform. Here, the unique pure Nash equilibrium is where both agents take the preferred road. In the case of a quasi-independent potential game, the unique pure Nash equilibrium is where the agent with the greater preference takes their desired road, and the other agent yields and takes the less preferred road. Lastly, in the case of a dependent potential game, there are two pure Nash equilibria, both where only one agent takes the preferred road.

The purpose of this short subsection is to show how the assumptions in common interest games can be directly observed in the parameters developed so far in this chapter.

2.3.7 Ordinal Potential Games

In this section we briefly extend what has been done to the more general class of ordinal potential games.

We already know that a game is a potential game when there is no component from $(\mathbb{G}_P^N)^\perp$. When there is a component from $(\mathbb{G}_P^N)^\perp$, however, the game may still have a strong essence of a potential game– it can be an ordinal potential game. We remind that reader that a game $\mathcal{G}$ is an ordinal potential game if and only if there exists a function that preserves the sign of the differences in unilateral deviations. Here we use a theorem by Voorneveld[5] that roughly says that a game $\mathcal{G}$ is an ordinal potential game if and only if it contains no weak improvement cycles.[6]

Hence it suffices for us to show that the matching pennies component of the game coming from $(\mathbb{G}_P^N)^\perp$ does not dominate the other components coming from $\mathbb{G}_P^N$ in such a way that it brings about unilateral deviation cycles. We will state this as a theorem.

Theorem 2.13. *A game $\mathcal{G}$ is an ordinal potential game if and only if the parameter $\widetilde{\gamma}$ representing the payoff from the $(\mathbb{G}_P^N)^\perp$ component of $\mathcal{G}$ is bounded as follows*

$$\min\{-|\alpha_1| - \gamma, -|\alpha_2| + \gamma\} < \widetilde{\gamma} < \max\{|\alpha_1| - \gamma, |\alpha_2| + \gamma\} \tag{2.5}$$

Proof. $\mathcal{G}$ is an ordinal potential game if and only if, according to the theorem by Voorneveld, $\mathcal{G}$ has no weak improvement cycle. A cycle exists if and only if one of the sets of inequalities in Table 2.11 is satisfied. The first and second sets of inequalities mean, respectively, that,

$\alpha_1 + \gamma + \widetilde{\gamma} \geq 0$		$\alpha_1 - \gamma - \widetilde{\gamma} \geq 0$
$-\alpha_1 + \gamma + \widetilde{\gamma} \geq 0$	or,	$-\alpha_1 - \gamma - \widetilde{\gamma} \geq 0$
$\alpha_2 - \gamma + \widetilde{\gamma} \geq 0$		$\alpha_2 + \gamma - \widetilde{\gamma} \geq 0$
$-\alpha_2 - \gamma + \widetilde{\gamma} \geq 0$		$-\alpha_2 + \gamma - \widetilde{\gamma} \geq 0$

Table 2.11: Cycle-Inducing Inequalities

[5]page 51 of https://pure.uvt.nl/ws/files/320704/voorneveld.pdf

[6]Cycles brought about by unilateral deviations, including cases where the unilateral difference between payoffs is zero.

$$\widetilde{\gamma} \geq \max\{|\alpha_1| - \gamma, |\alpha_2| + \gamma\} \tag{2.6}$$

$$\widetilde{\gamma} \leq \min\{-|\alpha_1| - \gamma, -|\alpha_2| + \gamma\}. \tag{2.7}$$

The complement of these regions is exactly the inequalities demanded in the theorem. Hence, $\mathcal{G}$ is an ordinal potential game if and only if $\widetilde{\gamma}$ is outside of the regions in (2.6) and (2.7), or, if and only if

$$\min\{-|\alpha_1| - \gamma, -|\alpha_2| + \gamma\} < \widetilde{\gamma} < \max\{|\alpha_1| - \gamma, |\alpha_2| + \gamma\}.$$

□

The class of ordinal potential games provides an important relaxation to the strict requirement of exact potential games. In an exact potential game, the shared γ parameter must be equal for both agents. In a perhaps overly strict way, this is capturing the idea of "common interest" between the agents. It is reasonable that this parameter need not be equal for the agents to still have a "common interest." This possibility is captured in ordinal potential games.

2.3.8 Discussion

In this section we explored the structure of 2×2 potential games. We saw that for a game to be an exact potential game its payoff structure cannot contain a matching pennies substructure, however small. Well behaved matching pennies components still preserve many aspects of exact potential games, which extends the space of allowed games to the class of ordinal potential games. Because potential games depend only on the Nash structure of a game, we did not include the externality component in our analysis. After extending these results to $2 \times 2 \times 2$ games, and then to $2 \times \ldots \times 2$ games, we discuss the externality structure of potential games. In the end of the chapter we identify a special relationship between potential games and identical play games, a relationship that depends on an alignment between the Nash and behavioral components. Let us now proceed to study $2 \times 2 \times 2$ potential games.

2.4 $2 \times 2 \times 2$ Potential Structure

2.4.1 Emergent Structures in the Utility Functions

The basis developed in the previous section naturally extends to situations with more agents. We skip the steps in the development of this since the process follows the same procedure as before, and because we prove theorems about this for n agents later in the chapter.

After computations we arrive at the following Nash components of the agents' payoff functions in a $2 \times 2 \times 2$ game,

$$\pi_1^N = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \gamma_{13} t_1 t_3 + \delta_{123} t_1 t_2 t_3$$
$$\pi_2^N = \alpha_2 t_2 + \gamma_{21} t_1 t_2 + \gamma_{23} t_2 t_3 + \delta_{231} t_1 t_2 t_3$$
$$\pi_3^N = \alpha_3 t_3 + \gamma_{31} t_1 t_3 + \gamma_{32} t_2 t_3 + \delta_{312} t_1 t_2 t_3.$$

This immediately reveals some emergent structures in the Nash component of the agents' utility functions. Like before, for $i, j \in N, i \neq j$, we have γ_{ij} parameters influencing the coordination structure of $\mathcal{G}$. This interaction is occuring between all pairs of agents, since each pair $i, j \in N$ has a γ_{ij} parameter. The same is true in the 2 agent case, of course, where the number of pairs is 1.

This structure in normal-form, with all γ_{ij} added for $i, j \in N$, $i \neq j$, is displayed in Table 2.12. What is clear from both the utility functions and normal-form representation of the

γ structure is that the γ_{ij} parameter for agents i and j has no dependence on agent k, for $i, j, k \in N$, $i \neq j, i \neq k, j \neq k$. Because of this we see that this is purely a coordination or anti-coordination force, depending on the sign, between agent i and j.

This structure is incredibly intuitive. For example, if agent 1 has a positive pressure to conform with agent 2, but wants to anti-coordinate with agent 3, the ideal profile for agent 1 is of the form $(t, t, -t)$. In any other profile, agent 1 must weigh the benefit of conforming with agent 2 against the cost of conforming with agent 3, or the cost of not conforming with agent 2 against the benefit of not conforming with agent 3. This opens up modeling in game theory to not only allow for more intricate assumptions, but to provide a mathematical framework in which to keep track of and analyze the consequences of assumptions!

Player 3: $+1$

	$+1$			-1		
$+1$	$\gamma_{12}+\gamma_{13}$	$\gamma_{21}+\gamma_{23}$	$\gamma_{31}+\gamma_{32}$	$-\gamma_{12}+\gamma_{13}$	$-\gamma_{21}-\gamma_{23}$	$\gamma_{31}-\gamma_{32}$
-1	$-\gamma_{12}-\gamma_{13}$	$-\gamma_{21}+\gamma_{23}$	$-\gamma_{31}+\gamma_{32}$	$\gamma_{12}-\gamma_{13}$	$\gamma_{21}-\gamma_{23}$	$-\gamma_{31}-\gamma_{32}$

$+1$

	$+1$			-1		
$+1$	$\gamma_{12}-\gamma_{13}$	$\gamma_{21}-\gamma_{23}$	$-\gamma_{31}-\gamma_{32}$	$-\gamma_{12}-\gamma_{13}$	$-\gamma_{21}+\gamma_{23}$	$-\gamma_{31}+\gamma_{32}$
-1	$-\gamma_{12}+\gamma_{13}$	$-\gamma_{21}-\gamma_{23}$	$\gamma_{31}-\gamma_{32}$	$\gamma_{12}+\gamma_{13}$	$\gamma_{21}+\gamma_{23}$	$\gamma_{31}+\gamma_{32}$

-1

Table 2.12: Emergent Structure in $2 \times 2 \times 2$ Games: γ

Another emergent structure in a $2 \times 2 \times 2$ game is the one composed of the δ_{ijk} parameters, for $i, j, k \in N, i \neq j, i \neq k, j \neq k$. For clarity, we describe how the δ indices should be interpreted. Given δ_{ijk} for $i, j, k \in N, i \neq j, i \neq k, j \neq k$, we have that i denotes the player whose utility function includes this δ parameter, whereas j and k represent other agents in

N. The order of the other agents does not matter, *i.e.*, $\delta_{123} = \delta_{132}$, and both simply mean that this is agent 1's δ parameter in relation to agents 2 and 3. In other words, the order of the first index matters but the other two are permutable. As a convention, in this thesis, we use δ_{123}, δ_{231}, and δ_{312}, for agents 1, 2, and 3, respectively.

The question is to interpret the δ-structure. For clarity, this structure is given in normal-form in Table 2.13. To begin our analysis, notice that if all δ parameters are positive, this component has Nash equilibria at the profiles $(+1, +1, +1)$, $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$. On the other hand, if they are all negative, we have equilibria at $(-1, -1, -1)$, $(-1, +1, +1)$, $(+1, -1, +1)$, and $(+1, +1, -1)$. Furthermore, we notice that if there is any difference between the signs of δ_{123}, δ_{231}, and δ_{312}, then there are weak-improvement cycles. We can already tell that the restriction of potential games will prevent this.

	$+1$	-1
$+1$	δ_{123} δ_{231} δ_{312}	$-\delta_{123}$ $-\delta_{231}$ $-\delta_{312}$
-1	$-\delta_{123}$ $-\delta_{231}$ $-\delta_{312}$	δ_{123} δ_{231} δ_{312}

$+1$

	$+1$	-1
$+1$	$-\delta_{123}$ $-\delta_{231}$ $-\delta_{312}$	δ_{123} δ_{231} δ_{312}
-1	δ_{123} δ_{231} δ_{312}	$-\delta_{123}$ $-\delta_{231}$ $-\delta_{312}$

-1

Table 2.13: Emergent Structure in $2 \times 2 \times 2$ Games: δ

It is common in the literature to interpret 2-strategy games as binary option games. For example, one may choose to vote yes or vote no, or one may wear their seatbelt or not wear their seatbelt. This component, when all parameters are of the same sign, is showing a pure Nash equilibrium structure where it is good for either all agents or only a single agent to choose a strategy. Alternatively, it could be good for either no agents or exactly two agents to choose a strategy. This exact pure Nash equilibrium structure is behind the legislator game, which we analyze after developing the potential structure of $2 \times 2 \times 2$ games. Let us proceed in doing this.

2.4.2 Potential Structure

Following the same procedure as in the previous section, we arrive at the below potential function for a $2 \times 2 \times 2$ game,

$$P(t_1, t_2, t_3) = \alpha_1 t_1 + \alpha_2 t_2 + \alpha_3 t_3 + \gamma'_{12} t_1 t_2 + \gamma'_{13} t_1 t_3 + \gamma'_{23} t_2 t_3 + \delta t_1 t_2 t_3. \tag{2.8}$$

Similar to the 2×2 case, we have collapsed each pair of corresponding γ components, *e.g.*, γ_{12} and γ_{21}, turning 6 initial γ parameters into 3. We also collapsed the 3 δ parameters into 1. More precisely, we have $\gamma_{12} = \gamma_{21} = \gamma'_{12}$, $\gamma_{13} = \gamma_{31} = \gamma'_{13}$, $\gamma_{23} = \gamma_{32} = \gamma'_{23}$, and $\delta_{123} = \delta_{231} = \delta_{312} = \delta$. This reduces the original 12 dimensional Nash structure to 7 dimensions, namely, $\alpha_1, \alpha_2, \alpha_3, \gamma'_{12}, \gamma'_{13}, \gamma'_{23}$, and δ.

As before, we want to find the class of games orthogonal to the class of $2 \times 2 \times 2$ potential games. This entails finding a 5-dimensional subspace of $\mathbb{G}^N$ that is orthogonal to $\mathbb{G}^N_P$.

Each of the collapsed γ pairs produces a corresponding matching pennies game, giving three orthogonal matching pennies components. What is left to find are two vectors orthogonal to the collapsed δ component. We skip the details of this, as it is standard linear algebra involving cross products and dot products. The resulting two vectors together describe a 3-agent matching pennies game. In other words, there are three vectors describing 2×2 matching pennies games between all pairs of agents, and two vectors that together describe a $2 \times 2 \times 2$ generalized matching pennies game between all three agents. However, because of the non-intuitive nature of these vectors, we offer an alternative basis for the orthogonal class of games, which we will later generalize to n agents.

To develop this basis, we establish a standard with which we can write basis vectors and avoid writing too many normal-form $2 \times 2 \times 2$ game matrices. Consider an arbitrary Nash component of a $2 \times 2 \times 2$ game in normal-form as shown in Table 2.14. This game can be represented with the vector $(a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, c_1, c_2, c_3, c_4)$.

	$+1$	-1		$+1$	-1
$+1$	$a_1 \quad b_1 \quad c_1$	$a_2 \quad -b_1 \quad c_3$	$+1$	$a_3 \quad b_3 \quad -c_1$	$a_4 \quad -b_3 \quad -c_3$
-1	$-a_1 \quad b_2 \quad c_2$	$-a_2 \quad -b_2 \quad c_4$	-1	$-a_3 \quad b_4 \quad -c_2$	$-a_4 \quad -b_4 \quad -c_4$
	$+1$			-1	

Table 2.14: Arbitrary $2 \times 2 \times 2$ Nash Component

A potential game can then be written as

$$
\begin{aligned}
&\alpha_1(1,1,1,1,0,0,0,0,0,0,0,0) + \alpha_2(0,0,0,0,1,1,1,1,0,0,0,0)+\\
&\alpha_3(0,0,0,0,0,0,0,0,1,1,1,1) + \gamma'_{12}(1,-1,1,-1,1,-1,-1,1,0,0,0,0)+\\
&\gamma'_{13}(1,1,-1,-1,0,0,0,0,1,-1,-1,1) + \gamma'_{23}(0,0,0,0,1,1,-1,-1,1,1,-1,-1)+\\
&\delta(1,-1,-1,1,1,-1,-1,1,1,-1,-1,1)
\end{aligned}
$$

For a fixed strategy of any given agent, we can define a matching pennies game between the two remaining agents. In a $2 \times 2 \times 2$ game, this gives rise to 6 different matching pennies games. The vectors describing each of these matching pennies games are linearly dependent, and one can be eliminated. We write below 5 matching pennies games, omitting the one between agent 2 and 3 with agent 1's strategy fixed at -1.

One can verify that the below vectors are linearly independent to the potential structure basis vectors above, and to each other.

$$m_1 = (1, -1, 0, 0, -1, 1, 0, 0, 0, 0, 0, 0)$$

$$m_2 = (0, 0, 1, -1, 0, 0, -1, 1, 0, 0, 0, 0)$$

$$m_3 = (1, 0, -1, 0, 0, 0, 0, 0, -1, 1, 0, 0)$$

$$m_4 = (0, 1, 0, -1, 0, 0, 0, 0, 0, 0, -1, 1)$$

$$m_5 = (0, 0, 0, 0, 1, 0, -1, 0, -1, 0, 1, 0)$$

In the 2×2 case, we had matching pennies as the subspace orthogonal to the space of potential games. There this class of games was spanned by a single vector. Here we have 5 vectors that together span the space of generalized matching pennies, where each payoff is zero-sum. It makes sense that the games orthogonal to common interest games imply that one agent gains at the cost of the other(s).

We can state a theorem that verifies that this basis spans the generalized class of matching pennies game, which is done in the next section for n agents.

As in the 2×2 case, we can transcend the classical cycle arguments used to determine when games are not exact potential games. Here we simply need the presence of any combination of these matching pennies components. We state this as a theorem, before analyzing the Nash equilibrium structures in $2 \times 2 \times 2$ symmetric potential games.

Theorem 2.14. *A $2 \times 2 \times 2$ game $\mathcal{G}$ is an exact potential game if and only if the coefficients in front of m_i are zero for $i = 1, 2, 3, 4, 5$.*

Proof. The proof is immediate from the fact that each m_i for $i = 1, 2, 3, 4, 5$ is orthogonal to $\mathbb{G}^P$. □

2.4.3 Symmetric $2 \times 2 \times 2$ Games

For simplicity, in this section we consider the symmetric case of $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, and $\gamma'_{12} = \gamma'_{13} = \gamma'_{23} = \gamma$. This reduces the dimensionality of $2 \times 2 \times 2$ potential games from 7 to 3, affording us a more manageable exposition. The symmetric $2 \times 2 \times 2$ game is given in normal-form in Table 2.15.

	$+1$	-1	$+1$	-1
$+1$	$\alpha + 2\gamma + \delta$	$\alpha - \delta$	$\alpha - \delta$	$\alpha - 2\gamma + \delta$
-1	$-\alpha - 2\gamma - \delta$	$-\alpha + \delta$	$-\alpha + \delta$	$-\alpha + 2\gamma + \delta$
	$+1$		-1	

Table 2.15: Symmetric $2 \times 2 \times 2$ Game

Since this is a Nash component, we know that in order for a strategy profile to be a pure strategy Nash equilibrium, we need all payoffs to be positive. Hence, of concern are the regions between the planes $\alpha + 2\gamma + \delta = 0$, $\alpha - \delta = 0$, and $-\alpha + 2\gamma - \delta = 0$. These regions are given in Figure 2.1, where, without loss of generality, we take $\alpha > 0$. Here the horizontal line is $\delta = \alpha$, the line with positive slope is $\delta = -\alpha + 2\gamma$, and the line with negative slope is $\delta = -\alpha - 2\gamma$.

These regions generate payoff matrices of the forms shown in Tables 2.16 - 2.22.

	$+1$	-1	$+1$	-1
$+1$	$+,+,+$	$-,-,-$	$-,-,-$	$-,+,+$
-1	$-,-,-$	$+,+,-$	$+,-,+$	$+,+,+$
	$+1$		-1	

Table 2.16: Region 1; $|\alpha + \delta| < 2\gamma$ & $\delta > \alpha$

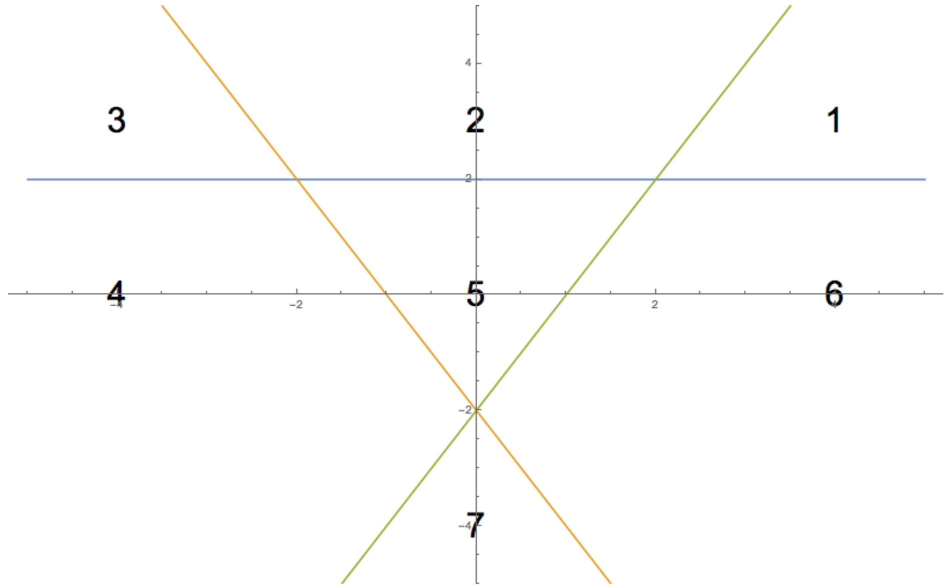


Figure 2.1: $\gamma\delta$-plane with $\alpha > 0$

	$+1$	-1	$+1$	-1
$+1$	$+,+,+$	$-,-,-$	$-,-,-$	$+,+,+$
-1	$-,-,-$	$+,+,+$	$+,+,+$	$-,-,-$
	$+1$		-1	

Table 2.17: Region 2; $\alpha + \delta > 2|\gamma|$ & $\delta > \alpha$

	$+1$	-1	$+1$	-1
$+1$	$-,-,-$	$-,+,-$	$-,-,+$	$+,+,+$
-1	$+,-,-$	$+,+,+$	$+,+,+$	$-,-,-$
	$+1$		-1	

Table 2.18: Region 3; $|\alpha + \delta| < -2\gamma$ & $\delta > \alpha$

Just as we did for the 2×2 case, we can define the classes of $2 \times 2 \times 2$ potential games (restricted to symmetric games). In regions 1 and 6 we have *pure coordination*, where there are exactly two pure Nash equilibria, at $(+1, +1, +1)$ and $(-1, -1, -1)$. In regions 3 and 4

	$+1$	-1	$+1$	-1
$+1$	$-,-,-$	$+,+,+$	$+,+,+$	$+,-,-$
-1	$+,+,+$	$-,-,+$	$-,+,-$	$-,-,-$
	$+1$		-1	

Table 2.19: Region 4; $|\alpha + \delta| < -2\gamma$ & $\delta < \alpha$

	$+1$	-1	$+1$	-1
$+1$	$+,+,+$	$+,-,+$	$+,+,-$	$-,-,+$
-1	$-,+,+$	$-,-,+$	$-,+,-$	$-,-,-$
	$+1$		-1	

Table 2.20: Region 5; $\alpha + \delta > 2|\gamma|$ & $\delta < \alpha$

	$+1$	-1	$+1$	-1
$+1$	$+,+,+$	$+,-,+$	$+,+,-$	$-,-,-$
-1	$-,+,+$	$-,-,-$	$-,-,-$	$+,+,+$
	$+1$		-1	

Table 2.21: Region 6; $|\alpha + \delta| < 2\gamma$ & $\delta < \alpha$

	$+1$	-1	$+1$	-1
$+1$	$-,-,-$	$+,+,+$	$+,+,+$	$-,-,-$
-1	$+,+,+$	$-,-,-$	$-,-,-$	$+,+,+$
	$+1$		-1	

Table 2.22: Region 7; $\alpha + \delta < -2|\gamma|$

we have *strict anti-coordination*, where there are exactly three pure Nash equilibria, either at $(+1, +1, -1)$ and its permutations in region 4, or at $(+1, -1, -1)$ and its permutations in region 3. Regions 2 and 7 have exactly four pure Nash equilibria, reflecting the sign of δ. When $\delta > 0$, in region 2, these pure Nash equilibria are $(+1, +1, +1)$ and all permutations of $(+1, -1, -1)$, and when $\delta < 0$, in region 7, they are $(-1, -1, -1)$ and all permutations of $(+1, +1, -1)$. We call this class δ-*dominant* potential games. Lastly, in region 5 we have a unique pure strategy Nash equilibrium reflected by the sign of α. Like in the 2×2 case, we call these *independent* potential games.

2.4.4 An Application: The Legislator Game

The legislator game involves three legislators who can either vote for a pay raise, strategy A, or not vote, strategy B. In order for the vote to go through, a majority of the legislators need to vote for the raise. None of the legislators want to vote in order to save face with their voting population. At the same time, the legislators want the pay raise. This game is commonly built with the value b to denote the benefit from the raise, and the value c to denote the cost from voting, where $b > c$. The game in normal-form is displayed in Table 2.23.

	A	B
A	$b-c$ $b-c$ $b-c$	$b-c$ b $b-c$
B	b $b-c$ $b-c$	0 0 $-c$

A

	A	B
A	$b-c$ $b-c$ b	$-c$ 0 0
B	0 $-c$ 0	0 0 0

B

Table 2.23: The Legislator Game

To get a better view into the game, let us take $b = 2$ and $c = 1$, producing the normal-form game in Table 2.24.

	A			B		
A	1	1	1	1	2	1
B	2	1	1	0	0	−1

A

	A			B		
A	1	1	2	−1	0	0
B	0	−1	0	0	0	0

B

Table 2.24: Legislator Game with $b = 2$ and $c = 1$

This game has four Nash equilibria– nobody votes, and majority (but not all) vote. This makes sense because none of the legislators want to unilaterally deviate from the strategy profile where none are voting (they do not want to be the only one to vote), and if they are already in a profile where the majority is voting for the pay raise, it is in their unilateral interest to stay there. Let us use the decomposition here to see where the payoffs are coming from.

	A			B		
A	−1	−1	−1	1	1	1
B	1	1	1	−1	−1	−1

A

	A			B		
A	1	1	1	−1	−1	−1
B	−1	−1	−1	1	1	1

B

Table 2.25: Nash Component of Legislator Game

	A			B		
A	2	2	2	0	2	0
B	2	0	0	0	0	−2

A

	A			B		
A	0	0	2	−2	0	0
B	0	−2	0	−2	−2	−2

B

Table 2.26: Behavioral Component of Legislator Game

As the construction of these games now proves, there are no personal preference components,

α_i for $i = 1, 2, 3$, and no coordinative pressure components γ_{ij} between pairs of agents i and j. But let us take a closer look using the coordinate system. We see that this reflects precisely the pure Nash equilibrium structure obtained in Region 7, which is given by $\delta < -\alpha$ and $\alpha + \delta < -2|\gamma|$. Hence, the example is simply a special case where $\delta = -1$ and $\alpha = \gamma = 0$. The pure Nash structure of the legislator game shows that it is *almost* a coordination game. However, instead of $(+1, +1, +1)$ being a pure Nash equilibrium, it is split into the equilibria at $(+1, +1, -1)$ and its permutations. One interpretation of this is that it is a coordination game with a threshold. It is fine for all agents to play -1 and coordinate on $(-1, -1, -1)$, however, it is too much to have them all play $+1$; it is better that only two of them play $+1$.

The Nash structure of the legislator game only provides half of the picture. The other half comes from the behavioral component. Analyzing the behavioral component in Table 2.26 shows that the idea of thresholds is present. That is, only when two others vote does a legislator receive a positive externality. The behavioral term only kicks in to effect when two other agents choose a specific strategy. We show the behavioral structure in more detail for $2 \times 2 \times 2$ games in the end of this chapter.

We are now ready to generalize the results to n-agent potential games.

2.5 $2 \times \ldots \times 2$ Potential Structure

Here we describe how the payoff structure observed in 2×2 and $2 \times 2 \times 2$ games blossoms to that in $2 \times \ldots \times 2$ games.

In the 2×2 Nash structure, the structure revealed individual preference parameters for the two agents, α_1 and α_2, and the coordination parameters between the two agents, γ_{12} and γ_{21}. Going to $2 \times 2 \times 2$ uncovered individual preference parameters for all three agents, α_1, α_2, and α_3, coordination parameters between all ordered pairs of agents, γ_{12}, γ_{13}, γ_{21},

γ_{23}, γ_{31} and γ_{32}, and an emergent parameter for each agent that is dependent on the three agents, δ_{123}, δ_{231}, and δ_{312}. In the case of $2 \times \ldots \times 2$ games, all agents have individual preference parameters, all pairs of agents have a coordination parameter, all 3-tuples have the δ parameter, and so on, until the parameter that is dependent on all agents.

To see this, let us first define our game. Let $\mathcal{G} = \langle \mathcal{N}, S, \Pi \rangle$ be a game where $\mathcal{N} = \{1, 2, \cdots, n\}$ is the set of agents, $S = \{S_1, S_2, \cdots, S_n\}$ is the set containing all agents' strategy sets, $S_i = \{+1, -1\}$ is the strategy set for each agent i, and $\Pi = \{\pi_1, \pi_2, \cdots, \pi_n\}$ is the set of payoff functions for all agents i. In other words, $\mathcal{G}$ is an n-agent $2 \times \cdots \times 2$ game.

Define $\xi^j_{i_1 \ldots i_j}$ to be agent i_1's Nash parameter that is dependent on the strategies of the j agents $i_1, \ldots i_j$. For example, in the $2 \times 2 \times 2$ case, $\xi^1_1 = \alpha_1$, $\xi^2_{12} = \gamma_{12}$ and $\xi^3_{123} = \delta_{123}$.

In the n-agent case, there are n total individual preferences ξ^1_i. For a fixed i, there is a total of $n-1$ pairs (i, j) with first entry i and $j \neq i$. This gives $n-1$ coordination components γ_{ij} for agent i, and a total of $n(n-1)$ coordination parameters for all agents. For the 3-tuples (i, j, k), assuming the order does not matter except for the first entry, which designates the agent whose payoff structure includes the parameter, gives $\binom{n-1}{2}$ parameters for each agent i, and, for all agents, a total of $n\binom{n-1}{2}$. Continuing in this fashion there will eventually be a single, or $\binom{n-1}{n-1}$, parameter $\xi_{i,j_1,\ldots,j_{n-1}}$ for each agent that includes their interaction with all remaining agents. In total there are $n\binom{n-1}{n-1}$ such terms. Hence there are $\sum_{m=0}^{n-1} n\binom{n-1}{m} = n2^{n-1}$ Nash parameters.

Count the externality parameters as follows. Agent i will receive an externality from all other individual agents, from each pair of agents, and so on. Hence, for each agent, there are $n - 1 + \binom{n-1}{2} + \ldots + \binom{n-1}{n-1} = 2^{n-1} - 1$ externality parameters. Since this is so for every agent, the total count for the game is $n(2^{n-1} - 1)$.[7]

The kernel, a constant value for each agent, will contribute an additional n parameters.

[7]The externality structure is explained in more detail at the end of this chapter.

Summing these gives $n2^{n-1} + n + n(2^{n-1} - 1) = n2^n$, which is precisely the dimensionality of a $2 \times \ldots \times 2$ game.

Without loss of generality, and because it simplifies the indices of the summations, we show only agent 1's Nash, behavioral, and kernel components of the utility function.

$$\pi_1^N(t_1, \ldots, t_n) = \xi_1^1 t_1 + \sum_{\substack{\text{unordered} \\ \text{pairs } (1,j)}} \xi_{1j}^2 t_1 t_j + \sum_{\substack{\text{unordered} \\ \text{3-tuples } (1,j,k)}} \xi_{1jk}^3 t_1 t_j t_k + \ldots$$
$$\ldots + \sum_{\substack{\text{unordered} \\ \text{(n-1)-tuples} \\ (1,\ldots,i_{n-1})}} \xi_{1\ldots i_{n-1}}^{n-1} t_1 \ldots t_{i_{n-1}} + \xi_{12\ldots n}^n t_1 \ldots t_n$$

$$\pi_1^B(t_1, \ldots, t_n) = \sum_{i \neq 1}^{n} \zeta_{1i}^1 t_i + \sum_{\substack{\text{unordered} \\ \text{pairs } (i,j) \\ i,j \neq 1}} \zeta_{1ij}^2 t_i t_j + \sum_{\substack{\text{unordered} \\ \text{3-tuples } (i,j,k) \\ i,j,k \neq 1}} \zeta_{1ijk}^3 t_i t_j t_k + \ldots$$
$$\ldots + \sum_{\substack{\text{unordered} \\ \text{(n-2)-tuples} \\ (i_1,\ldots,i_{n-1}) \\ \text{none of which}=1}} \zeta_{i_1 \ldots i_{n-2}}^{n-2} t_{i_1} \ldots t_{i_{n-1}} + \zeta_{12\ldots n-1}^{n-1} t_2 \ldots t_n$$

$$\pi_1^K(t_1, \ldots, t_n) = \kappa_1$$

Where here the parameters $\xi_{i_1 \ldots i_j}^j$ denote the Nash parameters dependent on the strategies of the j agents $i_1, \ldots i_j$, as described earlier in the section. Similarly $\zeta_{i_1 \ldots i_j}^j$ denotes the externality parameter with same interpretation of indices as $\xi_{i_1 \ldots i_j}^j$.

2.5.1 $2 \times \ldots \times 2$ Potential Games

Following what was observed in the 2×2 and $2 \times 2 \times 2$ case, we collapse all the shared Nash components. In the $2 \times 2 \times 2$ game this meant taking $\xi_{ij}^2 = \xi_{ji}^2$, or $\gamma_{ij} = \gamma_{ji}$, for all unordered pairs of agents (i, j), and $\xi^3_{ijk} = \xi^3_{jki} = \xi^3_{kij}$, or $\delta_{ijk} = \delta_{jki} = \delta_{kij}$, for the three agents i, j, and k.

Since the ξ^1_i have been left alone for each $i \in \mathcal{N}$, there are n of these parameters. Then, because we are collapsing ξ^2_{ij} and ξ^2_{ij} this is equivalent to saying that the order of the subscripts now does not matter. Hence, calculating this value before gave $n\binom{n-1}{1}$, and now it is $\binom{n}{2}$. In this manner, all Nash parameters in the $2 \times \ldots \times 2$ potential game can be calculated to be $\sum_{m=1}^{n} \binom{n}{m} = 2^n - 1$, where before it was $n \sum_{m=0}^{n-1} \binom{n-1}{m} = n2^{n-1}$ Nash parameters. Consequently, the subspace of the Nash space that is orthogonal to potential games has dimensionality $n2^{n-1} - (2^n - 1) = 2^{n-1}(n-2) + 1$. This is the generalized matching pennies subspace.

In a similar fashion as for the 2×2 and $2 \times 2 \times 2$ potential games, the potential function can be written in the form,

$$P(t_1, \ldots, t_n) = \sum_{i=1}^{n} \xi^1_i t_i + \sum_{\substack{\text{unordered} \\ \text{pairs } (i,j)}} \xi^2_{ij} t_i t_j + \sum_{\substack{\text{unordered} \\ 3\text{-tuples } (i,j,k)}} \xi^3_{ijk} t_i t_j t_k + \ldots$$
$$\ldots + \sum_{\substack{\text{unordered} \\ (\text{n}-1)\text{-tuples} \\ (i_1,\ldots,i_{n-1})}} \xi^{n-1}_{i_1 \ldots i_{n-1}} t_{i_1} \ldots t_{i_{n-1}} + \xi^n_{12\ldots n} t_1 \ldots t_n \quad (2.9)$$

The structure given by each m-tuple parameter is easier to analyze than it first may seem. This is because they are given by the product of the strategies t_i for every i in the m-tuple, and each t_i can be either $+1$ or -1. When this is a pair, we only have $t_i t_j$ for i, j in the pair. This will be positive when exactly both t_i and t_j are positive or both are negative, and it will be negative otherwise. This gives the coordination structure between the pair of agents. Now, for any m-tuple we have the product $t_{i_1} \ldots t_{i_m}$. For even m this will be positive exactly when there are an even number of agents playing $+1$ and an even number of agents playing -1. For odd m this will be positive exactly when there are an even number (or zero) of agents playing -1. Each structure, in absence of the other ones, has pure strategy Nash equilibria in the cases where all agents involved in the structure play $+1$ or an even number play -1. Multiplying these structures by -1 switches this around.

For example, take $n = 5$. Then for the each agent i the ξ_i^1 parameter will contribute its value to all agent i's payoffs where agent i plays $+1$, and will contribute the negative of its value to all profiles where agent i plays -1. This is clear since ξ_i^1 is affected by only t_i. For each pair of agents (i, j), the ξ_{ij}^2 parameter will contribute its value to the payoff of i precisely when i and j both play $+1$ or both play -1, and its negative when i and j play different strategies. For each 3-tuple of agents, ξ_{ijk}^3 will contribute its value to the payoffs of agent i in all cases where i, j, and k, play $(+1, +1, +1)$ and all permutations of $(+1, -1, -1)$, and its negative value in all other strategy combinations. For each 4-tuple of agents, the ξ_{ijkl}^4 parameter will contribute to agent i's payoff when the four agents play $(+1, +1, +1, +1)$ and all permutations of $(+1, +1, -1, -1)$, and will contribute its negative in all other strategy combinations of the four agents. Finally, for the 5-tuple consisting of all agents, the ξ_{ijklm}^5 parameter will contribute to agent i's payoff when the five agents play $(+1, +1, +1, +1, +1)$ and all permutations of $(+1, +1, +1, -1, -1)$ and $(+1, -1, -1, -1, -1)$, and will contribute its negative value in all other strategy choices.

An important feature of the components and the payoffs they contribute to the overall payoff structure of the game is captured in the below theorem.

Theorem 2.15. *A ξ^m component cannot be created from a linear combination of the ξ^k components for $k < m$.*

Proof. This is immediate from the orthogonality of the terms. □

An important implication of this theorem is more easily seen considering the $2 \times 2 \times 2$ case. Here the shared δ component cannot be built from a linear combination of $\alpha_1, \alpha_2, \alpha_3, \gamma_{12}, \gamma_{13}$, and γ_{23}. In other words, modeling a $2 \times 2 \times 2$ game as an aggregate of smaller 2×2 games cannot capture the complete structure of $2 \times 2 \times 2$ games. This idea is behind the following corollary to Theorem 2.15.

Corollary 2.15.1. *In terms of Nash structure, an n-agent $2 \times \ldots \times 2$ potential game can be represented as the sum of all smaller m-agent interactions if and only if $\xi^k = 0$ for all $k > m$.*

Proof. Let $\mathcal{G}$ be an n-agent $2 \times \ldots \times 2$ potential game that can be represented as the sum of all smaller m-agent interactions. Then $\xi^k = 0$ for all $k > m$ because otherwise the sum of all smaller m-agent interactions would not be able to produce the effect of ξ^k on the payoff structure of $\mathcal{G}$. For the other direction, let $\mathcal{G}$ be an n-agent $2 \times \ldots \times 2$ potential game with $\xi^k = 0$ for all $k > m$. Then, because there are no terms dependent on more than m agents, $\mathcal{G}$ can be represented as a sum of all smaller m-agent interactions. □

The above corollary concerns modeling and reducibility. Often times, research questions in game theory involve a large number of agents, but because of the complexity of in modeling these games, simplifying assumptions about the nature of the micro-interactions are made. For instance in [18], [19], and [12], an n-agent game on a network is modeled by defining each agent's payoff as the sum of their 2×2 interactions with their neighbors. Corollary 2.15.1 tells us that this is only an accurate representation if the situation being modeled has no important higher order interactions of 3 or more agents.

2.5.2 Generalized Matching Pennies

In this section, sketches for constructing a basis is offered for the subspace of the Nash space that is orthogonal to the potential space, the n-agent generalized matching pennies space of games. In the previous section the dimensionality of this space was calculated to be $2^{n-1}(n-2)+1$.

A $2\times\ldots\times 2$ game can be thought of as an n-dimensional cube where each vertex corresponds to a strategy profile. Each agent can move along the edges of the cube in their dimension by switching strategies. In the 2-agent case this is simply a square and agent 1 can move along the x-axis, and agent 2 along the y-axis, where each vertex is one of the strategy profiles $(+1,+1),(+1,-1),(-1,+1)$, and $(-1,-1)$. Here there is only one matching pennies component, the matching pennies game on the unique face of the square.

In the $2\times 2\times 2$ case, the generalized matching pennies component is 5-dimensional, and the basis for this was defined using a matching pennies game in 5 of the faces of the cube (a matching pennies game for the remaining face can be constructed using a linear combination of the other components).

For $2\times\ldots\times 2$ games this will follow a similar pattern. For agent 1, define a matching pennies game with all other agents, on every relevant face. For example, with agent 1 and 2 there will be a matching pennies game when the strategies of the remaining agents are fixed across all 2^{n-2} possibilities. Doing this for every pair of agents including agent 1 gives a total of $(n-1)2^{n-2}$ matching pennies vectors. With agent 2, follow a similar process. For agents $k=3$ up to $n-1$, construct a matching pennies vector on all faces of the cube where agent 2 and k are playing with fixed strategies for the remaining agents. For agent n we only include one such vector. In total this contributes $(n-3)2^{n-2}+1$. Then, $(n-1)2^{n-2}+(n-3)2^{n-2}+1=2^{n-1}(n-2)+1$, which is the dimensionality of the subspace.

The results regarding the class of games orthogonal to potential games can be extended. Firstly, any $2\times\ldots\times 2$ game can be written as the sum of a potential game plus a generalized matching pennies game. Consequently, any game whose dot product with the vectors of this class is not zero is not an exact potential game because its payoff structure is built from vectors from this space. With this, instead of classifying such games using the cyclic conditions often seen in the literature, we can interpret them as containing any one of the plethora of conditions that are constructible using the basis described.

Now that the structure of 2-strategy potential games for any number of agents has been exposed, let us open the first door to extending these results to any number of strategies. In this thesis we develop the structure only for 3×3 potential games.

2.6 3×3 Potential Structure

In the 2×2 case things are nice because the agents can pick between strategies $+1$ and -1, which allowed for a very natural construction of utility functions. In the three agent case, although it is not as simple, results can be derived.

The Nash, Behavioral, and Kernel structures of a 3×3 game are given in Tables 2.27, 2.28, and 2.29, respectively.

	A	B	C
A	η^1_{11} η^2_{11}	η^1_{12} η^2_{21}	η^1_{13} η^2_{31}
B	η^1_{21} η^2_{12}	η^1_{22} η^2_{22}	η^1_{23} η^2_{32}
C	η^1_{31} η^2_{13}	η^1_{32} η^2_{23}	η^1_{33} η^2_{33}

Table 2.27: 3×3 Nash Structure

Where for the Nash component we must have $\eta^i_{k3} = -\eta^i_{k1} - \eta^i_{k2}$ for each agent $i = 1, 2$ and for $k = 1, 2, 3$ [8]. For the behavioral component we must have $\beta_{i3} = -\beta_{i1} - \beta_{i2}$ for each agent $i = 1, 2$ [8].

	A		B		C	
A	β_{11}	β_{21}	β_{12}	β_{21}	β_{13}	β_{21}
B	β_{11}	β_{22}	β_{12}	β_{22}	β_{13}	β_{22}
C	β_{11}	β_{23}	β_{12}	β_{23}	β_{13}	β_{23}

Table 2.28: 3×3 Behavioral Structure

	A		B		C	
A	κ_1	κ_2	κ_1	κ_2	κ_1	κ_2
B	κ_1	κ_2	κ_1	κ_2	κ_1	κ_2
C	κ_1	κ_2	κ_1	κ_2	κ_1	κ_2

Table 2.29: 3×3 Kernel

2.6.1 A Change of Basis

The Nash Component

We reference Table 2.27. Because $\eta^i_{k3} = -\eta^i_{k1} - \eta^i_{k2}$ for each agent $i = 1, 2$ and for $k = 1, 2, 3$, there are six independent parameters per agent in the Nash component of a 3×3 game. Thus, in total, the Nash component of a 3×3 game is 12-dimensional. We can represent these 12 independent parameters with a canonical 12-dimensional basis, where the first 6 dimensions are reserved for agent 1, and the last 6 for agent 2. These are shown below, first for agent 1.

$$e_{11} = (1,0,0,0,0,0,0,0,0,0,0,0), \;\; e_{12} = (0,1,0,0,0,0,0,0,0,0,0,0)$$

$$e_{13} = (0,0,1,0,0,0,0,0,0,0,0,0), \;\; e_{14} = (0,0,0,1,0,0,0,0,0,0,0,0)$$

$$e_{15} = (0,0,0,0,1,0,0,0,0,0,0,0), \;\; e_{16} = (0,0,0,0,0,1,0,0,0,0,0,0)$$

For agent 2,

$$e_{21} = (0,0,0,0,0,0,1,0,0,0,0,0), \quad e_{22} = (0,0,0,0,0,0,0,1,0,0,0,0)$$
$$e_{23} = (0,0,0,0,0,0,0,0,1,0,0,0), \quad e_{24} = (0,0,0,0,0,0,0,0,0,1,0,0)$$
$$e_{25} = (0,0,0,0,0,0,0,0,0,0,1,0), \quad e_{26} = (0,0,0,0,0,0,0,0,0,0,0,1)$$

The Nash component in Table 2.27 can then be written as $\eta^1_{11}e_{11} + \eta^1_{12}e_{12} + \eta^1_{13}e_{13} + \eta^1_{21}e_{14} + \eta^1_{22}e_{15} + \eta^1_{23}e_{16}$ for agent 1, and $\eta^2_{11}e_{21} + \eta^2_{12}e_{22} + \eta^2_{13}e_{23} + \eta^2_{21}e_{24} + \eta^2_{22}e_{25} + \eta^2_{23}e_{26}$ for agent 2.

Following a similar procedure as in the 2×2 case, we look for vectors representing individual preferences, and vectors representing social pressures (like the pressure to conform in the 2×2 case). For simplicity and without loss of generality we only consider agent 1's vectors and omit the last 6 zeros. We propose the basis below,

$$e^*_{11} = (1,1,1,0,0,0)$$
$$e^*_{12} = (0,0,0,1,1,1)$$
$$e^*_{13} = (2,0,0,-1,0,0)$$
$$e^*_{14} = (0,-1,0,0,2,0)$$
$$e^*_{15} = (0,0,-1,0,0,-1)$$
$$e^*_{16} = (0,-1,1,1,0,-1)$$

The vectors e^*_{11} and e^*_{12} span the space of all rankings over strategies A, B, and C, where the sum is equal to zero, and where this ranking is invariant over the other agent's strategy. Hence we immediately associate these vectors with the individual preference components we have discussed at length in the previous sections.

This invariance is not present in vectors $e^*_{13}, e^*_{14}, e^*_{15}$, and e^*_{16}, but together, these vectors have the essence of coordination. The first three, e^*_{13}, e^*_{14}, and e^*_{15}, are in the perspective of pure coordination. If these are positive, then there are pressures to coordinate on (A, A), (B, B), and (C, C). If these are negative, then there are pressures to anti-coordinate. Without the remaining vector we are not able to distribute the anti-coordinative weight. For example, it might be better to play C when the other agent plays A, than to play B.

We represent this basis as matrices to make the definition of the agents' utility functions simpler. We have $e^*_i = A^i$ for $i = 1, 2, 3, 4, 5, 6$. Here the superscript of the matrix is not an exponent but simply indexes the different basis matrices. The subscript will be saved to designate specific row-column entries. More precisely, A^k_{ij} means the entry in the i^{th} row and j^{th} column of A^k. If a matrix is invariant over its columns, then A^k_i represents the entry in the i^{th} row of A^k.

$$A^1 = \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -1 & -1 \\ \hline \end{array} \quad A^2 = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline -1 & -1 & -1 \\ \hline \end{array} \quad A^3 = \begin{array}{|c|c|c|}\hline 2 & 0 & 0 \\ \hline -1 & 0 & 0 \\ \hline -1 & 0 & 0 \\ \hline \end{array}$$

$$A^4 = \begin{array}{|c|c|c|}\hline 0 & -1 & 0 \\ \hline 0 & 2 & 0 \\ \hline 0 & -1 & 0 \\ \hline \end{array} \quad A^5 = \begin{array}{|c|c|c|}\hline 0 & 0 & -1 \\ \hline 0 & 0 & -1 \\ \hline 0 & 0 & 2 \\ \hline \end{array} \quad A^6 = \begin{array}{|c|c|c|}\hline 0 & -1 & 1 \\ \hline 1 & 0 & -1 \\ \hline -1 & 1 & 0 \\ \hline \end{array}$$

Now we let us rename the strategies A, B, and C, to the numbers 1, 2, and 3, respectively.

Hence we can write the Nash component of agent 1's utility function as, where $t_1, t_2 = 1, 2, 3$,

$$\pi^N_1(t_1, t_2) = \alpha_{11}A^1_{t_1} + \alpha_{12}A^2_{t_1} + \gamma_{11}A^3_{t_1t_2} + \gamma_{12}A^4_{t_1t_2} + \gamma_{13}A^5_{t_1t_2} + \gamma_{14}A^6_{t_1t_2} \tag{2.10}$$

This is nice because agent 2's utility function has the same underlying matrices, but with t_1

and t_2 swapped. This gives,

$$\pi_2^N(t_1,t_2) = \alpha_{21}A^1_{t_2} + \alpha_{22}A^2_{t_2} + \gamma_{21}A^3_{t_2t_1} + \gamma_{22}A^4_{t_2t_1} + \gamma_{23}A^5_{t_2t_1} + \gamma_{24}A^6_{t_2t_1} \tag{2.11}$$

The Behavioral and Kernel Components

Consider the behavioral component given in Table 2.28. One notices that the behavioral values have the same underlying structure as the individual preferences, except with the agents switched. For example, an agent can receive one of three individual preference values from choosing a strategy, and at the same time, impose one of three externality values on the other agent, both with dimensionality 2. Mathematically, this means that we can use the same underlying matrices A^1 and A^2 for both the individual preferences and the externalities. This gives

$$\pi_1^B(t_1,t_2) = \beta_{11}A^1_{t_2} + \beta_{12}A^2_{t_2} \tag{2.12}$$

$$\pi_2^B(t_1,t_2) = \beta_{21}A^1_{t_1} + \beta_{22}A^2_{t_1} \tag{2.13}$$

The kernel component is trivial as it is invariant over any strategy profile. In other words,

$$\pi_1^K(t_1,t_2) = \kappa_1 \tag{2.14}$$

$$\pi_2^K(t_1,t_2) = \kappa_2 \tag{2.15}$$

We now state a theorem.

Theorem 2.16. *A* 3×3 *game* $\mathcal{G}$ *can be represented by the utility functions for agents 1 and*

2 shown below,

$$\pi_1(t_1,t_2) = \alpha_{11}A^1_{t_1} + \alpha_{12}A^2_{t_1} + \gamma_{11}A^3_{t_1t_2} + \gamma_{12}A^4_{t_1t_2} + \gamma_{13}A^5_{t_1t_2} + \gamma_{14}A^6_{t_1t_2} + \beta_{11}A^1_{t_2} + \beta_{12}A^2_{t_2} + \kappa_1 \quad (2.16)$$

$$\pi_2(t_1,t_2) = \alpha_{21}A^1_{t_2} + \alpha_{22}A^2_{t_2} + \gamma_{21}A^3_{t_2t_1} + \gamma_{22}A^4_{t_2t_1} + \gamma_{23}A^5_{t_2t_1} + \gamma_{24}A^6_{t_2t_1} + \beta_{21}A^1_{t_1} + \beta_{22}A^2_{t_1} + \kappa_2 \quad (2.17)$$

Proof. This follows from the nine simple, but tedious, computations over all combinations of values $t_1, t_2 = 1, 2, 3$. □

2.6.2 Potential Game Requirements

For $\mathcal{G}$ to be a potential game we need the existence of a potential function that reflects all possible unilateral deviations. Begin by defining the potential function $P(t_1,t_2) = \pi_1(t_1,t_2) + \pi_2(t_1,t_2)$.

This means for player 1 we need

$$\pi_1(1,1) - \pi_1(2,1) = P(1,1) - P(2,1)$$
$$\pi_1(1,1) - \pi_1(3,1) = P(1,1) - P(3,1)$$
$$\pi_1(1,2) - \pi_1(2,2) = P(1,2) - P(2,2)$$
$$\pi_1(1,2) - \pi_1(3,2) = P(1,2) - P(3,2)$$
$$\pi_1(1,3) - \pi_1(2,3) = P(1,3) - P(2,3)$$
$$\pi_1(1,3) - \pi_1(3,3) = P(1,3) - P(3,3)$$

We also need this for player 2.

$$\pi_2(1,1) - \pi_2(1,2) = P(1,1) - P(1,2)$$
$$\pi_2(1,1) - \pi_2(1,3) = P(1,1) - P(1,3)$$
$$\pi_2(2,1) - \pi_2(2,2) = P(2,1) - P(2,2)$$
$$\pi_2(2,1) - \pi_2(2,3) = P(2,1) - P(2,3)$$
$$\pi_2(3,1) - \pi_2(3,2) = P(3,1) - P(3,2)$$
$$\pi_2(3,1) - \pi_2(3,3) = P(3,1) - P(3,3)$$

After going through these calculations, we see that in order for $\mathcal{G}$ to be a 3×3 potential game with the basis offered in this section, we need $\gamma_{1i} = \gamma_{2i}$ for $i = 1, 2, 3, 4$, which we relabel as γ_i. In addition, like in the 2-strategy case, the potential function must exclude the behavioral terms. We summarize these results in the following two theorems and accompanying corollary.

Theorem 2.17. *A 3×3 game $\mathcal{G}$ with utility functions given in (2.16) and (2.17) is a potential game if and only if $\gamma_{1i} = \gamma_{2i}$ for $i = 1, 2, 3, 4$.*

Proof. This follows from taking $P(t_1, t_2) = \pi_1(t_1, t_2) + \pi_2(t_1, t_2)$ and checking all possible unilateral deviations in P with the respective agent's payoff function. From here, it is immediate that $\gamma_{1i} = \gamma_{2i}$ for $i = 1, 2, 3, 4$ is needed. The entire process, though tedious, is straightforward. □

Theorem 2.18. *If $\mathcal{G}$ is a 3×3 potential game, then any potential function for $\mathcal{G}$ can be written as*

$$P(t_1, t_2) = \alpha_{11} A^1_{t_1} + \alpha_{12} A^2_{t_1} + \alpha_{21} A^1_{t_2} + \alpha_{22} A^2_{t_2} + \gamma_1 A^3_{t_1 t_2} + \gamma_2 A^4_{t_1 t_2} + \gamma_3 A^5_{t_1 t_2} + \gamma_4 A^6_{t_1 t_2} + c \quad (2.18)$$

where $c \in \mathbb{R}$ is an arbitrary constant.

Proof. Similar to the proof of Theorem 2.17, this amounts to checking that the differences in all unilateral deviations agree between P and the respective π_i. $\square$

Corollary 2.18.1. *The potential function of a* 3×3 *potential game* $\mathcal{G}$ *depends only on the Nash component of* $\mathcal{G}$*.*

Proof. This follows immediately from the form of $P(t_1, t_2)$ given in Theorem 2.18. $\square$

2.6.3 The 3×3 Orthogonal Subspace

We re-emphasize that the requirement on the Nash parameters that $\mathcal{G}$ be a potential game are that $\gamma_{1i} = \gamma_{2i}$ for $i = 1, 2, 3, 4$. This means combining the four vectors whose coefficients are γ_{1i} and γ_{2i} into single vectors with coefficients γ_i for $i = 1, 2, 3, 4$. To find the orthogonal vectors then, just as we did in the 2×2 case, we can define four vectors where the coefficient $\gamma_{1i} = -\gamma_{2i}$. It is a straightforward computation to verify that these are indeed orthogonal.

What this amounts to is defining the vectors $n_1 = e^*_{13} + e^*_{23}, n_2 = e^*_{14} + e^*_{24}, n_3 = e^*_{15} + e^*_{25}$ and $n_4 = e^*_{16} + e^*_{26}$. Similarly we define $\widetilde{n_1} = e^*_{13} - e^*_{23}, \widetilde{n_2} = e^*_{14} - e^*_{24}, \widetilde{n_3} = e^*_{15} - e^*_{25}$ and $\widetilde{n_4} = e^*_{16} - e^*_{26}$.

$$e^*_{11} = (1,1,1,0,0,0,0,0,0,0,0,0), \quad e^*_{12} = (0,0,0,1,1,1,0,0,0,0,0,0)$$

$$e^*_{13} = (2,0,0,-1,0,0,0,0,0,0,0,0), \quad e^*_{14} = (0,-1,0,0,2,0,0,0,0,0,0,0)$$

$$e^*_{15} = (0,0,-1,0,0,-1,0,0,0,0,0,0), \quad e^*_{16} = (0,-1,1,1,0,-1,0,0,0,0,0,0)$$

$$e^*_{21} = (0,0,0,0,0,0,1,1,1,0,0,0), \quad e^*_{22} = (0,0,0,0,0,0,0,0,0,1,1,1)$$

$$e^*_{23} = (0,0,0,0,0,0,2,0,0,-1,0,0), \quad e^*_{24} = (0,0,0,0,0,0,0,-1,0,0,2,0)$$

$$e^*_{25} = (0,0,0,0,0,0,0,0,-1,0,0,-1), \quad e^*_{26} = (0,0,0,0,0,0,0,-1,1,1,0,-1)$$

For clarity, we represent these vectors in normal-form below.

$n_1 =$

$2,\ 2$	$0, -1$	$0, -1$
$-1,\ 0$	$0,\ 0$	$0,\ 0$
$-1,\ 0$	$0,\ 0$	$0,\ 0$

$n_2 =$

$0,\ 0$	$-1,\ 0$	$0,\ 0$
$0, -1$	$2,\ 2$	$0, -1$
$0,\ 0$	$-1,\ 0$	$0,\ 0$

$n_3 =$

$0,\ 0$	$0,\ 0$	$-1,\ 0$
$0,\ 0$	$0,\ 0$	$-1,\ 0$
$0, -1$	$0, -1$	$2,\ 2$

$n_4 =$

$0,\ 0$	$-1, -1$	$1,\ 1$
$1,\ 1$	$0,\ 0$	$-1, -1$
$-1, -1$	$1,\ 1$	$0,\ 0$

$\widetilde{n}_1 =$

$2, -2$	$0,\ 1$	$0,\ 1$
$-1,\ 0$	$0,\ 0$	$0,\ 0$
$-1,\ 0$	$0,\ 0$	$0,\ 0$

$\widetilde{n}_2 =$

$0,\ 0$	$-1,\ 0$	$0,\ 0$
$0,\ 1$	$2, -2$	$0,\ 1$
$0,\ 0$	$-1,\ 0$	$0,\ 0$

$\widetilde{n}_3 =$

$0,\ 0$	$0,\ 0$	$-1,\ 0$
$0,\ 0$	$0,\ 0$	$-1,\ 0$
$0,\ 1$	$0,\ 1$	$2, -2$

$\widetilde{n}_4 =$

$0,\ 0$	$-1,\ 1$	$1, -1$
$1, -1$	$0,\ 0$	$-1,\ 1$
$-1,\ 1$	$1, -1$	$0,\ 0$

It is clear that the vector $\widetilde{n}_4 =$ is a rock-paper-scissors game. Extending what was done in the 2×2 case, we see that the rock-paper-scissors game is a one-dimensional subclass in the generalized 3×3 matching pennies games, which comprises the subspace of games orthogonal to potential games.

2.7 Externalities

The purpose of this section is to establish a basis for the behavioral component of games. Since potential games place no requirements on the behavioral component, the work done in this section applies to games in general. We use the terms behavioral component, externality component, and externality structure interchangeably.

2.7.1 2×2 Externality Structure

The $2{\times}2$ behavioral component is given in normal-form in Table 2.30. We offer no alternative basis for the 2×2 behavioral component since the intuition already offered is one we would like to adopt. This intuition is that the choice of each agent i has, as a consequence, $\beta_{\neg i}$ when they play $+1$ and $-\beta_{\neg i}$ when they play -1, where $\neg i$ represents the agent who is not agent i.

	$+1$		-1	
$+1$	β_1	β_2	$-\beta_1$	β_2
-1	β_1	$-\beta_2$	$-\beta_1$	$-\beta_2$

Table 2.30: 2×2 Externality Structure

2.7.2 $2 \times 2 \times 2$ Externality Structure

The $2 \times 2 \times 2$ behavioral component is given in normal-form in Table 2.31, where $\beta_{i4} = -\beta_{i1} - \beta_{i2} - \beta_{i3}$. It is immediate that $\dim \mathbb{G}^B = 9$.

In the section where we developed the potential structure of $2 \times 2 \times 2$ games, we discussed only the Nash component of the agents' payoff functions. We now extend these functions to include the externality terms. We propose the following behavioral component of the payoff functions, which can be verified to form an orthogonal spanning set of the behavioral component.

$$\pi_1^B(t_1, t_2, t_3) = \beta_{12}t_2 + \beta_{13}t_3 + \beta_{123}t_2t_3$$

$$\pi_2^B(t_1, t_2, t_3) = \beta_{21}t_1 + \beta_{23}t_3 + \beta_{231}t_1t_3$$

$$\pi_3^B(t_1, t_2, t_3) = \beta_{31}t_1 + \beta_{32}t_2 + \beta_{312}t_1t_2$$

	$+1$			-1		
$+1$	β_{11}	β_{21}	β_{31}	β_{12}	β_{21}	β_{33}
-1	β_{11}	β_{22}	β_{32}	β_{12}	β_{22}	β_{34}

$+1$

	$+1$			-1		
$+1$	β_{13}	β_{23}	β_{31}	β_{14}	β_{23}	β_{33}
-1	β_{13}	β_{24}	β_{32}	β_{14}	β_{24}	β_{34}

-1

Table 2.31: $2 \times 2 \times 2$ Externality Structure

For each agent the externality structure is 3-dimensional. We propose a more intuitive basis. Right now, for each agent, the basis is $(1, 0, 0, -1), (0, 1, 0, -1), (0, 0, 1, -1)$. We propose the basis $(1, -1, 1, -1), (1, 1, -1, -1), (1, -1, -1, 1)$. It is easy to verify that these vectors are orthogonal to each other and to the Nash and kernel terms.

In agent 1's perspective, the first vector isolates the externalities brought about by agent 2, the second vector are the externalities brought about by agent 3, and the last vector gives the externalities dependent on both agent 2 and agent 3. More specifically, this externality reflects the effects of agents 2 and 3 coordinating. For example, if both play $+1$ or both play -1, the result is the same. If both play different strategies, however, the result is the negative of when both agents coordinate.

In other words, this basis offers a way of understanding the externalities as the sum of the individual and combined actions of the other agents.

2.7.3 $2 \times \ldots \times 2$ Externality Structure

The externality structure extends to a higher number of agents in precisely the same way the Nash structure extended. Each agent will receive $n-1$ individual externalities from the other agents. Then every pair of agents has a coordination externality, this gives $\binom{n-1}{2}$ externalities for each agent. This continues until we have the externality coming from the combined strategies of the remaining $n-1$ agents.

Without loss of generality, we write the behavioral component of the payoff function for agent 1 to simplify the indices of the summations,

$$\pi_1^B(t_1, \ldots, t_n) = \sum_{i=2}^{n} \beta_{1i} t_i + \sum_{\substack{\text{ordered} \\ \text{pairs } (i,j) \\ i,j \neq 1, i<j}} \beta_{1ij} t_i t_j + \ldots + \beta_{12\ldots n} t_2 \ldots t_n \tag{2.19}$$

We make note that this, in many ways, mirrors the Nash structure. The Nash structure of the game, for each agent i, involves agent i's individual decision and all higher-order decisions with additional agents. The behavioral structure of the game, for each agent i, involves the individual decision of all other agents, in addition to all high-order decisions between these remaining agents.

2.7.4 3×3 Externality Structure

This was discussed in the 3×3 section, like the 2×2 case, we explore no additional bases.

2.8 Identical Play

The class of identical play games is a curious class of games where in each strategy profile, all agents receive the same payoffs. With the payoff functions developed in this chapter, a simple way to think about this is the fact that these payoff functions have to be equal. In the 2×2 case this is simply $\pi_1(t_1, t_2) = \pi_2(t_1, t_2)$. The only way for this to be the case is for the coefficients in front of each $t_1, t_2, t_1t_2, 1$ to be the same. The results of this are stated in the below theorem.

Theorem 2.19. *A 2×2 game is an identical play game if and only if $\alpha_1 = \beta_2$, $\alpha_2 = \beta_1$, $\gamma_{12} = \gamma_{21}$, and $\kappa_1 = \kappa_2$.*

Proof. This is immediate from setting $\pi_1(t_1, t_2) = \pi_2(t_1, t_2)$. □

The implication of Theorem 2.19 is that what an agent receives individually in terms of their α must be equal to the externality they produce. For example, if agent 1 receives α_1 from playing strategy $+1$, then agent 2 will receive the externality α_1 from this decision, too. In other words, what an agent receives must be equal to what they *give*. We now state an immediate corollary relating potential games and identical play games.

Corollary 2.19.1. *The space of 2×2 identical play games is a subspace of the space of potential games.*

Proof. From Theorem 2.19 it is immediate that identical play games are potential games because $\gamma_{12} = \gamma_{21}$. Furthermore, identical play games place restrictions on the externality terms, which potential games do not. Because of this, it follows that potential games include identical play games. □

In other words, a potential game (with equal kernel values for all agents) can be transformed into an identical play game from an appropriate choice of the externality terms. Let us see how this extends to 3 agents. In this case we will need $\pi_1(t_1, t_2, t_3) = \pi_2(t_1, t_2, t_3) = \pi_3(t_1, t_2, t_3)$. We state the results of this in the following theorem.

Theorem 2.20. *A* $2 \times 2 \times 2$ *game is an identical play game if and only if* $\alpha_1 = \beta_{21} = \beta_{31}$, $\alpha_2 = \beta_{12} = \beta_{32}$, $\alpha_3 = \beta_{13} = \beta_{23}$, $\gamma_{12} = \gamma_{21} = \beta_{312}$, $\gamma_{13} = \gamma_{31} = \beta_{231}$, $\gamma_{23} = \gamma_{32} = \beta_{123}$ *and* $\kappa_1 = \kappa_2 = \kappa_3$.

Proof. This is immediate after setting $\pi_1(t_1, t_2, t_3) = \pi_2(t_1, t_2, t_3) = \pi_3(t_1, t_2, t_3)$. □

Our interpretation of the 2×2 case immediately carries over, but here we must also include the coordination components. Now, in addition to the requirement that the externality imposed on others having to equal the amount gained individually, the amount gained from coordinating with another agent must produce the same externality to the agent originally excluded from the coordination. In agent 1's point of view this is captured succinctly by the equality $\gamma_{23} = \gamma_{32} = \beta_{123}$. Here, what agent 2 receives from coordinating with agent 3, γ_{23}, must equal what agent 3 receives from coordinating with agent 2, γ_{32}, which must also equal what agent 1 receives from the coordination of agents 2 and 3, β_{123}.

We now state the extension of Corollary 2.19.1 to the case of 3 agents.

Corollary 2.20.1. *The space of* $2 \times 2 \times 2$ *identical play games is a subspace of the space of potential games.*

Proof. The proof amounts to the same verifications involved in the proof of Corollary 2.19.1.

□

The pattern naturally extends to the case of n agents. Here we must have that for each agent i their individual preference α_i equals the externality produced from this decision to all agents, that is $\alpha_i = \beta_{ji}$ for all i and $j \neq i$. For the higher terms, the pattern is the same. For example each $\gamma_{ij} = \gamma_{ji}$ but is also equal to β_{kij} for all $i \neq j, i \neq k, j \neq k$. The implications of this are fascinating, as the consequences of each action– individual decisions, coordination decisions, and higher-order decisions– must be the same for all agents.

2.9 Common Interest

We have mentioned that potential games are often referred to as common interest games. We have shown that potential games depend only on the Nash structure of the game. Hence, it should follow that "common interest" depends only on the Nash structure of the game.

When taking the payoff functions for a general game and imposing the requirements of a potential game, we collapsed the parameters of all agents involved in a particular component. In the 2×2×2 case, for example, we had $\gamma_{12} = \gamma_{21}$, $\gamma_{13} = \gamma_{31}$, $\gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$. Because of this we associate the idea of "common interest" to these parameters becoming "common" parameters.

However, is this good enough?

In the 2×2 case, a potential game with all components equal to zero except α_1 and α_2 is still a potential game. Here it is arguable that although the game is a potential game, there is no common interest. Even if the two agents have the same individual preference, we cannot say this preference is common interest in the sense where obtaining this result is dependent on both of their actions. If the only parameters of the game are α_1 and α_2, agent 1 and 2 can receive the payoff for playing their preference independent of what the other agent does. For this reason, we do not consider α_1 and α_2 as belonging to the common interest structure of the game. If the individual preferences happen to be aligned, a more appropriate terminology may be *coincidental interest*.

This suggests that the shared parameters of the game need to be positive in order for there to be common interest in the Nash component. In addition, it is worth considering the payoff structure orthogonal to the Nash component, the externality structure.

As we pointed out, as it stands, the definition of common interest must rely only on the Nash component of the game. Nevertheless, the externality structure plays an important role, and we saw that when the Nash component and the externality structure align in a special way, this gives rise to identical play games. Should the externalities be included in the notion of common interest? A quick answer is no, which agrees with the current literature because the notion of potential games and hence common interest ignores the behavioral component.

Chapter 3

Innovation Diffusion on Networks

3.1 Introduction

Innovation diffusion on networks is often modeled as a 2×2 symmetric coordination game played with neighbors on a network [18] [19] [12]. Recall that a coordination game is where the players have incentive to play the same or corresponding strategies, *i.e.*, to coordinate. In the case of two strategies, this yields a game with two pure strategy Nash equilibria. An important question arises concerning equilibrium refinement– which equilibrium gets played? A common approach to this involves the notion of risk-dominance. Intuitively, a pure strategy Nash equilibrium is risk-dominant if it is less risky. In other words, if the agents playing the coordination game are unsure of each other's strategies, it is expected that they will play the risk-dominant equilibrium since, in the face of uncertainty, this equilibrium is "safe." In more technical terms, the risk-dominant Nash equilibrium has a larger basin of attraction.

Furthermore, the risk-dominant pure strategy Nash equilibrium describes the long-run behavior of several dynamics like log-linear learning [4]. Because coordination games are also potential games, a potential function exists, and, as it turns out, the global maximum of the potential function is the risk-dominant equilibrium of the game (a fact that is proved in the first theorem of the chapter). Examples exist where the potential maximizing strategy profile, hence the probable outcome of many dynamics, differs from the strategy profile that maximizes social welfare, according to whichever measure. Specifically, Young [18] shows an example of this using the utilitarian measure of social welfare, which sums all payoffs in a given strategy profile.

Young uses the potential function together with the notion of close-knittedness, a graph-theoretic property, to characterize when subsets of agents will increase the potential function through their collective deviation [18] [19]. Roughly, a set of agents is close-knit if no subset has too many of their interactions with outsiders. This is a useful characteristic, as it is used to find bounds on the time it takes for the innovation to diffuse. In [19], Young uses a 1-dimensional parametrization of coordinate games in such a way that the same strategy profile globally maximizes both potential and welfare.

Newton and Sercombe point out that a collective deviation that increases the payoff of all deviating agents may actually decrease potential [12]. This has the flavor as the example originally pointed out in Young [18]. To handle this, Newton *et. al.*, give an alternative, 2-dimensional model to Young's [19] in such a way that their underlying coordination game has the same potential function as [19]. This additional parameter, they show, is not picked up by the potential function. With it, they develop a series of results analogous to those by Young [19], but instead of potential-maximizing deviating subsets, they characterize those subsets whose deviation increases the payoff of all members of the subset. Here, the graph-theoretic notion of cohesion is used, which measures how often individuals within a group interact with outsiders.

Our work from chapter 2 echoes strongly here. For instance, it was shown in 2 that the potential function does not take into account the behavioral component of the game. The social welfare function, on the other hand, picks up the behavioral component that the potential function misses. Thanks to the decomposition we already know that this will be the reason for any disagreement.

In this chapter, we provide a full description of when there is agreement and disagreement between the potential function and the welfare function. We begin by analyzing symmetric 2×2 games, as used by Young in [18] and [19], and Newton *et. al.*, in [12]. Note that the coordinate system for potential games from chapter 2 reduces to 3 dimensions (4 with the kernel) under the assumption of symmetry, going beyond the 1-dimensional game in [19] and 2-dimensional game in [12]. All of this is summarized in Figures 3.2 and 3.3. Then, we follow a similar analysis for broken α and β symmetries, allowing each agent i a different individual preference α_i and a different externality term β_i. This, in turn, gives a more general 5-dimensional model (7 with the kernel). To conclude the chapter, we study the symmetric game on networks before putting the notions of close-knittedness and cohesion into the language of the coordinate system.

3.2 Disagreement in Potential Games

In this section we use the coordinate system developed in chapter 2 to examine potential disagreement between the potential function and the social welfare function of a potential game. We focus on symmetric 2×2 potential games to gain intuition. We provide a brief discussion of the coordinates[1] and then partition this coordinate space into regions that give rise to the different flavors of agreement and disagreement. After an exposition of the symmetric 2×2 case, we summarize the regions in Figures 3.2 and 3.3. Then, we break both the individual preference symmetry and the externality symmetry– both acceptable asymmetries in potential games. From here we proceed to extend the results to networks.

[1] A more complete discussion can be found in chapter 2.

3.2.1 Symmetric 2×2 Potential Games

First note that all 2×2 symmetric games are potential games. This is simply because, by default, $\gamma_{12} = \gamma_{21}$, a necessary and sufficient condition for a game to be a potential game (Thm. 2.8). In a symmetric 2×2 game, ignoring the kernel, the coordinate system reduces to three dimensions. Namely, both agents have the same individual preference, α, along with the usual same coordinative pressure, γ, and the same externality value β. As explained in chapter 2, the exclusion of the kernel means that all components are centered around zero. The game split into the components generated by these coordinates is given in Table 3.1.

	$+1$	-1
$+1$	α, α	α, $-\alpha$
-1	$-\alpha$, α	$-\alpha$, $-\alpha$

Individual Preference

	$+1$	-1
$+1$	γ, γ	$-\gamma$, $-\gamma$
-1	$-\gamma$, $-\gamma$	γ, γ

Coordinative Pressure

	$+1$	-1
$+1$	β, β	$-\beta$, β
-1	β, $-\beta$	$-\beta$, $-\beta$

Externalities

Table 3.1: Symmetric Potential Game

In chapter 2 we painted a detailed picture of the role each of these coordinates play in the payoff structure of the game. These features are highlighted here briefly. The individual preference component, α, gives agent i the payoff αt_i independent of what the other agent plays. In other words, for any particular agent, the utility from the individual preference component is dependent only on the action of the same agent.

The externality coordinate, β, like α, is dependent on only one agent's choice of strategy.

Unlike α, this agent is not the one receiving the payoff– hence the name externality. In other words, an agent does not have direct control over this component of their utility– it is invariant over their own action and instead entirely dependent on the action of who they are playing with.

Lastly, the coordinative pressure coordinate γ is dependent on the choice of both agents, and contributes to the payoff structure pressures to conform or not conform, depending on its sign. Indeed, when positive, γ enhances the coordinative profiles, and, when negative, detracts from them. This value is invariant over what is coordinated on. This payoff simply comes from the value inherent to conforming no matter what is actually being coordinated on.

What we would like to do now is explore the tension of individualistic and cooperative forces in the game. For the individualistic forces, we focus on unilateral deviations and make use of the potential function. The potential function is useful here because its local maxima are pure strategy Nash equilibria, and its global maximum has been shown to describe the long-run behavior of the noisy best response dynamic and log-linear learning [12]. The reason for this is that these dynamics oftentimes serve as equilibrium refinements tools that return the risk-dominant equilibrium, which globally maximizes the potential function. This is stated as a theorem for 2×2 potential games below.

Theorem 3.1. *Let $\mathcal{G}$ be a 2×2 potential game. Then the potential function of $\mathcal{G}$ is globally maximized at the strategy profile (t', t'') if and only if (t', t'') is the risk-dominant Nash equilibrium of $\mathcal{G}$.*

Proof. There are two cases to consider. In the first case, there is a unique pure strategy Nash equilibrium. This case is trivial, since by default this unique equilibrium is risk-dominant and is the unique, hence global, maximum of the potential function. The second case involves two pure strategy Nash equilibria. Here we borrow a result from Harsanyi and Selten, stating that one of two pure strategy Nash equilibria is risk-dominant if and only if the product of the deviation loss for each agent from this equilibrium is greater than the same product for the other equilibrium.

Suppose, then, that $\mathcal{G}$ is a 2×2 potential game with pure strategy Nash equilibria (t', t'') and $(-t', -t'')$. Furthermore, suppose that the potential function is globally maximized at (t', t''). This requires either $P(t', t'') > P(-t', -t'') > P(t', -t'') > P(-t', t'')$, or $P(t', t'') > P(-t', -t'') > P(-t', t'') > P(t', -t'')$, because $P(-t', -t'')$ is a local maximum of the potential function. Using the potential function offered in chapter 2, this is true if and only if $\alpha_1 t' + \alpha_2 t'' > 0$, and either $\gamma t' t'' > \alpha_1 t' > \alpha_2 t''$ or $\gamma t' t'' > \alpha_2 t'' > \alpha_1 t'$, which is true by default since the game has two Nash equilibria.[2]

Now, using the result of Harsanyi and Selten, the profile (t', t'') is the risk-dominant Nash equilibrium of the game if and only if $(-2\alpha_1 t' - 2\gamma t' t'')(-2\alpha_2 t' - 2\gamma t' t'') > (2\alpha_1 t' - 2\gamma t' t'')(2\alpha_2 t' - 2\gamma t' t'')$, which simplifies to $\gamma t' t''(\alpha_1 t' + \alpha_2 t'') > 0$. When $\gamma > 0$, the game is a coordination game so the product $t' t'' = 1$. Hence, $(\alpha_1 t' + \alpha_2 t'') > 0$. When $\gamma < 0$, the game is an anti-coordination game, so the product $t' t'' = -1$. Hence $\gamma t' t'' > 0$, which implies $(\alpha_1 t' + \alpha_2 t'') > 0$. □

The importance of this theorem is that we can use, as a measure of the outcome of the individualistic forces, the global maximum of the potential function.

On the other hand, the utilitarian social welfare function– which sums the payoffs in every strategy profile– is a measure of what a successfully cooperative group would obtain. Henceforth in this thesis we will refer to the utilitarian social welfare function as simply the social welfare function. The results in this chapter are coupled to the social welfare function but similar work can be done for alternative measures.

[2] These games fall into the dependent class of potential games, which was shown must have $|\gamma| > |\alpha_1|, |\alpha_2|$.

To explore this tension between individualistic and social forces, then, we analyze the regions in the parameter space that lead to agreement and disagreement between the global maximum of the potential function and the social welfare function. The potential function, $P(t_1, t_2)$, can be written as in (3.1) below using Theorem 2.7 of chapter 2. We re-emphasize that the potential function does not include the pure externality β.

$$P(t_1, t_2) = \alpha(t_1 + t_2) + \gamma t_1 t_2 \tag{3.1}$$

The social welfare function is defined to be the sum of all payoffs in each strategy profile. In other words, $w(t_1, t_2) = \pi_1(t_1, t_2) + \pi_2(t_1, t_2)$. We can then write,

$$w(t_1, t_2) = (\alpha + \beta)(t_1 + t_2) + 2\gamma t_1 t_2. \tag{3.2}$$

Before we begin, notice there is a connection between the potential function and the social welfare function.

Remark 3.1. If $\alpha = \beta$ then $w(t_1, t_2) = 2P(t_1, t_2)$.

In terms of the orderings induced by the different regions, when $\alpha = \beta$, w and P become indistinguishable.[3] That is, the two functions have the same orderings in the same regions. Immediately we can tell that any disagreement between the two functions must come from β influencing the welfare function and shifting its regions while leaving those of the potential function invariant.

This allows us to use Remark 3.1 to take a short-cut. By uncovering the regions that give rise to the possible orderings of the social welfare function, the results for the potential function will follow by simply taking $\alpha = \beta$. This causes no issues since the potential function is

[3] This is because for any two strategy profiles (t'_1, t'_2) and (t^*_1, t^*_2), when $\alpha = \beta$, $w(t'_1, t'_2) > w(t^*_1, t_2*) \Rightarrow P(t'_1, t'_2) > P(t^*_1, t_2*)$.

independent of β. Before proceeding, we describe some useful symmetries that simplify the analysis.

Let us define two maps. Define $F(\alpha, \gamma, \beta) = (-\alpha, \gamma, -\beta)$, which reflects the individual preferences and the externalities, and define $G(\alpha, \gamma, \beta) = (-\alpha, -\gamma, -\beta)$, which reflects all three parameters, α, γ, and β. It is immediate that both functions have order 2, *i.e.*, $F^2(\alpha, \gamma, \beta) = (\alpha, \gamma, \beta)$ and $G^2(\alpha, \gamma, \beta) = (\alpha, \gamma, \beta)$, where the superscript 2 is an exponent.

In [8], a map of the structure of all 2×2 games is provided, and here we translate this map to the coordinate system restricted to the subspace of potential games. This captures the pure strategy Nash equilibrium structure of all possible symmetric 2×2 games. Divide the $\alpha\gamma$-plane into 8 regions with the lines $\alpha = 0$, $\gamma = 0$, $\gamma = \alpha$, and $\gamma = -\alpha$. The first two lines are to provide insight into the effects on the game of changing the signs of α and γ. The other two lines, in the order that they are written, make the payoffs zero at $(+1, -1)$ and $(-1, +1)$, when $\gamma = \alpha$, or at $(+1, +1)$ and $(-1, -1)$, when $\gamma = -\alpha$. These regions are plotted in Figure 3.1, where the α-axis is horizontal, and the γ-axis is vertical.

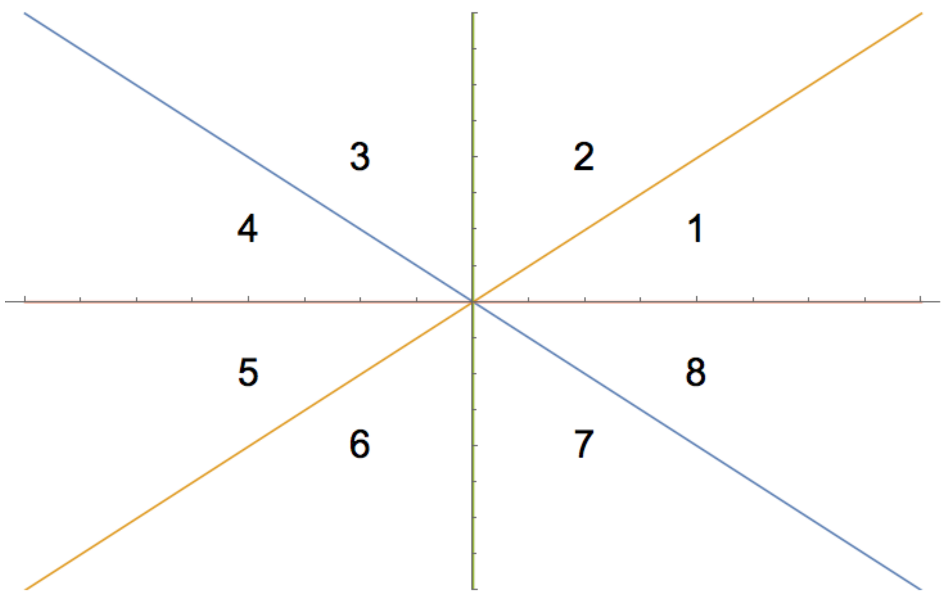


Figure 3.1: $\alpha\gamma$-plane

The regions are the following: (1) $\alpha > \gamma > 0$, (2) $\gamma > \alpha > 0$, (3) $\gamma > -\alpha > 0$, (4)

$-\alpha > \gamma > 0$, (5) $-\alpha > -\gamma > 0$, (6) $-\gamma > -\alpha > 0$, (7) $-\gamma > \alpha > 0$, and (8) $\alpha > -\gamma > 0$.

The effect of F on a Nash component is that the strategies get renamed, namely, $+1 \mapsto -1$ and $-1 \mapsto +1$. Applying G to a game multiplies the Nash component by -1. Alternatively, F switches both rows and columns for both agents, and G switches rows for agent 1 and columns for agent 2. Let us demonstrate this. Suppose that a fixed choice of (α, γ) produces the normal-form Nash component in Table 3.2.

	$+1$		-1	
$+1$	a	a	b	$-a$
-1	$-a$	b	$-b$	$-b$

Table 3.2: Arbitrary Nash Component of a Game $\mathcal{G}$ in Normal-Form

It is easily verified that the effects of F and G on the normal-form representation of $\mathcal{G}$ are as described in Table 3.3.

	$+1$		-1	
$+1$	$-b$	$-b$	$-a$	b
-1	b	$-a$	a	a

	$+1$		-1	
$+1$	$-a$	$-a$	$-b$	a
-1	a	$-b$	b	b

Table 3.3: $F(\mathcal{G})$, and $G(\mathcal{G})$

Repeated application of the functions F and G in the 8 regions creates two unique cycles. This gives $1 \leftrightarrow_F 4 \leftrightarrow_G 8 \leftrightarrow_F 5 \leftrightarrow_G 1$ and $2 \leftrightarrow_F 3 \leftrightarrow_G 7 \leftrightarrow_F 6 \leftrightarrow_G 2$. This means that if region 1 is understood, an application of F, or, equivalently, switching the strategies, extends this understanding to region 4. Then, whatever conclusions are obtained, an application of G, or, equivalently, multiplying all payoffs by -1, extends the understanding to region 8. Continuing with this, regions 1, 4, 8, and 5 are understood. Similarly with 2, 3, 7, and 6. Hence, we need only understand region 1, where $\alpha > \gamma > 0$, and region 2, where $\gamma > \alpha > 0$.

In region 1 there is a unique pure Nash equilibrium at $(+1, +1)$, and in region 2 there are two pure Nash equilibria, at $(+1, +1)$, and $(-1, -1)$. Applying F and G uncovers the Nash equilibrium structure for symmetric games in each region. The result of this is stated as a theorem, but first we prove in a lemma that F and G preserve the number of pure strategy Nash equilibria.

Lemma 3.1.1. *F and G preserve the number of pure strategy Nash equilibria.*

Proof. Because F has the effect of relabeling the strategy names, it is immediate that F preserves the number of pure strategy Nash equilibria. Now, notice that a game with one pure strategy Nash equilibria must have exactly one strategy profile with all negative entries in the Nash component of the game. Because G has the effect of multiplying all payoffs by -1, this unique all-negative strategy profile will become the unique Nash equilibrium of the game. When there are two pure strategy Nash equilibria, the remaining strategy profiles must have all negative entries. Because of this, G will simply swap the pure strategy Nash equilibria with the strategy profiles with all-negative entries. This completes the proof. □

Theorem 3.2. *In regions 1, 4, 8, and 5, the unique pure strategy Nash equilibrium is* $(\mathrm{sgn}(\alpha), \mathrm{sgn}(\alpha))$. *In regions 2, 3, 7, and 6, there are two pure strategy Nash equilibria,* $(+1, +1)$ *and* $(-1, -1)$ *when* $\gamma > 0$ *(regions 2 and 3), and* $(+1, -1)$ *and* $(-1, +1)$ *when* $\gamma < 0$ *(regions 6 and 7).*

Proof. In region 1 we have that $\alpha > \gamma > 0$. Hence the only strategy profile with all positive entries is $(+1, +1)$, and this is the unique pure strategy equilibrium of the game. Successive application of F and G will preserve the Nash equilibrium structure, implying that regions 4, 8, and 5 have a unique pure strategy Nash equilibrium. Region 2 is given by $\gamma > \alpha > 0$. Here there are two pure strategy Nash equilibria, at $(+1, +1)$ and $(-1, -1)$. Similar to before, successive application of F and G will preserve the Nash equilibrium structure. Because of this, regions 3, 7, and 6 also have two pure strategy Nash equilibria. In regions 2 and

3, because $\gamma > 0$, it is a simple computation to check that the equilibria are $(+1, +1)$ and $(-1, -1)$. Similarly, in regions 7 and 6, because $\gamma < 0$, the equilibria are $(+1, -1)$ and $(-1, +1)$. □

This theorem details the relationships between α and γ that give rise to all possible pure strategy Nash equilibrium structures in symmetric 2×2 potential games. This is important because of our goal of understanding the results of individualistic forces in a potential game, which we capture with the game's pure strategy Nash equilibrium structure.

Although 8 regions were defined in Figure 3.1, there are only four possible distributions of pure strategy Nash equilibria. More specifically, regions 1 and 8 both have the unique pure strategy Nash equilibrium $(+1, +1)$, and regions 4 and 5 have it at $(-1, -1)$. On the other hand, regions 2 and 3 both have pure strategy Nash equilibria at $(+1, +1)$ and $(-1, -1)$, while regions 6 and 7 have them at $(+1, -1)$ and $(-1, +1)$. This is showing that in regions 1 and 8, and 4 and 5, the mapping $\gamma \mapsto -\gamma$ does not change the Nash equilibrium structure. For regions 2 and 3, and 6 and 7, the mapping $\alpha \mapsto -\alpha$ does not change the structure.

Now that the Nash equilibrium structure of symmetric 2×2 games is understood, let us move on to the potential function and the social welfare function. We begin by stating a theorem describing the effects of F and G on these functions.

Theorem 3.3. *Let $\mathcal{G}$ be a symmetric (hence potential) game whose Nash and externality components are constructed with (α, γ, β), and where $\kappa = 0$. Denote by $F(\mathcal{G})$ the game obtained by applying F to (α, γ, β), and by $G(\mathcal{G})$ the game obtained by applying G to (α, γ, β). Denote the social welfare function of $\mathcal{G}$ by $w_{\mathcal{G}}$, and the following two by $w_{F(\mathcal{G})}$ and $w_{G(\mathcal{G})}$, respectively. Furthermore, denote the potential function of $\mathcal{G}$ by $P_{\mathcal{G}}$, and the following two by $P_{F(\mathcal{G})}$ and $P_{G(\mathcal{G})}$, respectively. Then*

$$w_{\mathcal{G}}(t_1, t_2) = w_{F(\mathcal{G})}(-t_1, -t_2) = -w_{G(\mathcal{G})}(t_1, t_2), \tag{3.3}$$

and

$$P_{\mathcal{G}}(t_1, t_2) = P_{F(\mathcal{G})}(-t_1, -t_2) = -P_{G(\mathcal{G})}(t_1, t_2). \tag{3.4}$$

Proof. This is a straightforward, though tedious, verification. □

Using Theorem 3.3, all possible strict orderings are covered by considering just the regions that induce $w(+1, +1) > w(+1, -1) > w(-1, -1)$ and $w(+1, +1) > w(-1, -1) > w(+1, -1)$, and then applying F and G to find the remaining regions.

The reason the strategy profile $(-1, +1)$ is omitted is that, because of the assumption of symmetry, both $w(+1, -1) = w(-1, +1)$ and $P(+1, -1) = P(-1, +1)$. This is stated as a remark.

Remark 3.2. $w(+1, -1) = w(-1, +1)$ and $P(+1, -1) = P(-1, +1)$

Hence, in this section, considering only the symmetric case, we refer to both $(+1, -1)$ and $(-1, +1)$ when we mention $(+1, -1)$.

The below theorem describes the regions that induce all possible strict orderings of the welfare function.

Theorem 3.4. *(i) The region where* $w(+1, +1) > w(+1, -1) > w(-1, -1)$ *is given by* $\alpha + \beta > 2|\gamma|$.
(ii) The region where $w(+1, +1) > w(-1, -1) > w(+1, -1)$ *is given by* $2\gamma > \alpha + \beta > 0$.
(iii) The region where $w(-1, -1) > w(+1, +1) > w(+1, -1)$ *is given by* $0 > \alpha + \beta > -2\gamma$.
(iv) The region where $w(-1, -1) > w(+1, -1) > w(+1, +1)$ *is given by* $\alpha + \beta < -2|\gamma|$.
(v) The region where $w(+1, -1) > w(-1, -1) > w(+1, +1)$ *is given by* $0 > \alpha + \beta > 2\gamma$.
(vi) The region where $w(+1, -1) > w(+1, +1) > w(-1, -1)$ *is given by* $-2\gamma > \alpha + \beta > 0$.

Proof. This involves checking the regions for $w(+1,+1) > w(+1,-1) > w(-1,-1)$ and $w(+1,+1) > w(-1,-1) > w(+1,-1)$, and then applying F and G to find the remaining regions, which is basic algebra. □

The importance of the above theorem is its detailing of the regions that induce different orderings of the welfare function. This is valuable because of our goal of understanding the results of cooperative forces in a potential game, in contrast to individual forces. These regions are defined by the quantity $(\alpha+\beta)$'s position in relation to $\pm 2\gamma$ and 0. What does this mean? This question is answered in the following corollaries.

Corollary 3.4.1. *When $\alpha+\beta > 0$ it is impossible for $w(-1,-1) > w(+1,+1)$. Alternatively, when $\alpha+\beta < 0$ it is impossible for $w(+1,+1) > w(-1,-1)$.*

Proof. This is an immediate consequence of Theorem 3.4 □

In regions (i), (ii), and (vi), $\alpha+\beta > 0$. This means that $\alpha+\beta$ is supporting the profile $(+1,+1)$, and in all of these cases, the social welfare at $(+1,+1)$ is superior to that at $(-1,-1)$. In region (vi), although $\alpha+\beta > 0$, it is bounded above by -2γ. This implies that $\gamma < 0$, meaning there are pressures to not conform, and, this pressure is greater than the benefit of $\alpha+\beta$. Hence, in (vi), although $(+1,+1)$ is superior to $(-1,-1)$, the profile that maximizes the welfare function is $(+1,-1)$. In regions (iii), (iv), and (v), $\alpha+\beta < 0$, which means that, no matter what, $(-1,-1)$ will be superior to $(+1,+1)$. Like before, in region (v), although $\alpha+\beta < 0$, it is bounded below by 2γ, and here it is $(+1,-1)$ that maximizes the welfare function.

Note that regions (v) and (vi) are the only regions where $(+1,-1)$ is the global maximum of the social welfare function. In these regions $\gamma < 0$. Consequently, in the symmetric case it is impossible to have $(+1,-1)$ maximize social welfare if the pressure to conform is positive. A corollary is stated in this regard.

Corollary 3.4.2. *If the social welfare function is globally maximized at* $(+1,-1)$ *then* $\gamma < 0$.

Proof. This is an immediate consequence of regions (v) and (vi) in Theorem 3.4. □

To transfer these results to the potential function, set $\alpha = \beta$. These results are summarized in the below theorem.

Theorem 3.5. *The region where* $P(+1,+1) > P(+1,-1) > P(-1,-1)$ *is given by* $\alpha > |\gamma|$ *(regions 1 and 8).*
The region where $P(+1,+1) > P(-1,-1) > P(+1,-1)$ *is given by* $\gamma > \alpha > 0$ *(region 2).*
The region where $P(-1,-1) > P(+1,+1) > P(+1,-1)$ *is given by* $0 > \alpha > -\gamma$ *(region 3).*
The region where $P(-1,-1) > P(+1,-1) > P(+1,+1)$ *is given by* $-|\gamma| > \alpha$ *(regions 4 and 5).*
The region where $P(+1,-1) > P(-1,-1) > P(+1,+1)$ *is given by* $0 > \alpha > \gamma$ *(region 6).*
The region where $P(+1,-1) > P(+1,+1) > P(-1,-1)$ *is given by* $-\gamma > \alpha > 0$ *(region 7).*

Proof. This is an immediate consequence of Remark 3.1 together with Theorem 3.4. □

Theorem above translates the results of Theorem , which details the structure of pure Nash equilibria in a potential game, into statements about the potential function.

The corollaries stated above for the welfare function naturally extend to the potential function. What follows are additional corollaries doing precisely this.

Corollary 3.5.1. *When* $\alpha > 0$ *it is impossible for* $P(-1,-1) > P(+1,+1)$*. Alternatively, when* $\alpha < 0$ *it is impossible for* $P(+1,+1) > P(-1,-1)$*.*

Proof. This is an immediate consequence of Theorem 3.2.1. □

Corollary 3.5.2. *If the potential function is globally maximized at* $(+1,-1)$ *then* $\gamma < 0$.

Proof. This is an immediate consequence of regions 6 and 7 in Theorem 3.2.1. □

Looking into Theorem 3.2.1, notice that the pairs of regions 1 and 8, and 4 and 5, have the same effect on the ordering of the potential function. This is because the potential function does not have a boundary at $\gamma = 0$. Another way to see this is that since the potential function is maximized at the pure strategy Nash equilibrium, its value there, at $(+1, +1)$ or $(-1, -1)$, must be greater than it's value at $(+1, -1)$ and $(-1, +1)$. Then, the agents have a unilateral incentive to deviate to $(+1, -1)$ and $(-1, +1)$ from $(-1, -1)$ when the equilibrium is $(+1, +1)$, and from $(+1, +1)$ when the equilibrium is $(-1, -1)$. Otherwise, these would be also be, in their respective cases, a pure strategy Nash equilibrium.

Games in regions 2 and 3, and 6 and 7, have two pure strategy Nash equilibria, $(+1, +1)$ and $(-1, -1)$ for 2 and 3, and $(+1, -1)$ and $(-1, +1)$ for 6 and 7. In region 2, $(+1, +1)$ is the superior Nash equilibrium, as it is supported by the agents' individual preferences. Hence, the incentive to deviate from $(+1, -1)$ or $(-1, +1)$ to $(+1, +1)$ is greater than that to $(-1, -1)$. For region 3 things are essentially the same, with the sign of the individual preferences flipped, making the equilibrium $(-1, -1)$ superior. Similarly, in region 6, although the pure strategy Nash equilibria are at $(+1, -1)$ and $(-1, +1)$, the individual preferences point in the direction of strategy -1, so the incentive to deviate from $(+1, +1)$ is greater than the incentive to deviate from $(-1, -1)$. This is exactly what's reflected in the ordering of the potential function. For region 7, the incentive to deviate from $(-1, -1)$ is greater than that from $(+1, +1)$.

Let us formalize the above arguments and connect the regions of the potential function with the Nash equilibrium structure of 2×2 symmetric games.

Theorem 3.6. *For a symmetric 2×2 game $\mathcal{G}$ the following four statements are equivalent.*

1. $t\alpha > |\gamma|$

2. *$P(t, t)$ is the unique maximum of the potential function.*

3. *(t, t) is the unique Nash equilibrium of $\mathcal{G}$.*

Proof. (1) is true if and only if the only strategy profile with all positive entries in the Nash component is (t, t). Hence, (t, t) is the unique Nash equilibrium of $\mathcal{G}$. Hence, (1) $\iff$ (3). Now, suppose (t, t) is the unique pure Nash equilibrium of $\mathcal{G}$. This is true if and only if (t, t) is a local maximum of the potential function. If there were additional local maxima, (t, t) would not be the unique pure Nash equilibrium. Hence, (t, t) is the global maximum of P. Hence, (3) $\iff$ (2), and this completes the proof. □

In other words, $(+1, +1)$ is the unique equilibrium when α is positive and is greater than the pressure to conform or not conform. On the other hand, $(-1, -1)$ is the unique equilibrium when α is negative and is greater in magnitude than the pressure to conform or not conform.

Theorem 3.7. *For a symmetric 2×2 game $\mathcal{G}$ the following three statements are equivalent.*

1. *$P(+1, +1)$ and $P(-1, -1)$ are local maxima of the potential function*

2. *$|\alpha| < \gamma$.*

3. *$(+1, +1)$ and $(-1, -1)$ are the only pure Nash equilibria of $\mathcal{G}$.*

Moreover, if $\alpha \neq 0$, then $(\mathrm{sgn}(\alpha), \mathrm{sgn}(\alpha))$ is the risk-dominant Nash equilibrium.

Proof. (1) is true if and only if $P(+1, +1) > P(+1, -1)$ and $P(-1, -1) > P(1, -1)$ if and only if $\alpha > -\gamma$ and $\alpha < \gamma$ if and only if $|\alpha| < \gamma$. Hence, (1) $\iff$ (2). Furthermore, this is true if and only if the entries in $(+1, -1)$ and $(-1, +1)$ of the Nash component are all negative. Hence (1) $\iff$ (2) $\iff$ (3).

Now suppose $\alpha \neq 0$. It is easy to verify that $P(\operatorname{sgn}\alpha, \operatorname{sgn}\alpha) > P(-\operatorname{sgn}\alpha, -\operatorname{sgn}\alpha)$. Hence the strategy profile $(\operatorname{sgn}\alpha, \operatorname{sgn}\alpha)$ is the global maximum of the potential function, which means it is the risk-dominant Nash equilibrium of $\mathcal{G}$. □

Here the magnitude of the individual preference is bounded by the pressure to conform, which is positive. This makes both $(+1, +1)$ and $(-1, -1)$ pure strategy Nash equilibria of the game. The individual preference, then, being positive or negative, can support either of these equilibrium; $(+1, +1)$ when positive, and $(-1, -1)$ when negative.

The below theorem considers the last case, where $(+1, -1)$ and $(-1, +1)$ are the pure Nash equilibria of the game.

Theorem 3.8. *For a symmetric* 2×2 *game* $\mathcal{G}$ *the following three statements are equivalent.*

1. $P(+1, -1)$ *and* $P(-1, +1)$ *globally maximize the potential function.*
2. $\gamma < 0$ *and* $|\alpha| < -\gamma$.
3. $(+1, -1)$ *and* $(-1, +1)$ *are the only pure Nash equilibria of* $\mathcal{G}$.

Proof. This proof follows the same process as the proof for Theorem 3.7. □

This theorem is similar to Theorem 3.7 where the magnitude of the individual preference is less than the magnitude of the pressure to conform, but here $\gamma < 0$. The effect of this is that the pure strategy Nash equilibria are $(+1, -1)$ and $(-1, +1)$, rather than $(+1, +1)$ and $(-1, -1)$.

We are now ready to describe the regions of agreement and disagreement. To begin, let us define the regions of *coordinative agreement*, *coordinative tension*, *unilateral tension*, and *anti-coordinative agreement*.

The region of *coordinative agreement* is where both the maximum of the potential function and the social welfare function agree on either $(+1, +1)$ or $(-1, -1)$. The region of *coordinative tension* contains the games where there is tension between the two coordinative equilibria, that is, $P(t_1, t_2)$ is maximized at $(+1, +1)$ while $w(t_1, t_2)$ is maximized at $(-1, -1)$, or vice-versa. On the other hand, the class of *unilateral tension* is when the functions maximize neighboring strategy profiles, like $(+1, +1)$ and $(+1, -1)$. Lastly, the class of *anti-coordination agreement* is when both $P(t_1, t_2)$ and $w(t_1, t_2)$ are maximized at $(+1, -1)$ and $(-1, +1)$.

The following theorem relates the regions defined above with the parameters α, γ, and β.

Theorem 3.9. *For $t \in \{-1, 1\}$, the region of coordinative agreement is given by both*

$$t\alpha > t\max(0, -\gamma)$$
$$t(\alpha + \beta) > t\max(0, -2\gamma)$$

Moreover, if $t\alpha > |\gamma|$ then this game has a unique pure strategy Nash equilibrium at (t, t). Otherwise the game has two pure strategy Nash equilibria, at (t, t) and $(-t, -t)$.

The region of coordinative tension is given by both

$$t\alpha > t\max(0, -\gamma)$$
$$t(\alpha + \beta) < t\max(0, -2\gamma)$$

The region of unilateral tension is given by $\gamma < 0$ and either both

$$t\alpha > -\gamma$$
$$|\alpha + \beta| < -2\gamma,$$

or both

$$|\alpha| < -\gamma$$

$$t(\alpha + \beta) > -2\gamma.$$

The region of anti-coordinative agreement is given by $\gamma < 0$ *and both*

$$|\alpha| < -\gamma$$

$$|\alpha + \beta| < -2\gamma$$

Where the first inequality in these regions describes the global maximum of the potential function, and the second describes the global maximum of the social welfare function.

Proof. The proof of this theorem amounts to combining the regions givens in Theorems 3.4 and 3.2.1 to give rise to the desired behaviors of P and w as dictated by the defined regions. □

As mentioned earlier, the disagreements come precisely when the β values shift the regions of the welfare function while leaving the regions of the potential function invariant. Because all regions in Theorem 3.9 are defined for arbitrary $t \in \{+1, -1\}$ let us take $t = +1$ for simplicity in our discussion of these results. In this case, let us, in addition, take $\alpha = \beta$. Here, the potential function and the welfare function agree, and the region defined by α, β, and γ must be either the region of coordinative agreement or the region of anti-coordinative agreement. An extensive discussion of the regions is given, and the results are summarized in Figure 3.2 for $\gamma \geq 0$, and in Figure 3.3 for $\gamma < 0$.

Starting with the region of coordinative agreement, it must be that $\alpha > \max(0, -\gamma)$ and $(\alpha + \beta) > \max(0, -2\gamma)$. Again, since we are taking $\alpha = \beta$, these inequalities are exactly the same. Increasing β will only increase the quantity $\alpha + \beta$ and hence, never change the regions. However, decreasing β enough will flip the sign of the second inequality, meaning that the welfare function will be maximized elsewhere.

When $\gamma > 0$ the only other possibility is for the welfare function to be maximized at the diametrically opposite strategy profile $(-1, -1)$. Here notice that once we decrease β enough so that $\beta < -\alpha$, the welfare function is maximized at $(-1, -1)$ while the potential function is maximized at $(+1, +1)$. In other words, there is coordinative tension. What is the story here? When $\beta < -\alpha$, the externality is not only opposite to the individual preference, its magnitude is also greater, *i.e.*, $|\beta| > |\alpha|$. In such a situation, agents who follow only their unilateral incentives, a process we are modeling with the potential function, will play a Nash equilibrium with lower welfare than the strategy profile diametrically opposite to it.

On the other hand, when $\gamma < 0$, there are several possibilities. Starting with $\alpha = \beta$, we must either have coordinative agreement or anti-coordinative agreement. From the region of coordinative agreement, decreasing β after an initial threshold will give rise to unilateral tension, where P is maximized at $(+1, +1)$ but w at $(+1, -1)$. Decreasing β further will cross the second threshold and give rise to coordinative tension, where P is maximized at $(+1, +1)$ but w at $(-1, -1)$.

This discussion is hinting at a bifurcation that happens when γ goes from $\gamma \geq 0$ to $\gamma < 0$. The reason for this bifurcation is that it is impossible for the potential and social welfare functions to be maximized at $(+1, -1)$ and $(-1, +1)$ when $\gamma \geq 0$. When $\gamma < 0$, however, this regions become possible. This is made evident in Figures 3.2 and 3.3.

We now have a good understanding of what selfish behavior can lead to, and when this does and doesn't agree with the utilitarian measure of social welfare for 2×2 games. The literature has tentatively recognized but has not characterized where there are differences in these measures, which we are motivating through the example given in [18] and the following

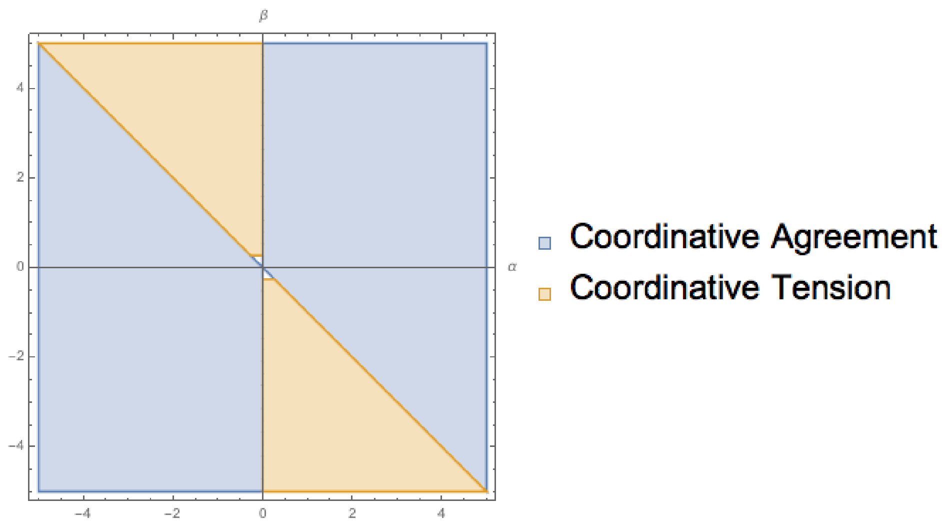


Figure 3.2: $\alpha\beta$-plane with $\gamma \geq 0$

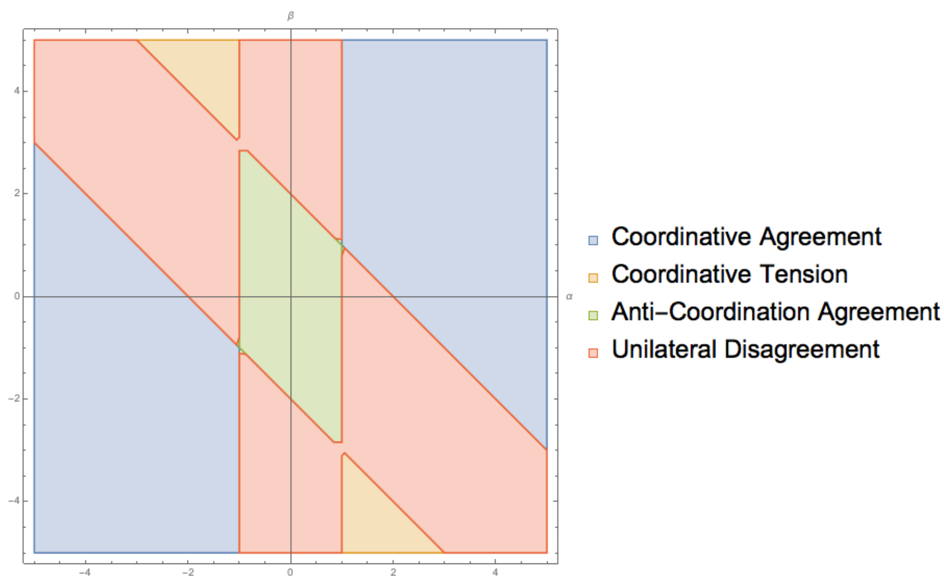


Figure 3.3: $\alpha\beta$-plane with $\gamma < 0$

work in [19] and [12]. In the theorems of this section we explicitly stated where the precise regions in the parameter space that give rise to these disagreements and their boundaries. All of this is culminated in Figures 3.2 and 3.3.

We will now break the α symmetry, before breaking the β symmetry, to understand the regions of agreement and disagreement for general potential games.

3.3 Breaking α Symmetry

Breaking the α symmetry opens up the modeling to more realistic situations without breaking free from exact potential games. For instance, it is entirely reasonable that two agents have different preferences. How do these differences change the perspective developed so far into potential games?

A game in normal-form with broken α symmetry is given in Table 3.4.

	$+1$		-1	
$+1$	$\alpha_1 + \gamma + \beta$	$\alpha_2 + \gamma + \beta$	$\alpha_1 - \gamma - \beta$	$-\alpha_2 - \gamma + \beta$
-1	$-\alpha_1 - \gamma + \beta$	$\alpha_2 - \gamma - \beta$	$-\alpha_1 + \gamma - \beta$	$-\alpha_2 + \gamma - \beta$

Table 3.4: 2×2 Broken α Symmetry

We begin with just the Nash component because this is the only pertinent information when calculating the potential function.

When $\gamma = 0$, the Nash component consists of only α_1 and α_2. The unique pure strategy Nash equilibrium is $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$. Increasing γ in either direction, positive or negative, has no effect until $|\gamma|$ surpasses $\min(|\alpha_1|, |\alpha_2|)$. At this point, the unique Nash equilibrium becomes the profile where the agent with the larger magnitude individual preference plays this preference, and the other agent follows ($\gamma > 0$) or plays the opposite of this preference ($\gamma < 0$). Once $|\gamma|$ surpasses both $|\alpha_1|$ and $|\alpha_2|$, the game becomes either a coordination game ($\gamma > 0$) or an anti-coordination game ($\gamma < 0$).

This echoes the three classes of potential games defined in chapter 2, independent, quasi-independent, and dependent potential games. The first case described, where the magnitude of γ is negligible compared to the magnitude of α_i, $i = 1, 2$, is an independent potential game. The second case, where $|\gamma|$ is between the magnitudes of α_1 and α_2, describes a quasi-independent potential game. Finally, the last case, where the magnitude of γ is supreme, belongs to the class of dependent potential games.

Let us make all of this precise.

There are 52 regions that emerge based on α_1 and α_2's relationship with themselves, and with γ and $-\gamma$. Both can be greater than $\max(\gamma, -\gamma)$, in between $\max(\gamma, -\gamma)$ and $\min(\gamma, -\gamma)$, or less than $\min(\gamma, -\gamma)$. Furthermore, they can be greater than or less than each other. 24 of the regions occur when $\gamma > 0$ and the other 24 when $\gamma < 0$. The regions in the $\alpha_1\alpha_2$-plane are plotted for arbitrary γ in Figure 3.4. When $\gamma = 0$ there are 4 additional regions, namely the quadrants of the $\alpha_1\alpha_2$-plane.

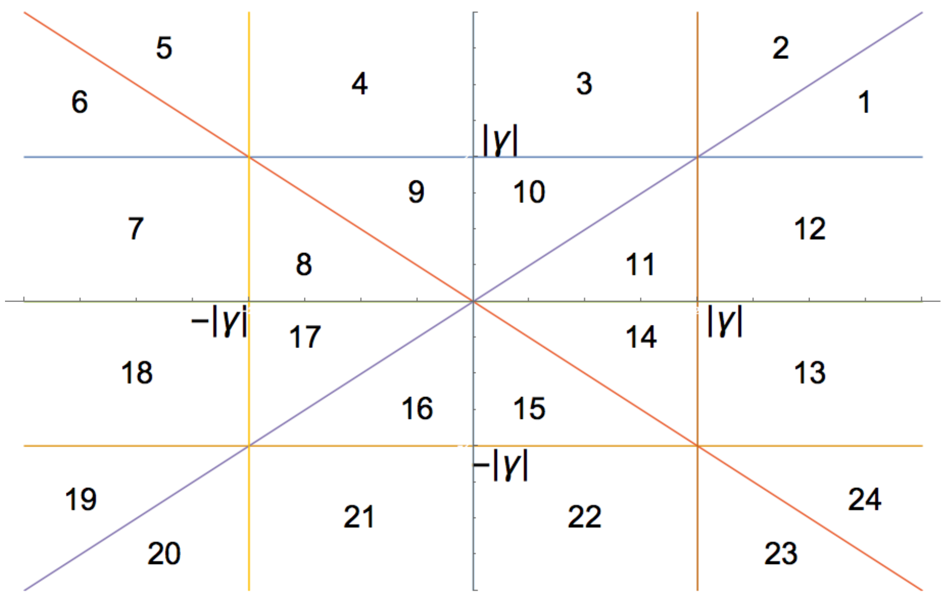


Figure 3.4: 24 regions for arbitrary γ

Suppose without loss of generality that the plot in Figure 3.4 is for $\gamma > 0$. Then, for $\gamma < 0$, re-label the regions by $1', 2', \ldots, 24'$ in the resulting plot. In the following four paragraphs by writing any number $i \in \{1, \ldots, 24\}$ we mean both i and i' because we use $|\gamma|$.

The class of independent potential games resides in regions 1, 2, 5, 6, 19, 20, 23, 24, where $|\alpha_i| > |\gamma|$ for $i = 1, 2$. In regions 1, 2, 19, and 20, the Nash equilibrium structure is qualitatively the same as in the symmetric case where $|\alpha| > |\gamma|$. In regions 5, 6, 23, and 24, α_1 and α_2 have different signs. Relabeling the strategies for one of the agents transforms games in these regions to those in regions 1, 2, 19, and 20. In all of these regions there is a unique Nash equilibrium given by $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$.

The class of quasi-independent potential games consists of regions 3, 4, 7, 12, 13, 18, 21, and 22. In regions 3, 12, 18, and 21, we have that $\operatorname{sgn}\alpha_1 = \operatorname{sgn}\alpha_2$. In these regions, when $\gamma > 0$, both α_1 and α_2 support $(+1, +1)$ when positive (regions 3 and 12), and $(-1, -1)$ when negative (regions 18 and 21). When $\gamma < 0$, the agent with the larger magnitude preference plays this preference while the other agent plays the opposite, which is also opposite to their own preference. In regions 12 and 18 the larger preference is α_1, and in 3 and 21 it is α_2. In the remaining regions, 4, 7, 13, and 22, $\operatorname{sgn}\alpha_1 \neq \operatorname{sgn}\alpha_2$. In these regions, when $\gamma > 0$, the agent with the larger preference gets to play this preference, while the other agent, having the opposite preference with magnitude dominated by $|\gamma|$, follows and conforms to the larger preference. When $\gamma < 0$, both agents play their preference.

Although the class of independent and quasi-independent potential games have a unique pure strategy Nash equilibrium, their Nash structures are qualitatively different. For the class of independent games, the strategy profile diametrically opposite to the Nash equilibrium consists of all negative payoffs. For the class of quasi-independent games, the diametrically opposite profile from the Nash equilibrium has a negative payoff for the agent with the larger preference, and a positive payoff for the agent with the smaller preference. The all-negative strategy profile is the one given by the unilateral deviation from the Nash equilibrium by the agent with the larger preference.

The class of dependent potential games is comprised of regions 8, 9, 10, 11, 14, 15, 16, and 17. In these regions, there are two pure strategy Nash equilibria. It is a coordination game when $\gamma > 0$ and an anti-coordination game when $\gamma < 0$. Because $\operatorname{sgn}\alpha_1 = \operatorname{sgn}\alpha_2$ in regions 10, 11, 16, and 17, these regions support a Nash equilibrium when $\gamma > 0$. On the other hand, when $\gamma < 0$, regions 10, 11, 16, and 17, although the agents have the same preference, the Nash equilibria are the anti-coordination profiles where only one agent plays their preference. In the remaining regions, 8, 9, 14, and 15, $\operatorname{sgn}\alpha_1 \neq \operatorname{sgn}\alpha_2$. Here, the agents can each play their preference when $\gamma < 0$. In other words, the strategy profile where the agents both play their preference is the risk-dominant Nash equilibrium. When $\gamma > 0$, only one of the agents plays their preference, and this situation represents the Bach and Stravinsky game.[4]

Because of the symmetries in taking $\gamma > 0$ and $\gamma < 0$ it turns out that most of these situations do not provide very different qualitative pictures of the Nash equilibrium structure. When alphas are different signs, the emergent situations are basically the same as in the symmetric case if we change the sign of gamma. The only new thing here is the class of quasi-independent games, which are unobtainable under the assumption of symmetry.

Lastly, we must consider the regions when $\gamma = 0$. In all of these cases, the regions give rise to independent potential games, where the unique pure strategy Nash equilibrium is $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$.

To capture all of this, we follow a similar progression as in the previous section. Suppose that a fixed choice of $(\alpha_1, \alpha_2, \gamma)$ produces the normal-form game in Table 3.5.

	$+1$		-1	
$+1$	a	b	c	$-b$
-1	$-a$	d	$-c$	$-d$

Table 3.5: Arbitrary Nash Component $\mathcal{G}^N$ in Normal-Form

We define three maps. Define $F(\alpha_1, \alpha_2, \gamma) = (\alpha_2, \alpha_1, \gamma)$. This function swaps α_1 and α_2. Define $G(\alpha_1, \alpha_2, \gamma) = (-\alpha_1, \alpha_2, \gamma)$. This function changes the sign of the first individual preference α_1. Define $H(\alpha_1, \alpha_2, \gamma) = (\alpha_1, \alpha_2, -\gamma)$. This function changes the sign of γ. The

[4] In the literature this game is most often called Battle of the Sexes. However, in an effort to disassociate this game from gender, we use the name Bach and Stravinsky.

effects of these functions on the normal-form representation of pure Nash component $\mathcal{G}^N$ are given in Table 3.6.

	$+1$	-1
$+1$	b a	d $-a$
-1	$-b$ c	$-d$ $-c$

,

	$+1$	-1
$+1$	$-c$ b	$-a$ $-b$
-1	c d	a $-d$

,

	$+1$	-1
$+1$	c d	a $-d$
-1	$-c$ b	$-a$ $-b$

Table 3.6: $F(\mathcal{G})$, $G(\mathcal{G})$, and $H(\mathcal{G})$

In other words, F switches agent 1 with agent 2, G swaps row player's payoffs and multiplies them by -1, and H swaps the strategies names and multiplies all payoffs by -1.

Now, we have $1 \leftrightarrow_F 2 \leftrightarrow_G 5 \leftrightarrow_F 24 \leftrightarrow_G 19 \leftrightarrow_F 20 \leftrightarrow_G 23 \leftrightarrow_F 6 \leftrightarrow_G 1$

$3 \leftrightarrow_F 12 \leftrightarrow_G 7 \leftrightarrow_F 22 \leftrightarrow_G 21 \leftrightarrow_F 18 \leftrightarrow_G 13 \leftrightarrow_F 4 \leftrightarrow_G 3$

$8 \leftrightarrow_F 15 \leftrightarrow_G 16 \leftrightarrow_F 17 \leftrightarrow_G 14 \leftrightarrow_F 9 \leftrightarrow_G 10 \leftrightarrow_F 11 \leftrightarrow_G 8$

Then we also have $i \leftrightarrow_H i'$ for every $1 \leq i \leq 24$.

This means that if we understand region 1, we can simply swap agents to understand region 2. Then, whatever conclusions are obtained, switching strategies of agent 2, and all positives to negatives and all negatives to positive, extends the conclusions to region 5. Continuing with this, regions 1, 2, 5, 24, 19, 20, 23, and 6 are understood. Similarly with regions 3, 12, 7, 22, 21, 18, 13, and 4, and with regions 8, 15, 16, 17, 14, 9, 10, and 11. To understand the results when $\gamma < 0$, can take any already established result, switch both agents strategies, and flip all positives to negatives and all negatives to positives. This matches exactly our discussion earlier in the section.

Let us now prove that F, G, and H do not qualitatively change the Nash equilibrium structure of a game.

Theorem 3.10. *F, G, and H do not qualitatively change the Nash equilibrium structure of a game.*

Proof. Refer to the Nash component in Table 3.5 and the effects of F, G, and H, as shown in Table 3.6. Because F has the effect of switching agent 1 and agent 2, it is immediate that F preserves the qualitative Nash equilibrium structure. Similarly, because G has the effect of swapping row agent's payoffs, G cannot change the qualitative Nash equilibrium structure of a game. Lastly, H does two things: swap the strategy names and multiply all payoffs by -1. Swapping the strategy names will produce no qualitative differences in the Nash equilibrium structure of the game. Multiplying all payoffs by -1 will turn all all-positive payoff profiles into all-negative, and vice-versa, and it will turn all profiles with a positive and a negative payoff into a profile with a negative and positive payoff, and vice-versa. Hence, H will not change the qualitative Nash equilibrium structure of the game. □

The discussion in this section is summarized in the below theorems. In the proofs we analyze only one region, and because of Theorem 3.10, F, G, and H, are used to extend the results to the remaining regions.

Theorem 3.11. *Independent potential games have a unique pure strategy Nash equilibrium, which is by default both risk-dominant and payoff-dominant, at the strategy profile* $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$.

Proof. As a representative of independent potential games, consider a game $\mathcal{G}$ in region 1 where $\alpha_1 > \alpha_2 > \gamma$. Because of this inequality, the only payoff profile with all positive entries in the Nash component is $(+1, +1)$, or $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$. Hence, this is the unique pure strategy Nash equilibrium of $\mathcal{G}$. Successive applications of the functions F, G, and H, transfers these results to the remaining regions of independent potential games, and this completes the proof. The risk and payoff-dominance of the Nash equilibrium is immediate due its uniqueness. □

Theorem 3.12. *Let $\mathcal{G}$ be a quasi-independent potential game. Without loss of generality assume $|\alpha_1| > |\alpha_2|$. Then $\mathcal{G}$ has a unique pure strategy Nash equilibrium at $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_1)$ when $\gamma > 0$, and a unique pure strategy Nash equilibrium at $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$ when $\gamma < 0$. These Nash equilibria are, by default, both risk-dominant and payoff-dominant.*

Proof. Consider a game $\mathcal{G}$ in region 3, a representative of quasi-independent potential games. Here $\alpha_2 > \gamma > \alpha_1$ and because of this, the strategy profile $(\operatorname{sgn}\alpha_2, \operatorname{sgn}\alpha_2)$ is the only payoff profile with all positive entires. Hence, it is the unique pure strategy Nash equilibrium of $\mathcal{G}$. Successive applications of F, G, and H, translates these results to the remaining regions of quasi-independent potential games, completing the proof. Like in the previous proof, the risk and payoff-dominance of the Nash equilibrium is immediate due its uniqueness. □

Theorem 3.13. *Let $\mathcal{G}$ be a dependent potential game. Without loss of generality assume $|\alpha_1| > |\alpha_2|$. Then $\mathcal{G}$ has pure strategy Nash equilibria at $(+1, +1)$ and $(-1, -1)$ when $\gamma > 0$. Furthermore, the Nash equilibrium $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_1)$ is risk-dominant. When $\gamma < 0$ the pure strategy Nash equilibria of $\mathcal{G}$ are $(+1, -1)$ and $(-1, +1)$. Furthermore, the Nash equilibrium $(\operatorname{sgn}\alpha_1, \operatorname{sgn}\alpha_2)$ is risk-dominant.*

Proof. Take region 8 as a representative of dependent potential games. In region 8, $\gamma > -\alpha_1 > -\alpha_2 > 0$. This relationship between α_1, α_2, and γ, defines two pure strategy Nash equilibria at $(+1, +1)$ and $(-1, -1)$. To show that $(-1, -1)$ is risk-dominant amounts to verifying that $(-\alpha_1+\gamma)(-\alpha_2+\gamma) > (\alpha_1+\gamma)(\alpha_2+\gamma)$, recalling a result by Harsanyi and Selten [6]. The inequality simplifies to $(\alpha_1 + \alpha_2)\gamma < 0$. The truth of this inequality is evident in region 8 due to $\gamma > 0$ and $\alpha_1, \alpha_2 < 0$. Hence, the strategy profile $(-1, -1)$ is risk-dominant. Successive applications of F, G, and H, extend these results to the remaining regions of dependent potential games, completing the proof. □

Because the potential function is globally maximized at the risk-dominant Nash equilibrium of the game, the Nash structure is fully described. What is left to do is study the welfare function with broken β symmetry.

3.4 Breaking β Symmetry

For a game with broken α and β symmetry, the payoff functions becomes

$$\pi_1(t_1, t_2) = \alpha_1 t_1 + \gamma t_1 t_2 + \beta_1 t_2$$
$$\pi_2(t_1, t_2) = \alpha_2 t_2 + \gamma t_1 t_2 + \beta_2 t_1$$

Then,

$$w(t_1, t_2) = \pi_1(t_1, t_2) + \pi_2(t_1, t_2) = (\alpha_1 + \beta_2)t_1 + (\alpha_2 + \beta_1)t_2 + 2\gamma t_1 t_2. \tag{3.5}$$

The below theorem details when the welfare function is maximized at all possible strategy profiles of a 2×2 game $\mathcal{G}$.

Theorem 3.14. *The social welfare function given in (3.5) is maximized at* $(+1, +1)$ *if and only if* $\alpha_1 + \beta_2 + \alpha_2 + \beta_1 > 0$ *and* $\beta_i + \alpha_{\neg i} > -2\gamma$*, where* $i = 1, 2$ *and* $\neg i$ *denotes the agent who is not* i*. The welfare function is maximized at* $(+1, -1)$ *if and only if* $\alpha_2 + \beta_1 < -2\gamma, \alpha_1 + \beta_2 > 2\gamma$*, and* $\alpha_1 + \beta_2 > \alpha_2 + \beta_1$*. The welfare function is maximized at* $(-1, +1)$ *if and only if* $\alpha_1 + \beta_2 < -2\gamma, \alpha_2 + \beta_1 > 2\gamma$*, and* $\alpha_2 + \beta_1 > \alpha_1 + \beta_2$*. Finally, the welfare function is maximized at* $(-1, -1)$ *if and only if* $\alpha_1 + \beta_2 + \alpha_2 + \beta_1 < 0$ *and* $\beta_i + \alpha_{\neg i} < 2\gamma$*, for* $i = 1, 2$*.*

Proof. This amounts to checking and simplifying the inequalities generated by requiring that each strategy profile is greater than the others. □

To interpret the theorem, first notice that the quantities $\alpha_1+\beta_2$ and $\alpha_2+\beta_1$ are the guaranteed payoff consequences of agent 1 and 2's decisions, respectively. In other words, in any strategy profile where agent 1 is playing $+1$, for example, the values α_1 and β_2 are present and hence picked up by social welfare function. To be precise, let us define the quantities $c_1 = \alpha_1 + \beta_2$ and $c_2 = \alpha_2 + \beta_1$ to be agent 1 and 2's guaranteed payoff contributions when they play $+1$. We can now rewrite the conditions in Theorem 3.14 using c_1 and c_2. Namely, w is maximized at $(+1, +1)$ if and only if $c_1 + c_2 > 0$ and $c_i > -2\gamma$ for $i = 1, 2$, it is maximized at $(+1, -1)$ if and only if $c_2 < -2\gamma$, $c_1 > 2\gamma$, and $c_1 > c_2$. Moreover, w is maximized at $(-1, +1)$ if and only if $c_1 < -2\gamma$, $c_2 > 2\gamma$, and $c_2 > c_1$. Lastly, w is maximized at $(-1, -1)$ if and only if $c_1 + c_2 < 0$ and $c_i < 2\gamma$ for $i = 1, 2$. In the symmetric case, $c_1 = c_2 = \alpha + \beta$.

The sum of the guaranteed contribution determines the ordering of $(+1, +1)$ and $(-1, -1)$. When positive, $w(+1, +1) > w(-1, -1)$ and when negative, the opposite. Each guaranteed contribution's relationship to -2γ determines the ordering of $(+1, +1)$ and the anti-coordination profiles $(+1, -1)$ and $(-1, +1)$. For $(-1, -1)$ and the anti-coordination profiles, it is the guaranteed contributions' relationship with 2γ that determines their ordering. In both of these cases, when the guaranteed contributions each exceed -2γ and 2γ, in their respective cases, then the welfare function favors the associated coordination profile. Finally, the guaranteed contributions relationship between each other determine the ordering of $(+1, -1)$ and $(-1, +1)$. When agent 1's contribution is greater than agent 2's, it is $(+1, -1)$ that is superior, and vice-versa.

Now, with Theorems 3.11, 3.12, 3.13, and 3.14, it is a matter of mixing-and-matching to determine the regions of coordinative agreement, coordinative tension, unilateral tension, and anti-coordinative agreement. To keep the exposition manageable, consider only the cases of coordinative agreement and coordinative tension in coordination games.

For the game to be a coordination game, it is necessary that $\gamma > |\alpha_1|, |\alpha_2|$. The risk dominant Nash equilibrium, according to Theorem 3.13, is given by the sign of the α_i such that $|\alpha_i| > |\alpha_{\neg i}|$ where $\neg i$ denotes the agent who is not i. Without loss of generality, assume that $|\alpha_1| > |\alpha_2|$. If $\alpha_1 > 0$, then the risk-dominant Nash equilibrium is $(+1, +1)$, and if $\alpha_1 < 0$, the risk-dominant Nash equilibrium is $(-1, -1)$. Consider just $\alpha_1 > 0$. In this case, it is necessary that $\alpha_1 + \beta_2 + \alpha_2 + \beta_1 > 0$ and $\beta_i + \alpha_{\neg i} > -2\gamma$ for coordinative agreement, and $\alpha_1 + \beta_2 + \alpha_2 + \beta_1 < 0$ and $\beta_i + \alpha_{\neg i} < 2\gamma$ for coordinative tension, where $i = 1, 2$.

To summarize, $(+1, +1)$ is the risk-dominant equilibrium in the coordination game because $\gamma > \alpha_1 > |\alpha_2|$. The welfare function agrees with this profile if and only if the sum of guaranteed contributions is greater than zero, and each guaranteed contribution is greater than -2γ. On the other hand, if the sum of guaranteed contributions is negative, and each guaranteed contribution is less than 2γ, then the game has coordinative tension.

3.5 Potential and Welfare on Networks

The complexity of potential games on networks, because of the emergent structures identified in chapter 2 and the large number of individual preferences and externalities, makes a full understanding of network potential games for large networks a big challenge. In chapter 4, some analysis is offered for 3-agent networks. General analysis of asymmetric potential games with emergent structures is beyond the scope of this thesis, and is saved for future work. A useful, though restrictive, way around this is to sum symmetric 2×2 games between all pairs of neighbors. This is done in [18] [19] [12], and as a starting point, this thesis follows the same path. An effect of this is that all higher order emergent structures are ignored. Hence, an entire n-agent game on a network can be described by the three parameters α, γ, and β, and network structure.

We borrow most of the network notation used by Newton *et al.* Suppose $\Gamma = (V, E)$ is a simple finite connected graph with set of vertices, or players, V, and set of edges E. The elements of E are the ordered pairs (i,j) where i and j are players in V and the existence of $(i,j) \in E$ means that agent i is connected to agent j in Γ. Moreover, assume that $i < j$ for every $(i,j) \in E$, so that if $(i,j) \in E$, then $(j,i) \notin E$. This causes the model to lose no generality and instead makes writing summations simpler. When $(i,j) \in E$, i and j are said to be *neighbors*, and the *degree* of an agent i is the sum of i's neighbors. For a subset $S \subseteq V$, the sum of degrees in S is denoted by $d(S)$. For two subsets $S, T \subseteq V$, the sum of edges from S to T is denoted by $d(S,T)$. Continuing in the fashion of Newton *et al.*, instead of referring to the degree of i by $d(\{i\})$ and the number of neighbors of i in S by $d(\{i\}, S)$, we use $d(i)$ and $d(i,S)$ respectively.

Each agent i plays the game $\mathcal{G}$ with all of their neighbors, where $\mathcal{G}$ is shown decomposed into the coordinate system in Table 3.7.

	$+1$	-1
$+1$	$\alpha \quad \alpha$	$\alpha \quad -\alpha$
-1	$-\alpha \quad \alpha$	$-\alpha \quad -\alpha$

Individual Preference

	$+1$	-1
$+1$	$\gamma \quad \gamma$	$-\gamma \quad -\gamma$
-1	$-\gamma \quad -\gamma$	$\gamma \quad \gamma$

Coordinative Pressure

	$+1$	-1
$+1$	$\beta \quad \beta$	$-\beta \quad \beta$
-1	$\beta \quad -\beta$	$-\beta \quad -\beta$

Externalities

Table 3.7: Coordinate System

Without loss of generality, write the payoff function for agent 1 understanding that the choice of agent 1 is arbitrary and that this form of the payoff function holds for any agent $i \in V$. The payoff function for agent 1 is

$$\pi_1(t_1, \ldots, t_n) = \alpha_1 t_1 + \gamma \sum_{(1,j)\in E} t_i t_j + \sum_{(1,j)\in E} \beta_{ij} t_j \tag{3.6}$$

Notice that the agent only receives α once. The γ term is summed across all neighbors.

The potential function for this game is given by

$$P(t_1, \dots, t_n) = \alpha \sum_{i \in V} t_i + \gamma \sum_{(i,j) \in E} t_i t_j \tag{3.7}$$

Let us verify that this is indeed a potential function for $\mathcal{G}$. The payoff for agent 1 playing $+1$ is given by $\pi_1(+1, \dots, t_n) = \alpha + \gamma \sum_{(1,j) \in E} t_j + \beta \sum_{(1,j) \in E} t_j$, while their payoff for playing the strategy -1 is given by $\pi_1(-1, \dots, t_n) = -\alpha - \gamma \sum_{(1,j) \in E} t_j + \beta \sum_{(1,j) \in E} t_j$. The difference is then,

$$\pi_1(+1, \dots, t_n) - \pi_1(-1, \dots, t_n) = 2\alpha + 2\gamma \sum_{(1,j) \in E} t_j$$

The potential function, when agent 1 plays $+1$, takes the value $P(+1, \dots, t_n) = \alpha + \alpha \sum_{i \in V, i \neq 1} t_i + \gamma \sum_{(1,j) \in E} t_j + \gamma \sum_{(i,j) \in E, i \neq 1} t_i t_j$, and when they play -1 the potential function has the value $P(-1, \dots, t_n) = -\alpha + \alpha \sum_{i \in V, i \neq 1} t_i - \gamma \sum_{(1,j) \in E} t_j + \gamma \sum_{(i,j) \in E, i \neq 1} t_i t_j$. The difference is then,

$$P(+1, \dots, t_n) - P(-1, \dots, t_n) = 2\alpha + 2\gamma \sum_{(1,j) \in E} t_j$$

Hence $\pi_1(+1, \dots, t_n) - \pi_1(-1, \dots, t_n) = P(+1, \dots, t_n) - P(-1, \dots, t_n)$. Because the choice of agent 1 was arbitrary this holds for all agents, and the potential function in (3.7) is a potential function for $\mathcal{G}$.

The social welfare function for this game, as throughout this chapter, calculates the sum of payoffs in any strategy profile. Given a state of the game, all agents receive either α or $-\alpha$ depending on their strategy being $+1$ or -1. These payoffs are earned by each agent independent of the network structure or the strategy played by the others. Hence, in the global social welfare function we must have $\alpha \sum_{i \in V}$. In the given state of the game, in any of the sub-2 $\times$ 2 interactions, the two agents are either each receiving γ or each receiving $-\gamma$. Hence, in the welfare function, these can be taken into account with $2\gamma \sum_{(i,j)\in E} t_i t_j$. Finally, we must figure out how the β terms can be counted by the welfare function. Take an edge $(i, j) \in E$. In this sub-2 $\times$ 2 interaction, agent i is receiving the externality βt_j, and the agent j is receiving the externality βt_i. In this interaction, then, these components can be added to give $\beta(t_i + t_j)$. Since this is happening with every edge, the global social welfare function must include $\beta \sum_{(i,j)\in E}(t_i + t_j)$.

Therefore, the social welfare function of this game is

$$w(t_1, \ldots, t_n) = \alpha \sum_{i\in V} t_i + 2\gamma \sum_{(i,j)\in E} t_i t_j + \beta \sum_{(i,j)\in E} (t_i + t_j) \tag{3.8}$$

The states in this game can be categorized as "all +1," "all −1," or a mix of +1's and −1's. In the state "all +1," which we denote by $+\mathbf{1}$, $t_i = +1$ for every $i \in V$. Similarly, the state "all −1," denoted by $-\mathbf{1}$, means that $t_i = -1$ for every $i \in V$. The state with both +1's and −1's there is at least one agent playing +1 and at least one agent playing −1, the specific number of each is denoted by n_{+1} and n_{-1}, respectively. We refer to this state as $(+\mathbf{1}, -\mathbf{1})$.

This gives:

$$P(+\mathbf{1}) = n\alpha + |E|\gamma$$

$$w(+\mathbf{1}) = n\alpha + 2|E|\gamma + 2|E|\beta$$

$$P(-\mathbf{1}) = -n\alpha + |E|\gamma$$

$$w(-\mathbf{1}) = -n\alpha + 2|E|\gamma - 2|E|\beta$$

On the other hand,

$$
\begin{aligned}
P(+\mathbf{1},-\mathbf{1}) &= (n^{\mathbf{t}}_{+1} - n^{\mathbf{t}}_{-1})\alpha + (|E^{\mathbf{t}}_{+1,+1}| + |E^{\mathbf{t}}_{-1,-1}| - |E^{\mathbf{t}}_{+1,-1}|)\gamma \\
w(+\mathbf{1},-\mathbf{1}) &= (n^{\mathbf{t}}_{+1} - n^{\mathbf{t}}_{-1})\alpha + 2(|E^{\mathbf{t}}_{+1,+1}| + |E^{\mathbf{t}}_{-1,-1}| - |E^{\mathbf{t}}_{+1,-1}|)\gamma \\
&\qquad + 2(|E_{+1,+1}| - |E_{-1,-1}|)\beta
\end{aligned}
$$

Where $E^{\mathbf{t}}_{t',t''} \subseteq E$ denotes the subset of the edge set E containing all edges such that the first agent is playing t' and the second t'' in the strategy profile $\mathbf{t}$. In the calculations of this section it is assumed that $\mathbf{t}$ stands for strategy profiles of the type $(+\mathbf{1},-\mathbf{1})$, where we write $\mathbf{t}$ as the superscript instead of $(+\mathbf{1},-\mathbf{1})$ to reduce clutter. This superscript also shows up in n_{+1} and n_{-1} since these depend on the number of agents playing each strategy, hence strategy profile. Notice that $|E^{\mathbf{t}}_{+1,-1}| = |E^{\mathbf{t}}_{-1,+1}|$.

After calculations, the following regions in the parameter space are found that give rise to inequalities in the potential and welfare functions.

$$P(+\mathbf{1}) > P(-\mathbf{1}) \iff \alpha > 0 \tag{3.9}$$

$$P(+\mathbf{1}) > P(+\mathbf{1},-\mathbf{1}) \iff n^{\mathbf{t}}_{-1}\alpha + |E^{\mathbf{t}}_{+1,-1}|\gamma > 0 \tag{3.10}$$

$$P(-\mathbf{1}) > P(+\mathbf{1},-\mathbf{1}) \iff -n^{\mathbf{t}}_{+1}\alpha + |E^{\mathbf{t}}_{+1,-1}|\gamma > 0 \tag{3.11}$$

$$w(+\mathbf{1}) > w(-\mathbf{1}) \iff n\alpha + 2|E|\beta > 0 \tag{3.12}$$

$$w(+\mathbf{1}) > w(+\mathbf{1},-\mathbf{1}) \iff n^{\mathbf{t}}_{-1}\alpha + 2|E^{\mathbf{t}}_{+1,-1}|\gamma + (|E^{\mathbf{t}}_{+1,-1}| + 2|E^{\mathbf{t}}_{-1,-1}|)\beta > 0 \tag{3.13}$$

$$w(-\mathbf{1}) > w(+\mathbf{1},-\mathbf{1}) \iff -n^{\mathbf{t}}_{+1}\alpha + 2|E^{\mathbf{t}}_{+1,-1}|\gamma - (2|E^{\mathbf{t}}_{+1,+1}| + |E^{\mathbf{t}}_{+1,-1}|)\beta > 0 \tag{3.14}$$

These inequalities look very similar to the ones in Section 3.2.1, where 2×2 games were analyzed. One can verify that these inequalities reduce to the 2×2 case.

For a more manageable exposition, we only identify the cases of coordinative agreement and disagreement for coordination games. The assumption that the underlying game is a coordination game translates to $|\gamma| > \alpha$. Then, if $\alpha > 0$, it is immediate that (3.9) and (3.10) are satisfied, and that $+\mathbf{1}$ globally maximizes the potential function.

There is coordinative agreement when the welfare function is also globally maximized at $+\mathbf{1}$. This happens precisely when $n\alpha+2|E|\beta > 0$ and $n^{\mathbf{t}}_{-1}\alpha+2|E^{\mathbf{t}}_{+1,-1}|\gamma+(|E^{\mathbf{t}}_{+1,-1}|+2|E^{\mathbf{t}}_{-1,-1}|)\beta > 0$. On the other hand, there is coordinative tension then the welfare function is globally maximized at $-\mathbf{1}$. The requirements here are that $n\alpha+2|E|\beta < 0$ and $-n^{\mathbf{t}}_{+1}\alpha+2|E^{\mathbf{t}}_{+1,-1}|\gamma-(2|E^{\mathbf{t}}_{+1,+1}|+|E^{\mathbf{t}}_{+1,-1}|)\beta > 0$. The intricate relationship of these regions for the welfare function with the set of edges will be left unexplored.

3.5.1 Young's Example

In *The Diffusion of Innovations in Social Networks* [18], Young defines a game where the payoffs are separated into individual and social components. The individual component involves what he defines to be the agents' idiosyncratic preferences for strategies A and B. This preference gives the agents a payoff for choosing A or B that is independent of the other agents' strategies. The social component of the payoff is said to be the network externalities that are generated in the strategy profiles of the game, represented using a 2×2 game matrix. In other words, this payoff for each agent is dependent on all agents' strategies. These externalities, says Young, may arise from a variety of factors including demonstration effects, increasing returns, or the desire to conform. The agents play on a network, and their utility comes from their idiosyncratic payoff plus the sum of the result of their 2×2 game with their neighbors. Each agent receives their idiosyncratic payoff once, while the externality payoff is cumulative over the agent's neighbors. Let us make this precise.

Borrowing the graph theoretic network notation used by Young, suppose $\Gamma = (V, E, W)$ is a simple finite connected weighted directed graph with set of vertices, or players, V, set of edges, E, and set of weights W. The elements of E are ordered pairs (i, j) where i and j are players in V and the existence of $(i, j) \in E$ means that agent i is influenced by agent j. Each edge $(i, j) \in E$ has a weight $w_{ij} \in W$ that represents the strength of agent j's influence over agent i. To reduce complexity, Young assumes that $w_{ij} = w_{ji}$ for all i, j such that $(i, j) \in E$, *i.e.,* the influence is symmetric for all pairs of connected agents.

Each agent can play one of two strategies, A or B, which we refer to as $+1$ and -1 to maintain the convention established in this dissertation. The individual component of agent i's payoff is the utility of agent i's idiosyncratic preferences for $+1$ and -1, denoted by $v_i(t_i)$, for each $i \in V$. It is clear from the domain of v_i that the idiosyncratic preference for each i is independent of t_j for all $j \in V, j \neq i$. Young defines the social component of agent i's payoff to be $\sum w_{ij} u(t_i, t_j)$ for all j such that $(i, j) \in E$. Here $u(t_i, t_j)$ represents the utility agent i gains from the 2×2 network externality game played with j. Again, in this 2×2 game, the payoffs are an aggregate of network externalities like increasing returns and the desire to conform. The total payoff for agent i is then written as

$$U_i(\mathbf{t}) = \sum_{(i,j) \in E} w_{ij} u(t_i, t_j) + v_i(t_i) \tag{3.15}$$

In the same paper [18], Young gives an example of a game where the strategy profile that globally maximizes the potential function is not the same as the profile that globally maximizes social welfare. The story is of two competing technologies, one of which is easy to use while the other offers better networking capabilities. Young assumes that the payoffs have an individual and a social component, hinting at the structure of potential games detailed. The first technology's payoff is assumed to have a high individual component but low social component. The payoff for the latter technology, on the other hand, is assumed to have no individual component but a high social component. These strategies are denoted by $+1$ and -1, respectively. Young specifies these payoffs to be as in Table 3.8.

		$+1$		-1	
$+1$	1	1	0	0	$v(+1) = 8$
-1	0	0	4	4	$v(-1) = 0$

Table 3.8: Young's Example of Disagreement

We remind the reader that this utility quantifies the agents' idiosyncratic preference for strategies $+1$ and -1. The network externalities of the game are given in the game matrix of Table 3.8. Using $n_{+1}(\mathbf{t})$ to denote the number of agents in state $\mathbf{t}$ that play strategy $+1$, Young writes the potential function $P(\mathbf{t})$ and the social welfare function $w(\mathbf{t})$ as

$$P(\mathbf{t}) = w_{+1,+1}(\mathbf{t}) + 4w_{-1,-1}(\mathbf{t}) + 8n_{+1}(\mathbf{t}) \tag{3.16}$$

$$w(\mathbf{t}) = 2w_{+1,+1}(\mathbf{t}) + 8w_{-1,-1}(\mathbf{t}) + 8n_{+1}(\mathbf{t}) \tag{3.17}$$

Young states in [18] that it is clear that the state where all agents play -1 maximizes social welfare, and the state where all agents play $+1$ maximizes the potential function. Denote these states by $+\mathbf{1}$ and $-\mathbf{1}$, respectively. To demonstrate, take the total sum of edge weights in the social network to be n_w. Then, in $+\mathbf{1}$, $w_{+1,+1}(+\mathbf{1}) = n_w$, $w_{-1,-1}(+\mathbf{1}) = 0$, and $n_{+1}(+\mathbf{1}) = n$, where n is the total number of agents. Then, $P(+\mathbf{1}) = n_w + 8n$ and $w(+\mathbf{1}) = 2n_w + 8n$. On the other hand, in $-\mathbf{1}$, $w_{+1,+1}(-\mathbf{1}) = 0$, $w_{-1,-1}(-\mathbf{1}) = n_w$, and $n_{+1}(-\mathbf{1}) = 0$. Here, $P(\textbf{-1}) = 4n_w$ and $w(\textbf{-1}) = 8n_w$.

In $+\mathbf{1}$, Young claims potential is maximized but not welfare, while in $-\mathbf{1}$, welfare is maximized but not potential. This means it must be the case that $n_w + 8n > 4n_w$ and $8n_w > 2n_w + 8n$. These inequalities reduce to $n > \frac{3}{8}n_w$ and $n_w > \frac{4}{3}n$. Putting these together then gives $\frac{8}{3}n > n_w > \frac{4}{3}n$, an implicit assumption in Young's example.

This example is used by Young to demonstrate a situation in which the dynamic used in his model leads, with high probability, to a state where most agents do not adopt the favorable technology. The example, however, is ad-hoc and it is not shown precisely why these payoffs give rise to disagreement between the potential function and the social welfare function. Moreover, decomposing the game in Table 3.8 into the coordinate system reveals there is some additional α information in the part of the game that is supposed to represent purely the social component of the payoff. In light of this and the implicit assumption highlighted in the previous paragraph, it is clear that there is some obscurity in this example. The essence of it, however, is undeniably the tension between the potential function and the social welfare function brought about by the behavioral component that we have explored in this chapter.

Focusing on the underlying 2×2 game, we know that Young's example is simply of a potential game that lives in the region of coordinative tension. Putting this 2×2 game on networks simply means that the game parameters (α, γ, and β) are scaled by network parameters and variables (such as number of edges and degrees).

3.6 Autonomy

3.6.1 Potential Autonomy

In [19] Young discusses the dynamics of social innovation. The core of the model is a coordination game played on networks, where one equilibrium represents the status quo and the other, the better equilibrium, represents an innovation that increases the welfare of those who adopt it. Young parametrizes with α the benefit of the innovation over the status quo. This produces the game matrix shown in Table 3.9, where the strategy -1 denotes the status quo, and $+1$ denotes and the innovation. In this paper, the same strategy profile maximizes both the welfare function and the potential function. Referring to Table 3.9, when $\alpha > 0$ this strategy profile is $(+1, +1)$.

	+1		-1	
+1	$1+\alpha$	$1+\alpha$	0	0
-1	0	0	1	1

Table 3.9: Young's Social Innovation Game

In this model, Young uses the potential function to characterize the network properties under which a subset of agents has the collective incentive to deviate to the innovation. This collective incentive is interpreted to be collective deviations that increase the potential function. Subsets of agents whose collective deviation increases the potential function are said to be *potential autonomous*. Young relates *potential autonomous* subsets to the graph-theoretic notion of *close-knittedness*, which measures how well integrated each subset of the group is with the rest of the group. Formally, $CK(S) = \min_{S' \subseteq S} d(S, S')/d(S')$. Here $d(S', S)$ is the number of edges between S' and S, and $d(S')$ is the sum of degrees in S'.

We state one of Young's main theorems without proof and refer the reader to [18].

Theorem 3.15. *A subset of agents S is potential autonomous if and only if*

$$CK(S) > \frac{1}{2+\alpha}. \tag{3.18}$$

3.6.2 Agency Autonomy

Newton *et al.* explore the same situation as Young but change his model to include a parameter β^* that is not picked up by the potential function. In addition, Newton *et al.* use the parameter α but in a different way from Young. We show that β^*, as used in this model, is a parameter coming solely from the game's behavioral component[5]– to which the potential function is blind– and this is why, following from discussions in chapter 2, it is not

[5]It is also present in the kernel but we pay not attention to it, as described in chapter 2.

picked up by the potential function. Note that we use β^* to disambiguate between the β from the coordinate system.

The game used by Newton *et al.* is given in Table 3.10.

	+1		-1	
+1	$1+\beta^*$	$1+\beta^*$	0	$\beta^*-\alpha$
-1	$\beta^*-\alpha$	0	1	1

Table 3.10: Newton *et al.* 's Model

Decomposing this game shows us how the β^* parameter is only present in the behavioral component[6]. This is shown in Tables 3.11 and 3.12.

	+1		-1	
+1	$\frac{1+\alpha}{2}$	$\frac{1+\alpha}{2}$	$-\frac{1}{2}$	$-\frac{1+\alpha}{2}$
-1	$-\frac{1+\alpha}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Nash Component

Table 3.11: Nash Component of Newton *et al.* 's Model

	+1		-1	
+1	$\frac{2\beta^*-\alpha}{4}$	$\frac{2\beta^*-\alpha}{4}$	$-\frac{2\beta^*-\alpha}{4}$	$\frac{2\beta^*-\alpha}{4}$
-1	$\frac{2\beta^*-\alpha}{4}$	$-\frac{2\beta^*-\alpha}{4}$	$-\frac{2\beta^*-\alpha}{4}$	$-\frac{2\beta^*-\alpha}{4}$

Behavioral Component

Table 3.12: Behavioral Component of Newton *et al.* 's Model

From this, we already know that β^* will not be picked up by the potential function since it is not found in the Nash component of the game. Newton *et. al.*, show that the game they

[6]We mention one more that it is also present in the kernel.

use, in Table 3.10, has the same potential function as the game used by Young, in Table 3.9. Because of this, the same characterization of potential autonomy holds. That is, a subset S of agents is potential autonomous if and only if $CK(S) = 1/(1+\alpha)$.

Newton *et. al.*'s focus is on characterizing network properties wherein deviating subsets of agents all receive a higher payoff from the deviation. Such a subset is defined to be *agency autonomous*. It is shown that agency autonomous sets are related to the graph-theoretic notion of cohesion. Cohesion and close-knittedness are similar. Close-knittedness measures how well integrated each subset of the group is with the rest of the group. Cohesion, on the other hand, measures how well integrated each agent is to the rest of the group. In other words, a cohesive group is understood to be one where no agent has too many of their interactions with outsiders. Formally, $Co(S) = \min_{i \in S} d(i, S)/d(i)$. The main theorem is the following, which we state without proof.

Theorem 3.16. *A subset S of agents is agency autonomous if and only if*

$$Co(S) > \frac{1}{1+\beta^*}. \tag{3.19}$$

3.6.3 An Alternative Parametrization

Newton *et al.* introduce the parameters α and β but gives no intuitive justification for the role it is supposed to play in the game. It is simply used as part of a two dimensional parametrization of all symmetric 2×2 coordination games up to affine transformations. In addition, it has the same potential function as the 1-dimensional game originally used by Young, a convenient property. We offer a more intuitive parametrization based on the coordinate system developed in this thesis. To keep things symmetric, we take $\alpha_1 = \alpha_2$ and call it α. Furthermore, we use β to denote the positive externality generated by agents playing $+1$. We hope that our use of α and β is not confused with Newton *et al.*'s.

This new parametrization allows us to express the game as in Table 3.13.

	+1		-1	
+1	$\gamma+\alpha+\beta$	$\gamma+\alpha+\beta$	$-\gamma+\alpha-\beta$	$-\gamma-\alpha+\beta$
-1	$-\gamma-\alpha+\beta$	$-\gamma+\alpha-\beta$	$\gamma-\alpha-\beta$	$\gamma-\alpha-\beta$

Table 3.13: Parametrization of 2×2 Coordination Games

If it is wished to further reduce the dimensionality of this parametrization there are several options; among them, one can set γ to a constant value, or one can aggregate, in a meaningful way, the Nash parameters α and γ.

Newton *et al.* juxtapose Young's concept of potential autonomous with agency autonomy, a concept they introduce. Young defines a set of agents to be potential autonomous if the potential function increases from their collective deviation. Newton *et al.* define a set of agents to be agency autonomous if the payoff of every player in the set is increased under a collective deviation. The collective deviation of interest, since +1 is interpreted to be the innovation and -1 the status quo, is from $(-1,-1)$ to $(+1,+1)$.

Using the parametrization offered in Table 3.13, we calculate the difference in payoff from a collective deviation from $(-1,-1)$ to $(+1,+1)$ in a 2-agent case. This deviation gives the agents the payoff difference of $(\gamma+\alpha+\beta)-(\gamma-\alpha-\beta)=2\alpha+2\beta$. This difference is positive when $\alpha+\beta>0$. On a network, this will simply be scaled by the appropriate values of the degrees and edge sets.

With the new parametrization, where the parameters are given meaning, we see that in order for the collective deviation to yield a positive increase in payoff, the agent needs that the sum of their individual preference for the innovation plus the externality received from others playing the innovation to be positive. This allows many interesting situations.

Perhaps the agents prefer the status quo, in which case $\alpha<0$, but the collective deviation will still be profitable as long as the externality outweighs the cost of playing something the agent does not inherently prefer. On the other hand, perhaps their individual preference for the innovation outweighs a negative externality coming from others adopting it.

Ideally, the agent both prefers and receives a positive externality from the adoption of the innovation. This is the case in Newton *et al.*'s model, and they go to lengths to argue this with their introduction of β^*. We would like to point out that, with our parametrization coming from the decomposition, this is immediate and clear.

Finally, we add that the way in which both Young and Newton *et al.* extend the 2×2 game to networks is by defining the agents' payoff to be the sum of all 2×2 interactions with their neighbors. Hence, for example, an agent playing $+1$ with n agents playing $+1$ and m agents playing -1 will receive a payoff of $(\gamma+\alpha+\beta)n+(-\gamma+\alpha-\beta)m$. Similarly, an agent playing -1 with the same neighbors receives a payoff of $(-\gamma-\alpha+\beta)n+(\gamma-\alpha-\beta)m$.[7]

Now, suppose all agents in Γ are playing strategy -1. What does it mean, then, for a set of agents to be agency autonomous, as defined by Newton *et al.*, using the parametrization offered in Table 3.13?

Consider a subset $S \subseteq V$. For any agent $i \in S$, their payoff at the beginning is simply $(\gamma-\alpha-\beta)d(i,V)$, which can be expressed as $(\gamma-\alpha-\beta)(d(i,S)+d(i,V\backslash S))$. If every agent $i \in S$ deviates from strategy -1 to strategy $+1$, the payoff for every agent $i \in S$ becomes $(\gamma+\alpha+\beta)d(i,S)+(-\gamma+\alpha-\beta)d(i,V\backslash S)$. This gives a positive difference when

$$(\gamma+\alpha+\beta)d(i,S)+(-\gamma+\alpha-\beta)d(i,V\backslash S) > (\gamma-\alpha-\beta)(d(i,S)+d(i,V\backslash S))$$

[7]This breaks away from our original interpretation of potential games on networks, where the α term is received by each agent only once. Here, the α term is summed in every 2×2 interaction. The consequences of this are minor, and will be made precise in future work.

Which reduces to,

$$(\alpha+\beta)d(i,S)+(-\gamma+\alpha)d(i,V\backslash S)>0$$

Here we cleverly add and subtract $(-\gamma+\alpha)d(i,S)$ to the left hand side of the inequality, giving, and further simplifying into,

$$\begin{aligned}
&(\alpha+\beta)d(i,S)+(-\gamma+\alpha)d(i,V\backslash S)+(-\gamma+\alpha)d(i,S)-(-\gamma+\alpha)d(i,S)>0\\
&(\alpha+\beta+\gamma-\alpha)d(i,S)+(-\gamma+\alpha)(d(i,V\backslash S)+d(i,S))>0\\
&(\beta+\gamma)d(i,S)+(-\gamma+\alpha)(d(i,V\backslash S)+d(i,S))>0\\
&(\beta+\gamma)d(i,S)+(-\gamma+\alpha)d(i)>0
\end{aligned}$$

Rearranging and dividing by $(\beta+\gamma)$ gives

$$\frac{d(i,S)}{d(i)}>\frac{-\alpha+\gamma}{\beta+\gamma}. \tag{3.20}$$

We make a note that the externality for the innovation is assumed to be positive $(\beta>0)$, and that the common interest coordinative pressure is also assumed to be positive $(\gamma>0)$. Because of this, $\beta+\gamma>0$ and we can divide without worrying about changing the inequality. For alternative cases, the appropriate modifications can be easily made.

Using their parametrization, Newton *et al.* find that a set of agents $S\subseteq V$ is agency autonomous if and only if $Co(S)>\frac{1}{1+\beta^*}$. Taking the minimum over the inequality (5.1) we, with our parametrization, can replace their inequality with $Co(S)>\frac{-\alpha+\gamma}{\beta+\gamma}$. In other words, we replace Theorem 3.16 with Theorem 3.17 below.

Theorem 3.17. *A subset S is agency autonomous if and only if*

$$Co(S) > \frac{-\alpha + \gamma}{\beta + \gamma} = 1 - \frac{\alpha + \beta}{\beta + \gamma}. \tag{3.21}$$

Going over similar calculations for the potential function and the notion of potential autonomy allows us to rewrite Theorem 3.15 with Theorem 3.18 below.

Theorem 3.18. *A subset S is potential autonomous if and only if*

$$Co(S) > \frac{-\alpha + \gamma}{2\gamma} = \frac{1}{2} - \frac{\alpha}{2\gamma}. \tag{3.22}$$

Here we are immediately afforded a more intuitive understanding of this network property. We have the parameters for individual preference α, coordinative pressure γ, and externalities β. Because this is a coordination game with Nash equilibria at strategy profiles $(+1, +1)$ and $(-1, -1)$, we have that $\gamma > |\alpha|$. This means that the fraction is always positive. Hence, some level of cohesion is always necessary. We also see that with increasing β, less cohesion is necessary. There are only some examples of the possible analysis using the decomposition as the parametrization of coordination games.

On the other hand we see that the requirement for potential autonomy is independent of β, something we continue to stress. There, the close-knittedness is shown to be dependent only on α and γ. The greater the α term, the agents individual preference, the less restrictive the notion of close-knittedness needs to be.

It makes sense that if the agents are acting unilaterally, asking the question "what should *I* do?", that the result is dependent largely on the agents' individual preference α. On the other hand, if the agents are able to make decisions together, asking questions of the sort "what should *we* do?", then the result is dependent on all factors of the game– the individual preference α, the pressure to conform γ, and the externality β.

3.7 Conclusion

We summarize the main conclusions from this chapter in the list below.

- The potential function does not pick the behavioral component of the game.
- The behavioral component of the game is the culprit behind possible tension between the risk-dominant Nash equilibrium and social welfare.
- Newton *et. al.* indirectly probed into the structure of the behavioral component and showed its invariance to potential autonomy.
- As we have seen throughout this chapter, a large amount of what happens in the game is determined by the behavioral component. The potential function, Nash equilibria, and many learning dynamics ignore the behavioral terms.
- This subtlety was first identified in [8]. In this chapter we applied it to clarify the literature on the diffusion of innovation.

We put Young and Newton's model on the coordinate system for symmetric games. It is entirely reasonable however, that the agents have different preferences over the two strategies, and that each agent may cause a particular externality from using each strategy. Breaking the α and β symmetry in this model is important and is saved for future work. Breaking the γ symmetry makes way for more realism, too. For example, a certain pair of agents want to conform, say i and j, so that $\gamma_{ij} > 0$, but another pair, say i and k, do not, so that $\gamma_{ik} < 0$. Moreover, allowing higher order Nash and behavioral terms in the network model allows for more realistic situations. Furthermore, allowing higher order Nash and behavioral terms in the network model allows for more realistic situations. This would give the model the flavor of the legislator game, discussed in chapter 2, where the Nash equilibria are not pure coordination profiles, but rather, there are thresholds where a certain number n_1 of agents playing $+1$ and n_2 playing -1 is better than all agents playing $+1$.

Chapter 4

Collaboration on Networks

4.1 Introduction

We remind the reader of the motivation behind this thesis– the pursuit of a model of group creativity and collaboration. The first step in this was a set of collective improvisation experiments set up primarily as a source for intuition. To build a model, because of their association with common interest and coordination, we focused on potential games and exposed their structure using algebraic tools in chapter 2. In chapter 3, we used the coordinate system developed in chapter 2 to offer a general framework for current research in the diffusion of innovation. Now, in this chapter, we discuss the experiments briefly, and then go over several models we contribute.

Models of scientific collaboration and network epistemology are a recent and fruitful area of study [21] [22] [20]. To capture scientific collaboration, these models focus on a pursuit of an underlying "truth" by the scientific agents on the network. The flow of information represents the communication of results and theories. Typically, the networks analyzed come from the work of Bala and Goyal [1]. They include the complete network, the wheel network, and the star network. These will be the networks of focus in this chapter.

In our model, we view the agents as engaging in a creative process, where there is no "right" answer and hence no "truth." Instead, we assume the agents have to, collectively, solve many problems faced by a collaborative group– including, among others, coordination and division of labor– which we model using appropriate Nash equilibrium structures. In solving these problems, the group is able to move forward in the production of their creative work.

To begin the chapter, we describe the collective improvisation experiments. We save formal and detailed analysis of the experiments for a later paper where appropriate tools, that are under development, can be used. Then, using the structure of potential games developed in chapter 2, we offer a simple model of an improvising trio of musicians on the full network, the wheel network, and the star network. After this, we offer a model to explain the importance of a conductor in coordinating an orchestra, which fits into the theme of the chapter because of its focus on networks and coordination (and music!). Finally, we offer an alternative point of view into the structure of games– where the payoffs are interpreted as information generated and transmitted.

4.2 Collective Improvisation Experiments

The collective improvisation experiments were conducted during the summer of 2015 with four graduate students from UC Irvine's program in Integrated Composition, Improvisation, and Technology. The four musicians, at the time of the experiments, had experience improvising together for two years. There were two guitarists, a saxophonist, and a violinist.

The four participating musicians were isolated in separate spaces that were acoustically treated in order to prevent sound from traveling between the spaces. Each musician was given a pair of headphones through which they could hear themselves and whomever else we allowed them to hear. All musical output was first run through a studio system where it was controlled to which headphones the different audio streams went. The networks of information flow were enforced through this mechanism.

In order to prevent the music from purposely becoming too abstract,[1] we asked the improvisers to play in such a way that a general audience would appreciate the music produced, but to also not restrict themselves to basic music ideas.[2] They knew that the listening would be controlled, but they were not informed of the details of the networks underlying each session during the experiments. After all sessions had been recorded, the musicians were shown the networks. In addition, they were aware that all musicians were participating in all of the pieces. In other words, the musicians knew that if they could not hear someone, it did not mean that this person was not participating in the piece.

We make an important note that a preliminary model was not established before the experiments, and precise predictions were not formulated. This is because of the fleeting opportunity to run the experiments.[3] The collective improvisation experiments were set up, then, mostly to gain intuition on how collective phenomena pertaining to collaboration and creativity manifested under the constraints of the experiment.

We would like to acknowledge, however, the rich source of intuition these experiments proved to be in the development of the model, and look forward to continuing this work both experimentally and theoretically. This initial round of experiments and all consequent analysis has shown us, for example, how to better run such experiments in the future. In addition, a complete and thorough musical analysis has not yet been performed on the music generated from the experiment. Techniques are being developed for this purpose.

There are plans to create an interactive website where visitors can hear the pieces generated

[1] In the style of Ornette Coleman's Free Jazz: A Collective Improvisation.

[2] The music produced was still far from "pop" music.

[3] The experiments were ran in a single day, over roughly 9 hours including set-up and tear-down. Successfully finding a time despite busy schedules and having willing musicians around was not taken for granted.

in each network, with the ability to mute certain parts. This is to allow the listener to examine the music being created by two agents who are not directly connected to each other on the network.[4]

This dissertation is only scratching the surface of the information structure and dynamics in collective improvisations.

4.2.1 The Nine Sessions

A total of nine improvisational sessions, each around 7 to 10 minutes in length, were recorded with varying underlying networks of information flow. We refer to the first guitar player with g_1, the saxophonist with s, the violinist with v, and the second guitarist with g_2.

Session 1 represents the scientific control experiment, where all musicians can hear each other. This is a typical collective free improvisation. All other sessions involved changing the underlying network of information flow, the independent variable of the experiment. The motivation behind each session is briefly discussed, but this dissertation focuses on Sessions 1, 3, 4, 6, and 8. Session 1 manifests the complete network and, as previously mentioned, is the standard situation in collective improvisation. Sessions 3 and 8 have the wheel network as their underlying listening structure, and sessions 4 and 6 have listening conforming to the star network. The complete, wheel, and star networks are commonly used in modeling the dynamics of information flow [21] [22] [20], and their use originates from work by Bala and Goyal [1].

The empty network was excluded from the experiments in an effort to save time because the result, we assumed, would have been complete chaos with any ephemeral coordination having emerged solely by chance.

[4] Stay tuned!

Session 1

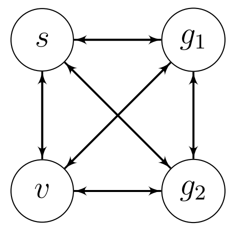


Figure 4.1: Session 1

Session 1, as described earlier, was the scientific control experiment. In this session everybody could hear everybody. The result of this session was unexpected[5] but simultaneously as expected.

Session 2

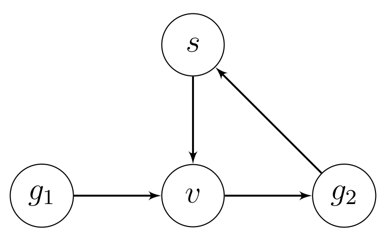


Figure 4.2: Session 2

In session 2, v, g_2, and s were in a wheel network where v, in addition, was able to hear g_1. On the other hand, g_1 could hear no one. The motivation here was to observe v's behavior, who was simultaneously receiving two possibly uncorrelated streams of information. One stream came from g_1, who received no feedback whatsoever, and another stream came from s, to whom v indirectly sent information.

Although v integrated information received from g_1 and sent it to g_2, by the time this came back to v through s, because of g_1 receiving no feedback, it is hard to say that it matched the music's evolutionary trajectory manifested through g_1. There are additional phenomena

[5]Thus is the nature of improvisation.

observed, especially in terms of the wheel sub-network in this session. However, a discussion of these observations is reserved to the wheel network sessions, Sessions 3 and 8.

Session 3

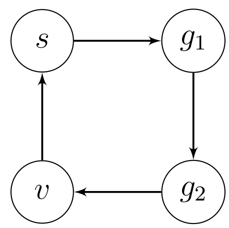


Figure 4.3: Session 3

Session 3 was a wheel network. Here g_1 received information from s, who received information from v, who received information from g_2, who finally received information from g_1. During the beginning of this session it is clear that the musicians experienced confusion. Usually there is immediate feedback in an improvisation as the musicians put forth their musical ideas. This immediate feedback adds to the processes involved that eventually lead to global levels of coordination in improvisation. In this network, however, because there is no bidirectional information flow, and hence no direct feedback, this essential component of musical coordination in improvisation was absent. Instead, in this session, musical ideas expressed by a musician only sometimes made their way back, and when they did, it was always with delay.

Regardless, after some time, the musicians seemed to integrate the odd nature of information flow in this session, and successfully reached a certain degree of coordination. The coordination observed by the end of this piece is different from the one observed in Session 1, the complete network. This reflects the multidimensionality of musical information, and how some of the information gets lost when there is not full communication between all participating improvisers– preventing full coordination.

Session 4

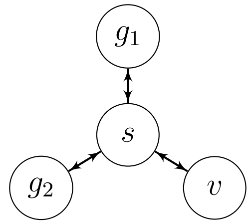


Figure 4.4: Session 4

In session 4 the agents were playing on a star network. In this network the central node s was connected to everyone. They sent information to all other agents and received information from all of them. The other agents g_1, v, and g_2 all received information from s and sent information back. The role imposed on s by the topology of this network is one of great responsibility in communicating and integrating all incoming streams of information. Agent s was observed to play softly throughout the piece, integrating information received from all other players.

The behavior of s in this session is different from the behavior of the central agent in Session 6, the other star network. This hints at many possible strategies for attaining global coordination. The strategy observed here involved playing an appropriate "average" of all information received. This strategy proved to be mostly successful during the session, but, as expected, only some level of coordination was reached among all musicians. Here, even less of the multidimensionality of musical information was able to be shared globally since there was a sole agent responsible for transmitting the information to all musicians.

Session 5

In session 5, the outside nodes of a star network were connected in a wheel. This network is referred to as "star-wheel". In this particular case, the central node g_1 heard everybody and sent information back to everybody. The other agents not only sent information to

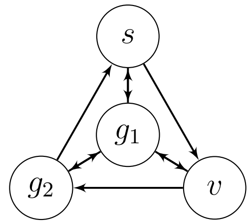


Figure 4.5: Session 5

and received information from g_1 but s also received information from g_2, who received information from v, who received information from s. It was difficult to tell this session apart from Session 1, the complete graph. It is clear that more sophisticated tools for analysis are needed for better discernment and conclusions.

Session 6

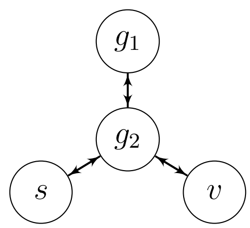


Figure 4.6: Session 6

Session six was another star network, where the central agent was g_2. The observed behavior of g_2 is very different from the behavior of s in the star network of Session 4. While s played softly and integrated the incoming information, in some sense "averaging" all musical ideas, agent g_2 seemed to disregard the incoming information and instead focused on reinforcing their own musical idea. This strategy makes sense. Due to the chaotic nature of the musical ideas put forth by the disconnected agents, a worthwhile attempt at coordination stems from continuously maintaining the same idea. This can go wrong, of course, if another participating agent does the same. In this session this strategy was observed to be mostly successful, and similar conclusions are made as in Session 4.

Sessions 4 and 6 together show the existence of at least two distinct strategies for reaching coordination. Roughly, one is to get on everybody's level and find a meaningful "center" or "average," and the other is to get everyone on your level. There is great difficulty in parsing the source of these distinct strategies employed by the agents in the center of the star. An agent may employ one of these strategies because of their musical personality, or it may be a direct response to who played first. If, for example, the improvisation is slow to start, and the center node is the one who begins putting forth musical ideas, then "getting everyone on your level" is a reasonable strategy to pursue. However, if the center node does not put forth the first musical idea, then to handle the incoming streams of musical ideas, a reasonable strategy is to "average" these ideas. In session 4, the center agent was not the first to start, and that in session 6 the center agent was the first to start. This supports our hypothesis.

Session 7

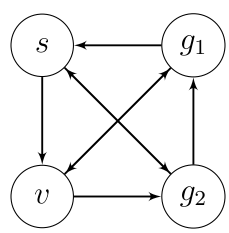


Figure 4.7: Session 7

The network underlying session 7 is almost complete. A wheel network was modified by connecting each pair of disconnected agents bidirectionally. In this particular case, since s and g_2 are disconnected in the wheel, they were connected bidirectionally. Similarly, v and g_1 were connected bidirectionally. Like Session 5, it was hard to tell this session apart from the complete network underlying Session 1. Because of this, no conclusions are made and further analysis is reserved for a later time.

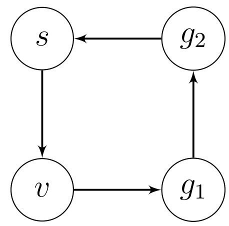


Figure 4.8: Session 8

Session 8

Session 8 is a permutation of Session 3. It is another wheel network where instead g_2 hears g_1, who hears v, who hears s, who finally hears g_2. In these sessions most comments made about Session 3 hold, although there seems to be less initial confusion and the speed with which global levels of coordination are reached is faster than in Session 3. The factors leading to this are numerous and without more sophisticated tools for analysis a proper discussion is withheld. Some hypotheses include the distribution of the agents on the network, the experience gained improvising on networks in the previous seven sessions, and the musicians being significantly more tired.[6]

Session 9

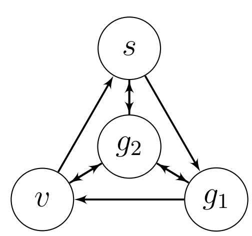


Figure 4.9: Session 9

Session 9 is another star-wheel network, like in Session 5. No additional comments are made.

[6]At this point, about 7 or 8 hours had passed since the experiments began.

4.2.2 Discussion

To reiterate, there is a lot to be extracted from these experiments and that this dissertation barely scratches the surface of the complexity found in improvisation. In the next section a simple model is developed to explore the multi-layered coordination present in improvisation. More specifically, in this model the agents have to coordinate not only rhythmically, but also melodically and harmonically. Later in the chapter, other possibilities faced by a collaborative group beyond coordination are investigated.

4.2.3 Next Experimental Steps

Software is (slowly) being developed to remotely connect musicians so that future experiments can be ran without a physical studio space or studio equipment. Instead, the musicians can play from their home using MIDI instruments[7]. Another convenience afforded by the MIDI interface is that it bypasses the main issue in remote musical collaboration– lag. When sending audio signals across the internet, due to their large size, there is unavoidable lag that would surely interfere with the creative process. Since MIDI is a sequence of numbers (representing note, velocity, echo, pedal, bend, etc.), the latency is minimal. Moreover, the software will include mechanisms to check for appropriate latency. The MIDI nature of the data generated using this software also produces more manageable data than audio signals from analog recordings. The first of which is numerical, and the latter type of data are wave files.

[7]This makes it especially easier since, for example, a computer keyboard can be used as a MIDI instrument.

4.2.4 Analytic Tools

Mathematics has proved a fruitful tool for exploring the rich structures and patterns in music. For example, there is a geometrical catalog of rhythms [16], mathematics describing distance between rhythms [15], and geometrical structures of chords [17]. However, nothing has been found that immediately aids our analysis of the improvisation experiments. It is clear that further work needs to be done to develop techniques specific to analyzing past and future experiments in a way that allows the developed models and theory to be validated, and their predictions to be tested.

4.3 Model: A Trio on Networks

A simple model of an improvising trio is offered using the coordinate system developed in chapter 2. This model is referred to as the *trio game*. The trio is composed of three agents, a saxophonist, a bassist, and a drummer. The saxophonist can play only the melody, the bassist the harmony, and the drummer the rhythm. The saxophonist and bassist still play *in* rhythm.

There are two melody lines, A_M and B_M, and two harmony lines A_H and B_H. Melody A_M sounds good with harmony A_H, and melody B_M sounds good with harmony B_H. Furthermore, there are two rhythms, C and D, and the three agents must coordinate on either rhythm. All melody and harmony lines can be adjusted to fit in with any particular rhythm. That is, as long as there is coordination in the melody and harmony, both rhythms C and D are acceptable.

The saxophonist must choose a melody line, A_M or B_M, and a rhythm, C or D. The bassist must choose a harmony line, A_H or B_H, and a rhythm, C or D. The drummer chooses only C or D. To maintain the convention established in this dissertation, refer to A_M, A_H, and

C, with $+1$, and to B_M, B_H, and D with -1. Effort is made in what follows to prevent any confusion to stem from this notation.

The improvisation is modeled with two games that happen at once– the melody and harmony game between the saxophonist and bassist, and the rhythm game between all three musicians. Refer to the saxophonist as agent s, the bassist as agent b, and the drummer as agent d. Assuming common interest and hence the structure of potential games, the payoff functions for the rhythm game are

$$\pi_s^R(r_s, r_b, r_d) = \alpha_s r_s + \gamma_{sb} r_s r_b + \gamma_{sd} r_s r_d + \delta r_s r_b r_d + \beta_{sb} r_b + \beta_{sd} r_d + \beta_{sbd} r_b r_d$$
$$\pi_b^R(r_s, r_b, r_d) = \alpha_b r_b + \gamma_{sb} r_s r_b + \gamma_{bd} r_b r_d + \delta r_s r_b r_d + \beta_{bs} r_s + \beta_{bd} r_d + \beta_{bds} r_s r_d$$
$$\pi_d^R(r_s, r_b, r_d) = \alpha_d r_d + \gamma_{sd} r_s r_d + \gamma_{bd} r_b r_d + \delta r_s r_b r_d + \beta_{ds} r_s + \beta_{db} r_b + \beta_{dsb} r_s r_b$$

The payoff functions for the melody and harmony game are

$$\pi_s^{MH}(t_s, t_b) = \alpha'_s t_s + \gamma'_{sb} t_s t_b + \beta'_{sb} t_b$$
$$\pi_b^{MH}(t_s, t_b) = \alpha'_b t_b + \gamma'_{sb} t_s t_b + \beta'_{bs} t_s$$
$$\pi_d^{MH}(t_s, t_b) = \beta'_{ds} t_s + \beta'_{db} t_b + \beta'_{dsb} t_s t_b$$

In the melody and harmony game, agent d's payoff is composed entirely of externality terms coming from agent s and b's individual efforts, β'_{ds} and β'_{db}, and their combined effort, β'_{dsb}. The drummer cannot play any melody or harmony strategies, but still hears the strategies played by the sax player and the bass player.

Since the agents are playing both games at the same time, we define their total payoffs to be

$$\pi_s((r_s, r_b, r_d), (t_s, t_b)) = \pi_s^R(r_s, r_b, r_d) + \pi_s^{MH}(t_s, t_b)$$
$$\pi_b((r_s, r_b, r_d), (t_s, t_b)) = \pi_b^R(r_s, r_b, r_d) + \pi_b^{MH}(t_s, t_b)$$
$$\pi_d((r_s, r_b, r_d), (t_s, t_b)) = \pi_d^R(r_s, r_b, r_d) + \pi_d^{MH}(t_s, t_b)$$

A strategy profile for the trio game $((r_s, r_b, r_d), (t_s, t_b))$ contains the ordered triplet (r_s, r_b, r_d) which details the rhythm strategy of agents s, b, and d, and the ordered pair (t_s, t_b) which contains the melody strategy of agent s, and the harmony strategy of agent b.

Our interest in when the agents are able to coordinate on both games motivates the following theorem.

Theorem 4.1. *The trio game on the full network is a coordination game with pure strategy Nash equilibria* $((+1,+1,+1),(+1,+1))$, $((+1,+1,+1),(-1,-1))$, $((-1,-1,-1),(+1,+1))$, $((-1,-1,-1),(-1,-1))$, *and no others, if and only if* $|\alpha_i + \delta| < \gamma_{ij} + \gamma_{ik}$ *for each agent* i *and agents* $j \neq i$ *and* $k \neq i$ *and* $|\alpha_i'| < \gamma_{sb}'$ *for* $i = s, b$.

Proof. First, for the rhythm game, $|\alpha_i + \delta| < \gamma_{ij} + \gamma_{ik}$ if and only if the Nash payoffs for each agent in $(+1,+1,+1)$ and $(-1,-1,-1)$ are positive. Hence, when these are pure strategy Nash equilibria. Because all other strategy profiles in a $2 \times 2 \times 2$ game are a unilateral deviation away from $(+1,+1,+1)$ or $(-1,-1,-1)$, there cannot be additional pure Nash equilibria. Similarly, for the melody and harmony game, $|\alpha_i'| < \gamma_{sb}'$ if and only if the Nash payoffs for each agent in $(+1,+1)$ and $(-1,-1)$ are positive. It is obvious that when this is the case, $(+1,-1)$ and $(-1,+1)$ cannot be pure Nash equilibria. Therefore, the trio game is a coordination game with pure strategy Nash equilibria $((+1,+1,+1),(+1,+1))$, $((+1,+1,+1),(-1,-1))$, $((-1,-1,-1),(+1,+1))$, $((-1,-1,-1),(-1,-1))$, and no others, if and only if $|\alpha_i + \delta| < \gamma_{ij} + \gamma_{ik}$ for each agent i and agents $j \neq i$ and $k \neq i$ and $|\alpha_i'| < \gamma_{sb}'$

for $i = s, b$. □

The pure strategy Nash equilibria described in the above theorem are those where there is coordination both in the rhythm game and in the melody and harmony game. In most improvisations we expect this to be the case; the agents are able to coordinate both melodically and harmonically, and rhythmically. The conditions that guarantee this should be familiar after chapters 2 and 3. When the agents' individual preferences are bounded by the sum of coordinative pressures, they are led by the pressure to coordinate rather than their individual preference. It is precisely in these cases that there is coordination in the trio game.

The heterogeneity of the agents in the model means that for agent i it need not be the case that $\gamma_{ij} = \gamma_{ik}$, where j and k are the remaining agents. In the case where $\gamma_{ij} > \gamma_{ik}$, most of the pressure to conform comes from agent j. This opens the model to interesting possibilities when the underlying network of listening is tampered with. For example, although the sum $\gamma_{ij} + \gamma_{ik}$ may be sufficient to push agent i to coordinate, the individual γ_{ij} and γ_{ik} may fall short of doing so. Or, it may be that γ_{ij} alone is sufficient to ensure coordination, but γ_{ik} is inadequate. In such a scenario, it is clear that a network where agent i hears only agent j will produce different music from a network where agent i hears only agent k.

Let us now analyze the trio game on the wheel and star networks. We assume that $\delta = 0$ to simplify the analysis.[8]

4.3.1 Wheel Network

There are two possibilities for the wheel network. One of which is where agent s hears agent b, who hears agent d, who then hears agent s, and the other is where agent s hears agent d, who hears agent b, who then hears agent s. Without any loss of generality, consider only

[8] Later in the chapter we study the effects of $\delta \neq 0$, although through a slightly different perspective.

the first case, shown below in Figure 4.10. All results for the other case are the same and carry over by swapping agent s and b's labels.

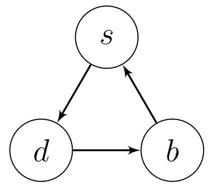


Figure 4.10: Wheel Network

The payoff functions become

$$\pi_s^R(r_s, r_b, r_d) = \alpha_s r_s + \gamma_{sb} r_s r_b + \beta_{sb} r_b$$
$$\pi_b^R(r_s, r_b, r_d) = \alpha_b r_b + \gamma_{bd} r_b r_d + \beta_{bd} r_d$$
$$\pi_d^R(r_s, r_b, r_d) = \alpha_d r_d + \gamma_{sd} r_s r_d + \beta_{ds} r_s$$

$$\pi_s^{MH}(t_s, t_b) = \alpha'_s t_s + \gamma'_{sb} t_s t_b + \beta'_{sb} t_b$$
$$\pi_b^{MH}(t_s, t_b) = \alpha'_b t_b$$
$$\pi_d^{MH}(t_s, t_b) = \beta'_{ds} t_s$$

Because agent s only hears agent b, their payoff functions for the rhythm and the melody and harmony games are adjusted to remove all terms that are dependent on agent d's strategy. This has no effect on the melody and harmony game since this game is played exclusively with agent b, but in the rhythm game it removes the $\gamma_{sd} t_s t_d$ and $\beta_{sd} t_d$ terms. Agent b, who only hears agent d, undergoes a similar transformation of their payoff function. The rhythm and the melody and harmony games lose all terms that are dependent on agent s's strategy. In the melody and harmony game, this results in agent b facing what is essentially a single-agent decision problem. Here it is obvious that if $\alpha_b > 0$, agent b will play $+1$, and if $\alpha_b < 0$, then agent b will play -1. The pattern continues for agent d, where all terms dependent on agent b are deleted from their payoff function of each game.

On the full network, we proved that $|\alpha_i + \delta| < \gamma_{ij} + \gamma_{ik}$ for each agent i and agents $j \neq i$ and $k \neq i$, makes the rhythm game is a coordination game. Taking $\delta = 0$ simplifies this to $|\alpha_i| < \gamma_{ij} + \gamma_{ik}$. On the wheel network, all agents lose one of the γ terms, resulting in the possibility that, even if the parameters give rise to a coordination game on the full network, their individual preference is larger in magnitude than the remaining γ. The possibilities resulting from this are described in the below Theorem 4.2.

Theorem 4.2. *Suppose agents i, j, and k are on a wheel network such that i hears j, who hears k, who hears i. A strategy profile for the rhythm game is written in the order i, j, k, i.e., $(\sigma_i, \sigma_j, \sigma_k)$.*

1. *If $|\alpha_i| < \gamma_{ij}$, $|\alpha_j| < \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$, then the rhythm game has pure strategy Nash equilibria $(+1, +1, +1)$ and $(-1, -1, -1)$.*
2. *If $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| < \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$, then the rhythm game has the unique pure strategy Nash equilibrium $(\operatorname{sgn}\alpha_i, \operatorname{sgn}\alpha_i, \operatorname{sgn}\alpha_i)$.*
3. *If $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| > \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$, then the rhythm game has unique pure strategy Nash equilibrium $(\operatorname{sgn}\alpha_i, \operatorname{sgn}\alpha_j, \operatorname{sgn}\alpha_i)$.*
4. *If $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| > \gamma_{jk}$, and $|\alpha_k| > \gamma_{ki}$, then the rhythm game has unique pure strategy Nash equilibrium $(\operatorname{sgn}\alpha_i, \operatorname{sgn}\alpha_j, \operatorname{sgn}\alpha_k)$.*

Proof. For case 1, suppose $|\alpha_i| < \gamma_{ij}$, $|\alpha_j| < \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$. Then each agent's Nash payoff is positive if and only if they play the same strategy as the agent from whom they receive information. In other words, all agents have a positive Nash payoff in the strategy profiles $(+1, +1, +1)$ and $(-1, -1, -1)$. Hence $(+1, +1, +1)$ and $(-1, -1, -1)$ are pure strategy Nash equilibria. No other profile can be a pure strategy Nash equilibrium because all other possibilities are a single unilateral deviation from $(+1, +1, +1)$ or $(-1, -1, -1)$, and hence must have a negative Nash payoff for the deviating agent.

For case 2, suppose $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| < \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$. Because $|\alpha_i| > \gamma_{ij}$, agent i will play the strategy $\operatorname{sgn} \alpha_i$. Agent i will never play $-\operatorname{sgn} \alpha_i$ so all Nash equilibria must have agent i playing $\operatorname{sgn} \alpha_i$. Agent k, who hears i, has $|\alpha_k| < \gamma_{ki}$, so they will have a positive Nash payoff just in case they play same thing as agent i, namely $\operatorname{sgn} \alpha_i$. Similarly, agent j, who hears k, will also play $\operatorname{sgn} \alpha_i$. Hence, $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i)$ is the unique pure strategy Nash equilibrium.

For case 3, suppose $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| > \gamma_{jk}$, and $|\alpha_k| < \gamma_{ki}$. Then, agent i will play $\operatorname{sgn} \alpha_i$ and agent j will play $\operatorname{sgn} \alpha_j$ independent of what the agent they hear plays. The remaining agent k, who hears i, has $|\alpha_k| < \gamma_{ki}$. Hence their payoff will be positive in the strategy profile where agent i plays $\operatorname{sgn} \alpha_i$. It follows that the unique pure strategy Nash equilibrium is $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_j, \operatorname{sgn} \alpha_i)$.

For case 4, suppose that $|\alpha_i| > \gamma_{ij}$, $|\alpha_j| > \gamma_{jk}$, and $|\alpha_k| > \gamma_{ki}$. Because each agent's individual preference is larger in magnitude than the pressure to conform, they receive a positive Nash payoff if and only if they play their preference. It follows immediately that the unique pure strategy Nash equilibrium is $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_j, \operatorname{sgn} \alpha_k)$. □

The above theorem describes the possible effects of the wheel network on the trio game. The goal of coordination is guaranteed in cases 1 and 2, but the last two cases yield coordination only by chance.

Case 1 preserves the coordination structure from the full network. Each agent's individual preference is bounded by the pressure to conform with the incoming agent's information. Because of this, each agent has positive Nash payoff whenever they coordinate with the neighbor from whom they receive information. It is possible then for all agents to coordinate on $(+1,+1,+1)$ or $(-1,-1,-1)$.

In case 2, only two of the agents, j and k, have their individual preferences bounded by the pressure to conform. The remaining agent i only receives a positive Nash payoff when they play their preference. Because the other agents are both playing coordination games with the incoming information, they will coordinate with agent i's preference.

In case 3, only one agent, agent k, has their individual preference bounded by the pressure to conform. Agent k will coordinate on the individual preference played by the agent they are listening to, agent i. The remaining agent j, because their individual preference supersedes their pressure to conform with k, will play their preference. Here there will be coordination if and only if $\operatorname{sgn}\alpha_i = \operatorname{sgn}\alpha_j$. Otherwise, both agents i and k will play $\operatorname{sgn}\alpha_i$ and agent j will play $\operatorname{sgn}\alpha_j = -\operatorname{sgn}\alpha_i$.

Lastly, in case 4, all agents have individual preference stronger than the pressures to conform. Because of this, they will each play their preference. The only chance for coordination in this case is if all agents happen to have the same preference.

In cases 1 and 2 there is guaranteed coordination. In cases 3 and 4 this coordination happens only by chance, if the appropriate preferences align. To apply this to the context of an improvisation, the relationships of α_s with γ_{sb}, α_b with γ_{bd}, and α_s with γ_{sd} must be considered. The first relationship involves the sax player's individual preference of a rhythm with their pressure to conform rhythmically with the bassist. The second relationship is between the bassist's individual preference and their pressure to conform with the drummer. Finally, the third relationship is that of the drummer's preference with their pressure to conform with the sax player.

To illustrate Theorem 4.2, suppose that the pressure of the sax player to conform with the bassist is not very strong, so that $|\alpha_s| > \gamma_{sb}$.[9] On the other hand, suppose that the

[9]Realistically, this is usually not the case. The bass is a very rhythmic instrument and it is expected

bassist has a strong pressure to conform with the drummer, and that the same is true of the drummer with the sax player, so that $|\alpha_b| < \gamma_{bd}$ and $|\alpha_d| < \gamma_{sd}$. According to Theorem 4.2, $(\operatorname{sgn}\alpha_s, \operatorname{sgn}\alpha_s, \operatorname{sgn}\alpha_s)$ is the unique pure strategy Nash equilibrium of the game. In other words, the agents coordinate uniquely on the sax player's preference. This happens roughly because of the sax player's low dependence on the bass player, and hence pursuit of their individual preference for the rhythm. The drummer conforms to the sax player's rhythm, and the bassist conforms to the drummer's rhythm, in turn conforming with the saxophonist.

Note that Theorem 4.2 made no assumptions about the nature of the game on the full network. The results hold even if the improvisation on the full network is not a coordination game.

What happens to the melody and harmony game on the wheel network? The answer to this question is given in the below Theorem 4.3.

Theorem 4.3. *Suppose agents i, j, and d are on a wheel network such that i hears j, who hears d, who hears i. A strategy profile for the melody and harmony game is written in the order i, j, i.e., (σ_i, σ_j).*

1. *If $|\alpha'_i| < \gamma'_{sb}$, then the melody and harmony game has pure strategy Nash equilibrium $(\operatorname{sgn}\alpha'_j, \operatorname{sgn}\alpha'_j)$.*
2. *If $|\alpha'_i| > \gamma'_{sb}$, then the melody and harmony game has pure strategy Nash equilibrium $(\operatorname{sgn}\alpha'_i, \operatorname{sgn}\alpha'_j)$.*

Proof. Agent j only receives information from agent d, who does not participate in the harmony and melody game. Because of this, j's payoff function does not include any terms

that in most cases the bassist conveys enough rhythmic information to the saxophonist to ensure rhythmic coordination.

dependent on agent i, and we can write $\pi_j^{MH}(t_s, t_b) = \alpha'_j t_j$. It is immediate that agent j's Nash payoff is positive if and only if agent j plays $\operatorname{sgn} \alpha'_j$. Agent i, on the other hand, who hears j, has utility function $\pi_i^{MH}(t_s, t_b) = \alpha'_i t_i + \gamma'_{sb} t_i t_j + \beta'_{ij} t_j$. If $|\alpha'_i| < \gamma'_{sb}$, then the Nash payoff of agent i is positive if and only if (t_s, t_b) is $(+1, +1)$ or $(-1, -1)$. In this case, the only strategy profile with positive Nash payoffs for both agents is $(\operatorname{sgn} \alpha'_j, \operatorname{sgn} \alpha'_j)$. Alternatively, if $|\alpha'_i| > \gamma'_{sb}$, then agent i's Nash payoff is positive if and only if agent i plays $\operatorname{sgn} \alpha'_i$. The only strategy profile where the payoff of both agents is positive is $(\operatorname{sgn} \alpha'_i, \operatorname{sgn} \alpha'_j)$. This completes the proof. □

Theorem 4.3 shows that there are two possibilities for the Nash structure of the harmony and melody game on the wheel network. When agent i, who hears agent j, for $i, j = s, b, i \neq j$, has individual preference greater than the pressure to conform with j, the unique pure strategy Nash equilibrium involves agent i and j each playing their individual preferences. Agent j does this regardless of the relation of their individual preference with the pressure to conform with agent i because agent j is, in their perspective, playing alone. The only chance of coordination here is if $\operatorname{sgn} \alpha'_i = \operatorname{sgn} \alpha'_j$. The other possibility is when agent i's individual preference is bounded by the pressure to conform. In this case, agent j, not hearing i, plays their preference $\operatorname{sgn} \alpha'_j$. Agent i, being pulled by the pressure to coordinate, has positive Nash payoffs in the strategy profiles $(+1, +1)$ and $(-1, -1)$. The only strategy profile with positive Nash payoffs for both agents, and hence a pure strategy Nash equilibrium, is $(\operatorname{sgn} \alpha'_j, \operatorname{sgn} \alpha'_j)$.

The intuition here is immediate. Since agent j receives no feedback from agent i, and because agent d cannot convey melodic and harmonic information, agent j is playing the melody and harmony game alone. Because of this, they will simply play their preference. Agent j only receives the rhythmic information coming from the drummer, so there are pressures to conform rhythmically, but freedom to follow their melodic and harmonic preference. Agent i, on the other hand, hears agent j's preference and conforms to it just in case $|\alpha'_i| < \gamma'_{sb}$. Agent j will never respond if agent i plays the alternative strategy in the melody and harmony game because this will never reach agent j.[10]

Theorems 4.2 and 4.3 allow for an interesting possibility. On the wheel network in Figure 4.10, if $|\alpha_s| > \gamma_{sb}$, $|\alpha_b| < \gamma_{bd}$, and $|\alpha_d| < \gamma_{sd}$ in the rhythm game, then the agents will coordinate on agent s's preference. At the same time, if $|\alpha'_s| < \gamma'_{sb}$, then agents s and b will coordinate on agent b's preference in the melody and harmony game. With this relationship between the coordinates, the unique pure strategy Nash equilibrium of the trio game is $((\operatorname{sgn}\alpha_s, \operatorname{sgn}\alpha_s, \operatorname{sgn}\alpha_s), (\operatorname{sgn}\alpha'_b, \operatorname{sgn}\alpha'_b))$.

A somewhat bizarre possibility is where $|\alpha_s| > \gamma_{sb}$, $|\alpha_b| > \gamma_{bd}$, and $|\alpha_d| < \gamma_{sd}$ in the rhythm game, and $|\alpha'_s| < \gamma'_{sb}$ in the melody and harmony game. Here, the unique pure strategy Nash equilibrium of the trio game is $((\operatorname{sgn}\alpha_s, \operatorname{sgn}\alpha_b, \operatorname{sgn}\alpha_s), (\operatorname{sgn}\alpha'_b, \operatorname{sgn}\alpha'_b))$. If $\operatorname{sgn}\alpha_s \neq \operatorname{sgn}\alpha_b$, then agents s and d are playing one rhythm, while agent b is playing another. At the same time, agents s and b are coordinating in the melody and harmony game. In other words, agents s and b agree on the melody and harmony, but disagree on the rhythm.

4.3.2 Star Network

There are three possibilities for the star network. One of which has agent s at the center, another has agent b at the center, and lastly, we can have agent d at the center. The cases where agent s or b are in the center are similar because both agents are involved in both the rhythm game and the melody and harmony game. Without loss of generality, consider the case where agent s is in the center, and the case where agent d is in the center. Both are displayed below in Figure 4.11. Furthermore, for a more concise exposition, assume the trio game on the full network is a coordination game.

[10] Assume there is some dynamic in which an agent can bring about change. For example, perhaps the agents have been playing rhythm $+1$ for far too long, and agent s wants shift the music to the rhythm -1. In this dynamic, according to some mechanism, the agents hearing this change would respond and follow the change or resist it. In a dynamic like this, on the wheel network, if agent s, who hears agent b, attempted to change what is played in the melody and harmony game, their lead would not propagate to agent b because

(a) s in Center (b) d in Center

Figure 4.11: Star Network

In the first case, where agent s is in the center of the star, the payoff functions become

$$\pi_s^R(r_s, r_b, r_d) = \alpha_s r_s + \gamma_{sb} r_s r_b + \gamma_{sd} r_s r_d + \beta_{sb} r_b + \beta_{sd} r_d$$

$$\pi_b^R(r_s, r_b, r_d) = \alpha_b r_b + \gamma_{sb} r_s r_b + \beta_{bs} r_s$$

$$\pi_d^R(r_s, r_b, r_d) = \alpha_d r_d + \gamma_{sd} r_s r_d + \beta_{ds} r_s$$

$$\pi_s^{MH}(t_s, t_b) = \alpha'_s t_s + \gamma'_{sb} t_s t_b + \beta'_{sb} t_b$$

$$\pi_b^{MH}(t_s, t_b) = \alpha'_b t_b + \gamma'_{sb} t_s t_b + \beta'_{bs} t_s$$

$$\pi_d^{MH}(t_s, t_b) = \beta'_{ds} t_s$$

In the second case, with agent d in the center of the star, the payoff functions are

$$\pi_s^R(r_s, r_b, r_d) = \alpha_s r_s + \gamma_{sd} r_s r_d + \beta_{sd} r_d$$

$$\pi_b^R(r_s, r_b, r_d) = \alpha_b r_b + \gamma_{bd} r_b r_d + \beta_{bd} r_d$$

$$\pi_d^R(r_s, r_b, r_d) = \alpha_d r_d + \gamma_{sd} r_s r_d + \gamma_{bd} r_b r_d + \beta_{ds} r_s + \beta_{db} r_b + \beta_{dsb} r_s r_b$$

$$\pi_s^{MH}(t_s, t_b) = \alpha'_s t_s$$

$$\pi_b^{MH}(t_s, t_b) = \alpha'_b t_b$$

$$\pi_d^{MH}(t_s, t_b) = \beta'_{ds} t_s + \beta'_{db} t_b$$

---

it would have to get past agent d who cannot convey melodic and harmonic information.

A striking difference between the two cases concerns the melody and harmony game. In the first case, an agent that plays both games is in the center. Coordination in the melody and harmony game transfers from the full network since both agents involved in the game are connected to each other bidirectionally. In the second case, where the drummer is in the middle, coordination in the melody and harmony game happens just in case the saxophonist and the drummer have the same preference. This is captured succinctly in the following theorem.

Theorem 4.4. *Suppose the melody and harmony game is a coordination game on the full network. Agents s and b will coordinate on the melody and harmony game on the star network if and only if agent d is not in the center vertex or* $\operatorname{sgn}\alpha'_s = \operatorname{sgn}\alpha'_b$.

Proof. We prove the equivalent expression that agents s and b do not coordinate on the melody and harmony game on the star network if and only if agent d is in the center and $\operatorname{sgn}\alpha'_s \neq \operatorname{sgn}\alpha'_b$.

Suppose agent d is in the center and $\operatorname{sgn}\alpha'_s \neq \operatorname{sgn}\alpha'_b$. This is true if and only if agent s's Nash payoff for the melody and harmony game is $\alpha'_s t_s$, and agent b's is $\alpha'_b t_b$. This means agent s will play $\operatorname{sgn}\alpha'_s$ and agent b will play $\operatorname{sgn}\alpha'_b$. Because these are not equal, agents s and b will not coordinate. This completes the proof. □

Theorem 4.4 is both deep and obvious. The drummer being at the center, who cannot convey information about the melody and harmony, can only help coordinate the rhythm between agents s and b. The melody and harmony parts played by agents s and b will only align if the agents happen to have the same preferences. Note that this is a highly idealized model with only 2 strategies for the melody and harmony game. In a more realistic situation, there are several options for the melody and harmony and the alignment of the agents' preferences is unlikely. Hence, we expect that in general, if the drummer is in the center of the star, coordination of harmony and melody does not occur.

The below theorem details the effect of the star network on the rhythm game that is a coordination game on the full network.

Theorem 4.5. *Suppose the rhythm game is a coordination game on the full network. Suppose agents i, j, and k are on a star network such that i and j hear each other, and j and k hear each other, but i and k are disconnected. A strategy profile is written in order i, j, k, i.e., $(\sigma_i, \sigma_j, \sigma_k)$.*

1. *If $|\alpha_i| < \gamma_{ij}$ and $|\alpha_k| < \gamma_{ki}$, then the game has exactly two pure strategy Nash equilibria, namely $(+1, +1, +1)$ and $(-1, -1, -1)$.*

2. *If $|\alpha_i| > \gamma_{ij}$ and $|\alpha_k| < \gamma_{ki}$, then the game has the unique pure strategy Nash equilibrium $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i)$.*

3. *If $|\alpha_i| > \gamma_{ij}$ and $|\alpha_k| > \gamma_{ki}$, then if $\operatorname{sgn} \alpha_i = \operatorname{sgn} \alpha_k$, then the game has unique pure strategy Nash equilibrium $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i)$. Otherwise, the game has unique pure strategy Nash equilibrium $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_k)$ if and only if $\operatorname{sgn} \alpha_i \cdot \alpha_j + \gamma_{ij} - \gamma_{jk} > 0$, or it has unique pure strategy Nash equilibrium $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_k, \operatorname{sgn} \alpha_k)$ if and only if $\operatorname{sgn} \alpha_k \cdot \alpha_j - \gamma_{ij} + \gamma_{jk} > 0$.*

Proof. Before beginning, note that agent j's Nash payoff of the rhythm game is identical to that in the full network. Consequently, $|\alpha_j| < \gamma_{ij} + \gamma_{jk}$.

Case 1 follows almost immediately. Because all agents have individual preference bounded by their total pressures to conform, they have positive Nash payoff in the strategy profiles $(+1, +1, +1)$ and $(-1, -1, -1)$. Hence, these profiles are pure strategy Nash equilibria.

Now, suppose $|\alpha_i| > \gamma_{ij}$ and $|\alpha_k| < \gamma_{ki}$. The first inequality implies that agent i will play $\operatorname{sgn} \alpha_i$ independent of what agent j plays, because i's preference is larger than the pressure to conform. Hence, any pure strategy Nash equilibrium must have agent i playing $\operatorname{sgn} \alpha_i$.

Agents j and k both have individual preferences smaller than total pressures to conform, so they will have positive Nash payoffs in $(+1, +1, +1)$ and $(-1, -1, -1)$. Then, because the Nash equilibrium must have agent i playing $\operatorname{sgn} \alpha_i$, the only possibility for the equilibrium is $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i)$. This proves case 2.

For case 3, agents i and k each have individual preference greater than pressure to conform, so agent i will play $\operatorname{sgn} \alpha_i$ and agent k will play $\operatorname{sgn} \alpha_k$, and any pure strategy Nash equilibrium must have agent i and k playing their preferences. If they are equal, then it is immediate that the unique pure strategy Nash equilibrium is $(\operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i, \operatorname{sgn} \alpha_i)$. If they are not equal, then agent j, who is playing a coordination game with agents i and k, is being pulled in two directions at once. Agent j will play $\operatorname{sgn} \alpha_i$ if and only if $\operatorname{sgn} \alpha_i \cdot \alpha_j + \gamma_{ij} - \gamma_{jk} > 0$. On the other hand, agent j will play $\operatorname{sgn} \alpha_k = -1$ if and only if $\operatorname{sgn} \alpha_k \cdot \alpha_j - \gamma_{ij} + \gamma_{jk} > 0$. This completes the proof. $\square$

A consequence of Theorem 4.5 is that, perhaps counter-intuitively, the center agent has little power in enforcing their preference. This power, instead, is coming from the agents at the extremes, which is made evident in cases 2 and 3 of the theorem. Case 3 displays an interesting situation where the center agent is forced to mediate between two disagreeing agents. The center agent can only coordinate with one of their neighbors. Here, the sum of the center agent's preference and the cost-benefit of conforming with one agent but not the other is what sways the decision.

Furthermore, Theorems 4.4 and 4.5 together give us the tools to understand the many possibilities of the trio game on the star network. For example, we may have the drummer at the center, preventing melodic and harmonic coordination when $\operatorname{sgn} \alpha'_s \neq \operatorname{sgn} \alpha'_b$. At the same time, assuming the drummer has a high pressure to conform with both the sax player and the bassist so that the rhythm game matches case 1 of Theorem 4.5, guarantees the agents will coordinate rhythmically.

4.3.3 Discussion

This simple model of an improvising trio alludes heavily to the notion of information transfer. In Section 4.5, we offer a general model of collaboration where instead of payoffs, the agents generate information. This thesis is providing only the initial stepping stones to this kind of research, and we provide few answers. However, we give suggestions as to how to interpret an agent's limited information transfer (like the drummer) in this alternative framework. Future development of the model will take into account possible ambiguity in how the information is transferred by some of the agents. For example, a sax player that hears a drummer and synchronizes with them rhythmically, may not be communicating adequately, or enough of this rhythmic information to other agents.

In addition, note that no assumptions were made about the externalities in this model. The model introduced in Section 4.5 incorporates the externalities in the perspective of information transfer rather than payoffs. We save all discussion of externalities to this section.

Finally, we assumed throughout this section that both the melody and harmony game and the rhythm game are coordination games on the full network. In reality, although it is speculated that this is required for an improvising group to produce "good" music, this may not always be the case. Work that is currently being developed, but that is presently beyond the scope of this thesis, details when, because of possible miscoordination on the full network, a limited listening network is *better* for the musicians.

As an additional study into the possibilities offered by the coordinate system, the following section offers a model of an orchestra, and makes an argument for the importance of the conductor in coordinating the musicians' interpretation of a written piece of music.

4.4 Model: An Orchestra and the Conductor

Research into orchestras as it pertains to the theme of this thesis has not, to the best of the author's knowledge, been addressed by academics. But there are useful entries in resources like The Concise Oxford Dictionary of Music [9], wherein the entry for "conducting" includes "a conductor is not merely responsible for the technical excellence of the performance but also for projecting his personal attitude to the composer's intentions," and "the art (or method) of controlling an orchestra [...] involving the beating of time, ensuring of correct entries, and the 'shaping' of individual phrasing."

The allusion to the conductor's role in coordination is immediate. As stated in the two quotes in the previous paragraph, among other coordinative responsibilities, the conductor helps the performing musicians coordinate their individual interpretations of the piece. To further motivate this, a brief discussion is provided below on some reasons why the coordination of interpretations is important, as commonly agreed upon by musicians and orchestra members.

A piece of written music contains several pieces of information to help the musician play the song. Some of these, like the key, are firm and objective. Other pieces of information, like the dynamics, are more subjective, and depend largely on each individual's interpretation.[11] This ambiguous side of music notation is widely recognized by musicians. An orchestra provides an interesting problem wherein there is a large number of musicians, each of whom have their own interpretation of the piece. Coordinating all interpretations to create a coherent piece of music is not trivial. It is no surprise that orchestras have a person dedicated to coordinating the interpretation of the piece– the conductor.

In terms of networks, without a conductor, an orchestra can be thought of as a complete weighted network. All musicians hear all other musicians, but those nearby are naturally

[11] More specifically, the dynamics consist of Italian words describing how loud or soft certain parts of the music should be played. Some examples are *pianissimo*, meaning "very soft," *forte*, meaning "loud," and *crescendo*, meaning "getting louder."

louder, and hence their link has a stronger weight. The conductor is in a position where all musicians are heard well, and where all musicians can see the conductor. Using the tools developed in this thesis, we offer an explanation for an orchestra's dependence on the conductor.

An orchestra is modeled as a 5-agent coordination game, comprised of four musicians and one conductor. Each musician has a preference for one of two interpretations of the piece, the first interpretation is represented by strategy $+1$, and the second by strategy -1. For simplicity, the model is limited to order two terms of the coordinate system, *i.e.,* only α's and γ's. To analyze the model, we make the additional simplifying assumption that the agents are on a line and can only hear their neighbors. This is a simplification of the assumption that agents are in a fully connected network with weights dependent on location. This network is shown in Figure 4.12.

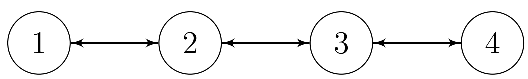


Figure 4.12: Orchestra without a Conductor

Assuming common interest, the agents' payoff functions can be written as

$$\pi_1(t_1, t_2, t_3, t_4) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2$$

$$\pi_2(t_1, t_2, t_3, t_4) = \alpha_2 t_2 + \gamma_{12} t_1 t_2 + \gamma_{23} t_2 t_3 + \beta_{21} t_1 + \beta_{23} t_3$$

$$\pi_3(t_1, t_2, t_3, t_4) = \alpha_3 t_3 + \gamma_{23} t_2 t_3 + \gamma_{34} t_3 t_4 + \beta_{32} t_2 + \beta_{34} t_4$$

$$\pi_4(t_1, t_2, t_3, t_4) = \alpha_4 t_4 + \gamma_{34} t_3 t_4 + \beta_{43} t_3$$

The assumption that this is a coordination game between for each agent and their neighbors means that the following inequalities must be true

- $|\alpha_1| < \gamma_{12}$

- $|\alpha_2| < \gamma_{12} + \gamma_{23}$
- $|\alpha_3| < \gamma_{23} + \gamma_{34}$
- $|\alpha_4| < \gamma_{34}$

Although each local game, that of each agent with their neighbors, is a coordination game, is the global game a coordination game? It turns out that this depends on the distribution of preferences. Which sets of preferences give rise to a global coordination game and which do not?

There are $2^4 = 16$ possible distributions of preferences. In two of these, all agents have the same preference of $+1$ in one case and -1 in the other. In eight of the distributions, three of the agents, a majority, have the same preference. The remaining six possibilities are those in which two agents prefer $+1$ and two prefer -1. An initial intuition might be that the first ten described cases pose no issues in coordination; that is, they give rise to a global coordination game, and that the remaining six cases may have pure strategy Nash equilibria beyond the pure coordinative profiles $(+1, +1, +1, +1)$ and $(-1, -1, -1, -1)$. As it turns out, this is not the case.

To detail the Nash equilibrium structure of the orchestra game without the conductor, two theorems are proved below.

Theorem 4.6. *The orchestra game with four agents on a line network, as in Figure 4.12, is a global coordination game unless*

1. $\alpha_2 + \gamma_{12} - \gamma_{23} > 0$ *and* $-\alpha_3 - \gamma_{23} + \gamma_{34} > 0$*, in which case the game has the additional Nash equilibrium* $(+1, +1, -1, -1)$.
2. $-\alpha_2 + \gamma_{12} - \gamma_{23} > 0$ *and* $\alpha_3 - \gamma_{23} + \gamma_{34} > 0$*, in which case the game has the additional Nash equilibrium* $(-1, -1, +1, +1)$.

Proof. We remind the reader that every agent is playing a coordination game with their neighbor(s). Because of this, the strategy profiles $(+1, +1, +1, +1)$ and $(-1, -1, -1, -1)$ have positive Nash payoffs for every agent. Consequently, any profile that is a unilateral deviation away from either $(+1, +1, +1, +1)$ and $(-1, -1, -1, -1)$ must have a negative Nash payoff for at least one agent and cannot be a Nash equilibrium. The only remaining strategy profiles that could be pure strategy Nash equilibria are those in which exactly two agents play $+1$ and the remaining two play -1. In all cases, the agents at the extremes, agents 1 and 4, have positive Nash payoffs just in case they coordinate with their neighbor, agents 2 and 3, respectively. Two candidate profiles remain, $(+1, +1, -1, -1)$ and $(-1, -1, +1, +1)$. It is left to analyze the conditions in which the Nash payoffs of agents 2 and 3 are positive in these profiles. In strategy profile $(+1, +1, -1, -1)$, agent 2's Nash payoff is $\alpha_2 + \gamma_{12} - \gamma_{23}$, and agent 3's is $-\alpha_3 - \gamma_{23} + \gamma_{34}$. In $(-1, -1, +1, +1)$ they are $-\alpha_2 + \gamma_{12} - \gamma_{23}$ and $\alpha_3 - \gamma_{23} + \gamma_{34}$. This completes the proof. □

Theorem 4.6 illuminates the importance of agents 2 and 3, the agents in the middle of the line network. Suppose, as a first point of analysis, that $\alpha_2 = \alpha_3 = 0$. Then the theorem shows that if agent 2's pressure to conform with agent 1 is greater than the pressure to conform with agent 3, and agent 3's pressure with agent 4 is greater than the pressure with agent 2, the game will have the additional equilibria $(+1, +1, -1, -1)$ and $(-1, -1, +1, +1)$. When $\alpha_2 \neq 0$ and $\alpha_3 \neq 0$, then the previous sentence must be slightly modified to accommodate these preferences. Namely, if agent 2's individual preference together with their pressure to conform with agent 1 is larger than their pressure to conform with 3, it is possible to have a Nash equilibrium where the agents don't globally coordinate. In addition, when the analogous statement for agent 3 is true (replacing agent 1 in agent 2's statement with agent 4), the game can result in mis-coordination.

In the case where $\gamma_{12} = \gamma_{23} = \gamma_{34}$ Theorem 4.6 reduces to the following corollary.

Corollary 4.6.1. *The orchestra game with four agents on a line network, as in Figure*

4.12, with $\gamma_{12} = \gamma_{23} = \gamma_{34}$ is a global coordination game unless agents 2 and 3 have opposing individual preferences. In this case the game has the additional pure strategy Nash equilibrium $(\operatorname{sgn}\alpha_2, \operatorname{sgn}\alpha_2, \operatorname{sgn}\alpha_3, \operatorname{sgn}\alpha_3)$

Proof. The assumption that $\gamma_{12} = \gamma_{23} = \gamma_{34}$ simplifies the condition for the game to have the additional Nash equilibrium $(+1, +1, -1, -1)$ to be $\alpha_2 > 0$ and $\alpha_3 < 0$. Similarly, the condition for the game to have $(-1, -1, +1, +1)$ as an additional equilibrium becomes $\alpha_2 < 0$ and $\alpha_3 > 0$. Each of these are true if and only if $\operatorname{sgn}\alpha_2 \neq \operatorname{sgn}\alpha_3$. The result follows immediately from this. □

The above corollary shows that in the case where the pressures to conform are identical across all agents, the determining factor in the Nash structure of the orchestra game is the relationship between $\operatorname{sgn}\alpha_2$ and $\operatorname{sgn}\alpha_3$. When these are the same, the global game is a coordination game. When they are not, in addition to the equilibria $(+1, +1, +1, +1)$ and $(-1, -1, -1, -1)$, the game has the equilibrium where agents 1 and 2 coordinate on agent 2's individual preference, and agents 3 and 4 coordinate on agent 3's individual preference.

This model is idealized in many ways. An entire orchestra (which often times has over 100 members) is modeled with only four agents. Furthermore, the complexities of the listening network are simplified to a line where each agent only hears their immediate neighbor(s). In addition, there are many leaders in an orchestra, like the first chair musicians, an important feature we are ignoring. Nevertheless, we believe the model captures how, in many cases, a global coordination game with heavy weight on local interactions has pure strategy Nash equilibria that are not coordinative strategy profiles. For an orchestra this could mean an incongruous performance, which is to be avoided.

4.4.1 The Conductor

The role, and even need, of the conductor is not entirely agreed upon in the orchestral world. Some say the conductor is not necessary at all, others say they are important during rehearsal attributing this to prepping the orchestra for coordination during performance, and others say the conductor is crucial in both the rehearsal and the performance.

We proceed in our simple model by considering the effect of a conductor only during the performance. We model the integration of a conductor as a new vertex that sends information to all members of the orchestra, but that does not receive information back. In reality this is certainly not the case; the conductor is not blindly conducting and responds to the performance in real-time. We ignore this additional complexity, assuming that the conductor has a fixed interpretation of the piece they will enforce on the orchestra. Any information the conductor receives, we assume, would be used solely to guide them in enforcing this fixed interpretation. The new network is displayed in Figure 4.13.

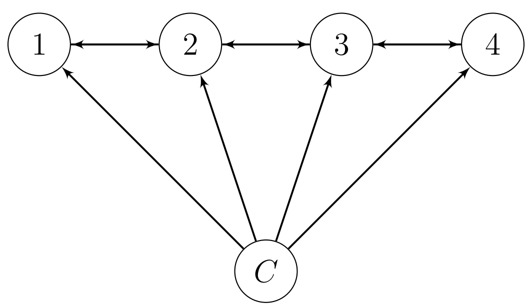


Figure 4.13: Orchestra with a Conductor

The payoff functions are now

$$\pi_1(t_1, t_2, t_3, t_4, t_C) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 + \gamma_{1C} t_1 t_C + \beta_{1C} t_C$$

$$\pi_2(t_1, t_2, t_3, t_4, t_C) = \alpha_2 t_2 + \gamma_{12} t_1 t_2 + \gamma_{23} t_2 t_3 + \beta_{21} t_1 + \beta_{23} t_3 + \gamma_{2C} t_2 t_C + \beta_{2C} t_C$$

$$\pi_3(t_1, t_2, t_3, t_4, t_C) = \alpha_3 t_3 + \gamma_{23} t_2 t_3 + \gamma_{34} t_3 t_4 + \beta_{32} t_2 + \beta_{34} t_4 + \gamma_{3C} t_3 t_C + \beta_{3C} t_C$$

$$\pi_4(t_1, t_2, t_3, t_4, t_C) = \alpha_4 t_4 + \gamma_{34} t_3 t_4 + \beta_{43} t_3 + \gamma_{4C} t_4 t_C + \beta_{4C} t_C$$

$$\pi_4(t_C) = \alpha_C t_C$$

Each member of the orchestra now has a pressure to conform with the conductor, and an externality term from the conductor's decision. Because the conductor receives no incoming edges, they are simply facing a single-agent decision problem between two options. If $\alpha_M > 0$, then the conductor has a preference for $+1$ and will undoubtedly play this. On the other hand, if $\alpha_M < 0$, then the conductor will play -1. We assume, without loss of generality, that $\alpha_M > 0$, and that the conductor plays $+1$. This allows us to simplify the payoff functions to

$$\pi_1(t_1, t_2, t_3, t_4, +1) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 + \gamma_{1C} t_1 + \beta_{1C}$$

$$\pi_2(t_1, t_2, t_3, t_4, +1) = \alpha_2 t_2 + \gamma_{12} t_1 t_2 + \gamma_{23} t_2 t_3 + \beta_{21} t_1 + \beta_{23} t_3 + \gamma_{2C} t_2 + \beta_{2C}$$

$$\pi_3(t_1, t_2, t_3, t_4, +1) = \alpha_3 t_3 + \gamma_{23} t_2 t_3 + \gamma_{34} t_3 t_4 + \beta_{32} t_2 + \beta_{34} t_4 + \gamma_{3C} t_3 + \beta_{3C}$$

$$\pi_4(t_1, t_2, t_3, t_4, +1) = \alpha_4 t_4 + \gamma_{34} t_3 t_4 + \beta_{43} t_3 + \gamma_{4C} t_4 + \beta_{4C}$$

$$\pi_4(+1) = \alpha_C$$

Or, collecting like terms,

$$\pi_1(t_1, t_2, t_3, t_4, +1) = (\alpha_1 + \gamma_{1C})t_1 + \gamma_{12}t_1t_2 + \beta_{12}t_2 + \beta_{1C}$$

$$\pi_2(t_1, t_2, t_3, t_4, +1) = (\alpha_2 + \gamma_{2C})t_2 + \gamma_{12}t_1t_2 + \gamma_{23}t_2t_3 + \beta_{21}t_1 + \beta_{23}t_3 + \beta_{2C}$$

$$\pi_3(t_1, t_2, t_3, t_4, +1) = (\alpha_3 + \gamma_{3C})t_3 + \gamma_{23}t_2t_3 + \gamma_{34}t_3t_4 + \beta_{32}t_2 + \beta_{34}t_4 + \beta_{3C}$$

$$\pi_4(t_1, t_2, t_3, t_4, +1) = (\alpha_4 + \gamma_{4C})t_4 + \gamma_{34}t_3t_4 + \beta_{43}t_3 + \beta_{4C}$$

$$\pi_4(+1) = \alpha_C$$

In chapter 2 we defined three classes of potential games for 2×2 games, namely independent, quasi-independent, and dependent potential games.[12] Each of these classes is defined by the relationship of α_i and α_j, for $i \neq j$, the individual preference parameters of the two agents i and j, with γ, the coordinative pressure. The class of independent potential games are those where the magnitude of both individual preferences is greater than $|\gamma|$. Quasi-independent potential games have $|\gamma|$ sandwiched between the magnitudes of the two preferences. Both of these classes have a unique pure strategy Nash equilibrium. Lastly, in the class of dependent potential games it is $|\gamma|$ that dominates the magnitudes of the individual preferences. These games have two pure strategy Nash equilibria.

Without the conductor, the orchestra game has the flavor of a 2×2 dependent potential game. This is because the magnitudes of all individual preferences are bounded by the corresponding γ. Integrating the conductor allows us to break free from dependent potential games. It is immediate from the payoff functions that the γ_{iC} term has direct influence on each α_i for $i = 1, 2, 3, 4$. When it is positive, it promotes a preference of $+1$ for the four musicians. If it is large enough, it will cause at least one agent i's t_i coefficient, which is $\alpha_i + \gamma_{iC}$, to exceed the bound given by the pressure to conform. When this happens, the game changes flavor to that of quasi-independent potential games. If the γ_{iC} for agents

[12]For an exposition see chapter 2.

$i = 1, 2, 3, 4$ is large enough for all agents, then the game becomes like the independent 2×2 potential games.

How strong does γ_{iC}, for $i = 1, 2, 3, 4$, have to be for the game to have the single Nash equilibrium where all agents play the conductor's preference? We answer this question in the following theorem.

Theorem 4.7. *A sufficient condition for the orchestra game with a conductor to have the unique pure strategy Nash equilibrium* $(\operatorname{sgn} \alpha_C, \operatorname{sgn} \alpha_C, \operatorname{sgn} \alpha_C, \operatorname{sgn} \alpha_C, \operatorname{sgn} \alpha_C)$ *is* $\gamma_{2C} > -\alpha_2 + \gamma_{12} + \gamma_{23}$ *and* $\gamma_{3C} > -\alpha_3 + \gamma_{23} + \gamma_{34}$.

Proof. Without loss of generality, let α_C be such that $\operatorname{sgn} \alpha_C = +1$. Now, note that if γ_{iC}, for $i = 2, 3$, is strong enough to make $+1$ a strictly dominant strategy for agents 2 and 3, then the agents at the extreme will follow their lead in playing $+1$, even if γ_{jC}, for $j = 1, 4$, is not strong enough to make $+1$ a strictly dominant strategy for agents 1 and 4. This is because of local coordinative nature assumed in the model. Agents 2 and 3, from their perspective, are each playing a 3-agent coordination game. It is sufficient to show the conditions in which γ_{2C} makes agent 2's Nash payoff in $(-1, +1, -1, \pm 1)$ positive. Agent 4's strategy in this case is written ± 1 because agent 2's payoffs are invariant to agent 4's strategy, since they are disconnected on the network. If γ_{2C} makes this Nash payoff positive, then it necessarily makes the payoffs in $(+1, +1, -1, \pm 1)$ and $(-1, +1, +1, \pm 1)$ positive, which is easily verified using agent 2's payoff function and the assumptions on γ_{12} and γ_{23}. Because the game is a local coordination game, agent 2's Nash payoff in $(+1, +1, +1, \pm 1)$ is automatically positive. Hence, values of γ_{2C} that make agent 2's payoff in $(-1, +1, -1, \pm 1)$ positive, make the strategy $+1$ strictly dominant for agent 2. Similarly for agent 3 and the strategy profile $(\pm 1, -1, +1, -1)$. This is the case precisely when $\gamma_{2C} > -\alpha_2 + \gamma_{12} + \gamma_{23}$ and $\gamma_{3C} > -\alpha_3 + \gamma_{23} + \gamma_{34}$. □

Admittedly, the statement of Theorem 4.7 does not cover all possibilities in which the conductor is able to successfully influence the orchestra to coordinate on their interpretation $+1$. The intention is to offer an initial characterization of such conditions, which already indicates great potential in furthering this research. The condition described by Theorem 4.7 informs of the importance of the conductor's influence on the middle agents 2 and 3, the next most influential agents. For both agents 2 and 3, individually, the conductor's influence must be larger than their preference for interpretation -1 together with the pressure to conform with both neighbors, who, in the strategy profile of concern, are playing -1. Work in progress focuses on an exhaustive understanding of the conditions that guarantee coordination on the conductor's interpretation.

As an example, take $\gamma_{12} = \gamma_{23} = \gamma_{34} = 2$, $\alpha_1 = \alpha_2 = 1$, and $\alpha_3 = \alpha_4 = -1$. Furthermore, take $\gamma_{1C} = \gamma_{2C} = \gamma_{3C} = \gamma_{4C}$, which we denote by γ_C. Without the conductor, this orchestra game has the pure strategy Nash equilibria $(+1, +1, +1, +1)$, $(-1, -1, -1, -1)$, and $(+1, +1, -1, -1)$, according to Corollary 4.6.1. Including a conductor such that $\gamma_C > 5$ makes $(+1, +1, +1, +1, +1)$ the unique pure strategy Nash equilibrium of the game, where all agents coordinate on the conductor's interpretation.

4.5 Model: Collaboration

In this section we expand on the idea of interpreting payoffs as information. To make the distinction precise, define a collaboration as below, almost identically to a game.

Definition 4.8. A collaboration $\mathcal{C}$ consists of a set of agents $\mathcal{N} = \{1, \ldots, n\}$ where each agent $i \in \mathcal{N}$ has a set of k_i strategies $S_i = \{\sigma_{i1}, \ldots, \sigma_{ik_i}\}$ and contribution function $\psi_i : \mathcal{S} \to \mathbb{R}$, where $\mathcal{S} = \prod_{i \in \mathcal{N}} S_i$. We write this as $\langle \mathcal{N}, \mathcal{S}, \Psi \rangle$ where $\Psi = \{\psi_1, \ldots, \psi_n\}$.

The definition above differs from the definition of a game only in the interpretation of ψ_i as contribution functions instead of payoff functions. Because of this similarity, in this section, we use the terms collaboration and game interchangeably. The contribution function of each agent maps the given strategy profile to a quantity of information generated in that profile. For example, a discussion can be modeled using the strategies *speak*, which we denote with $+1$, and *listen*, which we denote by -1. A naive approach in modeling a 2-agent discussion is to choose quantities of information generated and received in each profile, for example in Table 4.1.

	$+1$	-1
$+1$	1 1	4 4
-1	3 3	0 0

Table 4.1: 2×2 Discussion Model

We re-emphasize that the quantities in each cell of the normal-form collaboration are not payoffs, but information received. It immediately follows that if both agents are listening, strategy profile $(-1, -1)$, then no information is generated. On the other hand, when agent 1 is speaking and agent 2 listening, profile $(+1, -1)$, more information is generated for both agents than when agent 2 is doing the talking, profile $(-1, +1)$. In addition, the profile where both speak, $(+1, +1)$, generates some information, but less than when only one agent is speaking. The intuition here is that only some of the information gets through when the agents are trying to speak over each other. Although the quantities represent information, we assume the amount of information is directly proportional to the payoff associated with the information; more information is worth more. The agents, like in standard game theory, want to maximize the information produced.

We can do better than this.[13] Because the coordinate system developed in chapter 2 does not depend on the interpretation that the numerical quantities received are payoffs, the structure can be borrowed and the interpretation of payoffs swapped to that of information. Hence,

[13] The naive model is already alluding to an intuitive identical play structure to collaborations, which we will derive in the next section.

the contribution functions in a 2×2 game are

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 \tag{4.1}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 + \gamma_{21} t_1 t_2 + \beta_{21} t_2. \tag{4.2}$$

When agent 1 plays +1 they generate α_1 information for themselves, β_{21} information for agent 2, and $\gamma_{12} t_2$ information *with* agent 2, depending on t_2. The coordinate γ_{12} represents the information agent 1 is able to contribute from interacting with agent 2. The quantity γ_{12} represents the collaborative information generated by agent 1 from their feedback, or collaboration, with agent 2. Depending on the structure of the game, when this collaboration is fruitful, γ_{12} supports the collaboration (be it in the form of a coordination or anti-coordination game). On the other hand, when the collaboration is problematic, γ_{12} takes away from the total information contributed. The quantities α_2, γ_{21}, and β_{21} are interpreted in the same way, but with agents 1 and 2 swapped.

It is not necessarily the case that $\gamma_{12} = \gamma_{21}$. For example, perhaps agent 1 is inspired by agent 2 and is able to generate more information from their collaboration. At the same time, perhaps agent 2 is in some way debilitated from their interaction with agent 1, and generates less information from the collaboration. This would be reflected by $\gamma_{12} > 0$ and $\gamma_{21} < 0$.

We are inspired to modify the payoff functions (4.1) and (4.2). Because agent 1 and 2 are collaborating and communicating, the collaborative information agent 1 generates working with agent 2, γ_{12}, must reach agent 2, at least to some degree, in a way that is distinct from β_{21}. This is because it depends on both t_1 and t_2 whereas β_{21} depends only on t_1. This information is denoted by β_{212}, where the first index, 2, represents agent 2, who receives the information, and the following 12 represents that the information comes from agent 1 working with agent 2. Similarly, the information agent 2 generates with agent 1, γ_{21}, must reach agent 1 to some extent, which is denoted by β_{121}. Because this information depends on the actions of both agents, it is a coefficient of t_1t_2. The contribution functions are now

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 + \beta_{121} t_1 t_2 \tag{4.3}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 + \gamma_{21} t_1 t_2 + \beta_{21} t_2 + \beta_{212} t_1 t_2 \tag{4.4}$$

Collecting like terms,

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + (\gamma_{12} + \beta_{121}) t_1 t_2 + \beta_{12} t_2 \tag{4.5}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 + (\gamma_{21} + \beta_{212}) t_1 t_2 + \beta_{21} t_1 \tag{4.6}$$

The externality terms are assumed to be functions of their corresponding Nash components. When agent 1 plays +1, they generate α_1 information for themselves, and β_{21} information for agent 2, where β_{21} is a function of α_1. Similarly, β_{12} is a function of α_2, β_{121} is function of γ_{21}, and β_{212} is a function of γ_{12}.

4.5.1 Perfect Communication

Suppose that, in a highly idealized situation, the agents are able to perfectly communicate all information generated. When agent 1, for example, plays +1, they generate α_1 for themselves, and β_{21} for agent 2. If this information is perfectly communicated, then $\alpha_1 = \beta_{21}$. Similarly, $\alpha_2 = \beta_{12}$, $\gamma_{12} = \beta_{212}$, and $\gamma_{21} = \beta_{121}$. The contribution functions can be rewritten as

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + (\gamma_{12} + \gamma_{21}) t_1 t_2 + \alpha_2 t_2 \tag{4.7}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 + (\gamma_{21} + \gamma_{12}) t_1 t_2 + \alpha_1 t_1 \tag{4.8}$$

Because the shared coefficients, *i.e.*, those in front of t_1t_2, are the same for both agents, the collaboration is in the form of an exact potential game and hence has a potential function. Furthermore, it is clear that $\psi_1(t_1, t_2) = \psi_2(t_1, t_2)$ and that, in the language of game theory, this is an identical play game. This is stated as a theorem.

Theorem 4.9. *A collaboration with perfect communication has the form of an identical play game.*

Proof. The proof is immediate since equations (4.7) and (4.8) are equal. □

Furthermore, because the collaboration is identical play, as we saw in chapter 2, the social welfare function and potential function of the game align. Hence, the strategy profile that globally maximizes the potential function also globally maximizes the social welfare function. The global maximum of the potential function, the risk-dominant Nash equilibrium, is the strategy profile that maximizes the information generated by the agents individually. With perfect communication and identical play, this information is perfectly communicated and hence the strategy profile maximizes all information obtained, not only individually but collectively. This is captured in the theorem below.

Theorem 4.10. *The risk-dominant equilibrium of an identical play collaboration produces the maximum global information.*

Proof. Consider the risk-dominant equilibrium of an identical play collaboration. This strategy profile globally maximizes the potential function. Because of identical play, it also maximizes the social welfare function. Since the social welfare function computes the sum of payoffs in the strategy profiles of a game, it immediately follows that the risk-dominant Nash equilibrium produces the highest collective payoff. Changing the interpretation of payoffs to information completes the proof. □

An important implication of this theorem is that if the agents in the collaboration with perfect communication are choosing their strategies "safely" in the face of uncertainty of what the other will choose, they will play the risk-dominant equilibrium which maximizes the total information produced. This raises the important question of what, on the other hand, certainty of one another's actions will allow the agents to produce. This question, unfortunately, is saved for future work.

The next section extends the collaboration model to the case of 3 agents and 2 strategies. After this, we show the effect of networks on 2×2 collaborations before concluding the chapter with an analysis of $2 \times 2 \times 2$ collaborations on the three networks of interest– the star, wheel, and full networks.

4.6 $2 \times 2 \times 2$ Collaboration

The contribution functions in a $2 \times 2 \times 2$ collaboration are

$$\begin{aligned}\psi_1(t_1, t_2, t_3) = \alpha_1 t_1 + &(\gamma_{12} + \beta_{121}) t_1 t_2 + (\gamma_{13} + \beta_{131}) t_1 t_3 \\ &+ (\delta_{123} + \beta_{1231} + \beta_{1312}) t_1 t_2 t_3 + \beta_{12} t_2 + \beta_{13} t_3 + (\beta_{123} + \beta_{132}) t_2 t_3\end{aligned}$$

$$\begin{aligned}\psi_2(t_1, t_2, t_3) = \alpha_2 t_2 + &(\gamma_{21} + \beta_{212}) t_1 t_2 + (\gamma_{23} + \beta_{232}) t_2 t_3 \\ &+ (\delta_{231} + \beta_{2123} + \beta_{2312}) t_1 t_2 t_3 + \beta_{21} t_1 + \beta_{23} t_3 + (\beta_{213} + \beta_{231}) t_1 t_3\end{aligned}$$

$$\begin{aligned}\psi_3(t_1, t_2, t_3) = \alpha_3 t_3 + &(\gamma_{31} + \beta_{313}) t_1 t_3 + (\gamma_{32} + \beta_{323}) t_2 t_3 \\ &+ (\delta_{312} + \beta_{3231} + \beta_{3123}) t_1 t_2 t_3 + \beta_{31} t_1 + \beta_{32} t_2 + (\beta_{312} + \beta_{321}) t_1 t_2\end{aligned}$$

Each α_i is the information agent i is able to produce independent of the other agents. For a pair of agents i and j, the parameter γ_{ij} is the information agent i produces with agent j. The parameter δ_{ijk}, whose structure was detailed in chapter 2, represents the information agent i produces with both agents j and k in a way that is dependent on all their actions. The α_i depends only on agent i's action, the γ_{ij} depends on both agent i and j's actions, and the δ_{ijk} depends on the actions of all agents i, j, and k.

An interpretation of an externality of the type β_{ij} and β_{iji} in 2×2 collaborations has already been given. This interpretation carries over to $2 \times 2 \times 2$ collaborations. The first, β_{ij}, denotes the information agent i receives from agent j dependent only on t_j, and is a function of α_j. The second, β_{iji}, denotes the information agent i receives from agent j's collaboration with i, which is a function of γ_{ji}.

Emergent externality parameters in $2 \times 2 \times 2$ collaborations are β_{ijk} and β_{ijki}. The first, β_{ijk}, is the information agent i receives from agent j's interaction with agent k, which is a function of γ_{jk}. For example, agent 2's collaboration with agent 3 produces γ_{23} for agent 2, β_{123} for agent 1, and β_{323} for agent 3, all of which are coefficients of $t_2 t_3$. Agent 1 has no direct influence over this information that is dependent only on agent 2 and 3's actions. Lastly, β_{ijki} is the information agent i receives from agent j's interaction with both agents i and k. For example, agent 1, together with agents 2 and 3, generates δ_{123}, from which agent 2 then receives β_{2123} of this information, and agent 3 receives β_{3123}.

As before, let us examine the effects of perfect communication. This assumption results in

$$\alpha_1 = \beta_{21} = \beta_{31}, \qquad \alpha_2 = \beta_{12} = \beta_{32}, \qquad \alpha_3 = \beta_{13} = \beta_{23},$$

$$\gamma_{12} = \beta_{212} = \beta_{312}, \qquad \gamma_{13} = \beta_{313} = \beta_{213}, \qquad \gamma_{21} = \beta_{121} = \beta_{321}$$

$$\gamma_{23} = \beta_{323} = \beta_{123}, \qquad \gamma_{31} = \beta_{131} = \beta_{231}, \qquad \gamma_{32} = \beta_{232} = \beta_{132}$$

$$\delta_{123} = \beta_{2123} = \beta_{3123}, \qquad \delta_{231} = \beta_{1231} = \beta_{3231}, \qquad \delta_{312} = \beta_{1312} = \beta_{2312}$$

Because this makes all contribution functions equal, we write $\psi_1 = \psi_2 = \psi_3$ without the indices as ψ, as below,

$$\psi(t_1, t_2, t_3) = \alpha_1 t_1 + \alpha_2 t_2 + \alpha_3 t_3 + (\gamma_{12} + \gamma_{21}) t_1 t_2 + \\ (\gamma_{13} + \gamma_{31}) t_1 t_3 + (\gamma_{23} + \gamma_{32}) t_2 t_3 + (\delta_{123} + \delta_{231} + \delta_{312}) t_1 t_2 t_3$$

The possible Nash structures in the $2 \times 2 \times 2$ case are numerous. In this thesis we focus on a few situations that contain meaningful interpretations for our study of collaboration. These are pure coordination, division of labor, leadership, and the legislator game. Pure coordination is a situation of consensus. In a project, for example, the agents must decide on one, and only one, methodology. In an improvisation, the agents must agree on a particular style, key, or rhythm. On the other hand, in a collaboration it may be important for the agents to distribute themselves according to roles. In a scientific problem, one agent may be responsible for experimental work, and another on the theory. In an improvisation, the agents may adopt different harmonic and melodic roles. Our notion of division of labor captures these situations. Leadership is interpreted to mean those situations in which exactly one agent can lead. In a group project this could manifest as a group discussion where only one agent can speak at a time, while the others listen. In an improvisation, only one agent can solo while the others are responsible for holding the groove. This interpretation of leadership makes it a special case of division of labor.

Pure coordination involves all agents playing the same thing. Division of labor involves situations where the agents don't all play the same thing. Because we are considering three agents and two strategies, it is clear that it is impossible for all three agents to play different things. All of this is to say that division of labor, in many ways, is a generalization of anti-coordination but not pure anti-coordination, for which 3 strategies are needed. The last situation considered is the legislator game, which is between pure coordination and anti-coordination. In the below paragraph we review the legislator game briefly, and refer the reader to chapter 2 for more details.

In the legislator game, it is good for all agents to play a particular option, but not for all to play its alternative. Instead, it is good for exactly two of the agents to play the alternative. None of the legislators want to be the only one to vote for a raise, making not voting a good option for all. When exactly two of the legislators vote, the benefit occurs, so those who are voting would not deviate, and the one already not voting would also not deviate. In a collaboration this can be interpreted as a situation where there are limited resources. For example, in an improvisation, it is possible that two of the musicians engage in a "call and response" where the musicians involved riff off of each other's musical ideas, leaving the third musician to hold the groove.[14] In this case the pure strategy Nash equilibria involve profiles where exactly two musicians are engaged in call and response and the other holds the groove, in addition to the profile where all three musicians are maintaining the groove.

For pure coordination, we want to understand the region in the parameter space where the only two pure Nash equilibria are $(+1, +1, +1)$ and $(-1, -1, -1)$. For division of labor, in the $2 \times 2 \times 2$ collaboration, there are five cases. Here, essentially, we are curious about anti-coordination structures in $2 \times 2 \times 2$ collaborations. The following five sets of pure Nash

[14]There could very well be a call and response situation engaging more than two musicians, but for the purpose of this example, consider a case where only two can engage in the call and response.

equilibria are possible,

$$\{(+1,-1,-1),(-1,+1,-1),(-1,-1,+1)\}$$

$$\{(+1,-1,+1),(-1,+1,-1)\}$$

$$\{(-1,+1,+1),(+1,-1,-1)\}$$

$$\{(+1,+1,-1),(-1,-1,+1)\}$$

$$\{(+1,+1,-1),(+1,-1,+1),(-1,+1,+1)\}.$$

Finally, the legislator situation is one where it is either good for all to coordinate on a strategy, say $+1$, so that $(+1,+1,+1)$ is a Nash equilibrium, but it is also good for exactly two of them to coordinate on its alternative, -1, so that $(+1,-1,-1)$, $(-1,+1,-1)$, and $(-1,-1,+1)$ are Nash equilibria. Alternatively, the game could have equilibria $(-1,-1,-1)$, $(-1,+1,+1)$, $(+1,-1,+1)$, and $(+1,+1,-1)$.

The regions in the parameter space that give rise to pure coordination, division of labor, and the legislator game are described in Theorems 4.11 - 4.14 below, where in all cases take $\gamma^{12} = \gamma_{12} + \gamma_{21}, \gamma^{13} = \gamma_{13} + \gamma_{31}, \gamma^{23} = \gamma_{23} + \gamma_{32}$, and $\delta = \delta_{123} + \delta_{231} + \delta_{312}$ to simplify the exposition.

Theorem 4.11. *A $2 \times 2 \times 2$ collaboration with perfect communication has the form of a pure coordination game if and only if $|\alpha_i + \delta| < \gamma^{ij} + \gamma^{ik}$ for each distinct $i, j, k \in \{1, 2, 3\}$.*

Proof. Suppose $|\alpha_i + \delta| < \gamma^{ij} + \gamma^{ik}$ for each distinct $i, j, k \in \{1, 2, 3\}$. This is true if and only if the Nash entries in the strategy profiles $(+1,+1,+1)$ and $(-1,-1,-1)$ are positive for each agent. Hence these profiles are pure strategy Nash equilibria. Because all other strategy profiles are a unilateral deviation from $(+1,+1,+1)$ and $(-1,-1,-1)$, at least one Nash entry is negative. Therefore, no other strategy profile can be a pure strategy Nash equilibrium. □

Theorem 4.11 details when the collaboration has the form of a coordination game, and should seem familiar by this point in the thesis. The $2 \times 2 \times 2$ collaboration with perfect communication is a pure coordination game when each agent i's total collaborative information with the other agents j and k, $\gamma^{ij} + \gamma^{ik}$, dominates the other parameters, the individual preference α_i and δ. In this case, each agent has the incentive to generate information collaboratively rather than individually.

Theorem 4.12. *A $2 \times 2 \times 2$ collaboration with perfect communication has the form of a division of labor game with pure strategy Nash equilibria at the strategy profiles* $(+1, -1, -1)$, $(-1, +1, -1)$, *and* $(-1, -1, +1)$ *if and only if, for each* $i, j, k \in \{1, 2, 3\}, i \neq j, i \neq k, j \neq k$,

$$|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik}) \quad \text{and} \quad -\alpha_i + \delta > |\gamma^{ij} - \gamma^{ik}|$$

Alternatively, the collaboration has exactly three pure strategy Nash equilibria at $(+1, +1, -1)$, $(+1, -1, +1)$, *and* $(-1, -1, +1)$ *if and only if, for each* $i, j, k \in \{1, 2, 3\}, i \neq j, i \neq k, j \neq k$,

$$|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik}) \quad \text{and} \quad -\alpha_i + \delta < -|\gamma^{ij} - \gamma^{ik}|$$

Proof. The inequality $|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik})$ for each $i, j, k \in \{1, 2, 3\}, i \neq j, i \neq k, j \neq k$ is true if and only if all agents have negative Nash entries in the strategy profiles $(+1, +1, +1)$ and $(-1, -1, -1)$. An additional consequence is that each agent has a positive Nash entry in the strategy profile in which they are a unilateral deviation away from $(+1, +1, +1)$ and $(-1, -1, -1)$. In these remaining profiles there are exactly two agents coordinating. In one case, they are coordinating on $+1$, and in the other, on -1. It is easily verified that $-\alpha_i + \delta > |\gamma^{ij} - \gamma^{ik}|$ if and only if, in these remaining profiles, the agents have incentive to coordinate on $+1$, and when $-\alpha_i + \delta < -|\gamma^{ij} - \gamma^{ik}|$ if and only if the agents have incentive to coordinate on -1. This completes the proof. □

Theorem 4.12 above details when the collaboration has the form of an anti-coordination variant game with exactly 3 pure strategy Nash equilibria. In both cases, $|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik})$ for each $i, j, k \in \{1, 2, 3\}, i \neq j, i \neq k, j \neq k$. This, of course, entails $\gamma^{ij} + \gamma^{ik} < 0$, which means that the sum of collaborative information generated by agent i with each of the remaining agents has to be negative. This prevents profiles $(+1, +1, +1)$ and $(-1, -1, -1)$ from being equilibria, and gives a positive Nash entry to the agent who is a unilateral deviation away from pure coordination.

The second inequality in each case of Theorem 4.12 concern cases where for each agent i, the remaining agents j and k are mis-coordinating, and i must choose to coordinate with the agent playing $+1$ or with the agent playing -1. When this supports coordinating with the agent playing $+1$, the pure strategy Nash equilibria are $(+1, +1, -1)$, $(+1, -1, +1)$, and $(-1, -1, +1)$, and $-\alpha_i + \delta < -|\gamma^{ij} - \gamma^{ik}|$. On the other hand, when $-\alpha_i + \delta > |\gamma^{ij} - \gamma^{ik}|$, there is pressure to conform with the agent playing -1 in the profiles outside of pure coordination. Here, the pure strategy Nash equilibria are $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$.

Theorem 4.13. *A $2 \times 2 \times 2$ collaboration with perfect communication has the form of a division of labor game with two pure strategy Nash equilibria given by the strategy profiles where agents i and j coordinate with each other, and both anti-coordinate with agent k, if and only if $|\alpha_i - \delta| < \gamma^{ij} - \gamma^{ik}$, $|\alpha_j - \delta| < \gamma^{ij} - \gamma^{jk}$, and $|\alpha_k - \delta| < -\gamma^{ik} - \gamma^{jk}$, where the order of the superscripts of γ does not matter.*

Proof. First, notice that if the game has exactly two pure strategy Nash equilibria, as described in the statement of the theorem, with two agents coordinating with each other and mis-coordinating with the remaining agent, then there cannot be any more pure strategy Nash equilibria. The reason for this is similar to why, in the case of full coordination, as in Theorem 4.11, if $(+1, +1, +1)$ and $(-1, -1, -1)$ are pure strategy Nash equilibria, there cannot be any others. All strategy profiles besides the Nash equilibria are a unilateral deviation away from the Nash equilibria, and hence must have a negative Nash entry. Therefore, all that needs to be shown is that the strategy profiles claimed by the theorem to be pure strategy Nash equilibria are, in fact, Nash equilibria. This amounts to verifying that all Nash entries in these profiles are positive. For the profiles where agents i and j coordinate with each other, and both anti-coordinate with agent k, this is so if and only if $|\alpha_i - \delta| < \gamma^{ij} - \gamma^{ik}$, $|\alpha_j - \delta| < \gamma^{ij} - \gamma^{jk}$, and $|\alpha_k - \delta| < -\gamma^{ik} - \gamma^{jk}$. □

The above theorem describes when there are exactly two pure strategy Nash equilibria in a collaboration in the form of a division of labor game. There are three such sets of pure strategy Nash equilibria, namely $\{(+1, -1, +1), (-1, +1, -1)\}$, $\{(-1, +1, +1), (+1, -1, -1)\}$, and $\{(+1, +1, -1), (-1, -1, +1)\}$. In all of these cases, the same two agents are coordinating in both pure strategy Nash equilibria, and both are mis-coordinating with the remaining agent. According to Theorem 4.13, the agent who is mis-coordinating has their individual preference together with $-\delta$ less than, in magnitude, the sum of the negative collaborative information generated. For the other two agents, the same is true except the sum of the collaborative information includes the negative information generated with the mis-coordinating agent and the positive information generated with the agent with whom there is coordination.

Theorem 4.14. *A* $2\times2\times2$ *collaboration with perfect communication is a legislator game with pure strategy Nash equilibria at the strategy profiles* $(+1, +1, +1)$, $(+1, -1, -1)$, $(-1, +1, -1)$, *and* $(-1, -1, +1)$ *if and only if*

$$\alpha_i + \delta > |\gamma^{ij} + \gamma^{ik}| \quad \text{and} \quad -\alpha_i + \delta > |\gamma^{ij} - \gamma^{ik}|$$

Alternatively, the collaboration has only the four pure strategy Nash equilibria $(-1, -1, -1)$, $(-1, +1, +1)$, $(+1, -1, +1)$, *and* $(+1, +1, -1)$ *if and only if*

$$\alpha_i + \delta < -|\gamma^{ij} + \gamma^{ik}| \quad \text{and} \quad -\alpha_i + \delta < -|\gamma^{ij} - \gamma^{ik}|$$

Proof. It is a straightforward, though tedious computation to verify that the first inequality in case 1 is true if and only if $(+1, +1, +1)$ is a pure strategy Nash equilibrium, and that the first inequality in case 2 is true if and only if $(-1,-1,-1)$ is a pure strategy Nash equilibrium. Similarly, the second inequality is true in each case if and only if the corresponding division of labor strategy profiles are pure strategy Nash equilibria. □

The conditions in Theorem 4.14 above are almost the same as those in Theorem 4.12. The difference is that $\alpha_i + \delta > |\gamma^{ij} + \gamma^{ik}|$ in the first case, and $\alpha_i + \delta < -|\gamma^{ij} + \gamma^{ik}|$ in the second case, where in Theorem 4.12 these inequalities are, respectively, $|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik})$ and $|\alpha_i + \delta| < -(\gamma^{ij} + \gamma^{ik})$. This difference amounts to changing the signs of the Nash payoffs in the respective full coordination profile, $(+1,+1,+1)$ is the first case and $(-1,-1,-1)$ in the second case, from negative to positive.

We have the twelve parameters $\alpha_1, \alpha_2, \alpha_3, \gamma_{12}, \gamma_{21}, \gamma_{13}, \gamma_{31}, \gamma_{23}, \gamma_{32}, \delta_{123}, \delta_{231}$, and δ_{312}. A full analysis of the intricate relationships between these parameters, their effects on the collaboration, and their results on networks is beyond the scope of this thesis, and are reserved for future work. Going forward, make the simplifying assumption that $\alpha_1 = \alpha_2 = \alpha_3$, $\gamma_{12} = \gamma_{21} = \gamma_{13} = \gamma_{31} = \gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$. These are referred to as α, $\frac{\gamma}{2}$, and $\frac{\delta}{3}$, respectively. We use $\frac{\gamma}{2}$ because we want the sum of $\gamma_{12} + \gamma_{21}$, and of all other γ pairs, to be written simply as γ rather than 2γ. Similarly, $\frac{\delta}{3}$ is used.

In chapter 3 we gave a thorough analysis of asymmetric 2×2 potential games, a 3-dimensional space of games consisting of parameters α_1, α_2, and γ. In the $2 \times 2 \times 2$ case, our assumption of symmetry simplifies the space to 3-dimensions as well, namely the coordinates α, $\frac{\gamma}{2}$, and $\frac{\delta}{3}$. Note that although both cases have dimension 3, they are not equal representations.

This is because the $2 \times 2 \times 2$ case includes δ, which is an emergent parameter that, because of its orthogonality with all α and γ terms, cannot be expressed as a linear combination of the α and γ terms. Had the two 3-dimensional spaces been equal would have implied that our simplifying assumption of symmetry in the 3-agent case would be reducible to smaller 2×2 interactions, and hence would ignore anything specific and emergent that may happen for 3 agents. For example, it would not be possible to represent collaborations in the form of the legislator game.

The previous theorems are restated below, where the interpretations are omitted since they are almost identical, except with the assumption of symmetry, to those for Theorems 4.11 - 4.14. The proof of every theorem below amounts to setting $\alpha_1 = \alpha_2 = \alpha_3$, $\gamma_{12} = \gamma_{21} = \gamma_{13} = \gamma_{31} = \gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$, in the corresponding Theorem 4.11 - 4.14. Because of this, the proofs are omitted.

Theorem 4.15. *A symmetric $2 \times 2 \times 2$ collaboration with perfect communication has the form of a pure coordination game if and only if $2\gamma > |\alpha + \delta|$.*

Theorem 4.16. *A symmetric $2\times2\times2$ collaboration with perfect communication has the form of a division of labor with pure strategy Nash equilibria at the strategy profiles $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$ if and only if $|\alpha + \delta| < -2\gamma$ and $-\alpha + \delta > 0$. Alternatively, the collaboration has pure strategy Nash equilibria at $(+1, +1, -1)$, $(+1, -1, +1)$, and $(-1, -1, +1)$ if and only if $|\alpha + \delta| < -2\gamma$ and $-\alpha + \delta < 0$.*

Theorem 4.17. *A symmetric $2 \times 2 \times 2$ collaboration with perfect communication cannot have the form of a division of labor game with two pure strategy Nash equilibria where two agents coordinate with each other and both anti-coordinate with the third.*

Theorem 4.18. *A symmetric $2\times2\times2$ collaboration with perfect communication has the form of a legislator game with pure strategy Nash equilibria at the strategy profiles $(+1, +1, +1)$, $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$ if and only if $\alpha + \delta > |2\gamma|$ and $-\alpha + \delta > 0$.*

All theorems above, except Theorem 4.17, are natural simplifications of Theorems 4.11 - 4.14. The impossibility of a division of labor with two pure strategy Nash equilibria is a direct consequence of symmetry. Under symmetry, all agents are assumed to be identical. Because of this, if $(+1, +1, -1)$ is a pure strategy Nash equilibrium, the strategy profiles $(+1, -1, +1)$

and $(-1,+1,+1)$ must be equilibria, too. Hence it is impossible, with symmetry, to have the only two pure strategy Nash equilibria be $(+1,+1,-1)$ and $(-1,-1,+1)$.

The following section elaborates on the effects of networks on 2×2 and $2\times2\times2$ collaborations, through slightly different perspectives in each case. For 2×2 collaborations a preliminary analysis is provided for the case of symmetry, where the notion of the audience is introduced and the discussion pertains to the optimization of information received by the audience. For $2\times2\times2$ collaborations, symmetry is also assumed, and the networks of concern are the full, star, and wheel networks. There the results do not include an audience, and the analysis is focused on the effects of networks on various $2\times2\times2$ Nash equilibrium structures.

4.7 Collaboration on Networks

4.7.1 2×2

Up to isomorphism, there are three directed networks on two vertices. The complete network, the one-way network, and the empty network. The complete network is the standard case of a 2×2 collaboration. On the one-way network, one agent receives information from the other, who receives nothing in return. The empty network is the trivial case where there is no interaction between the agents.

We remind the reader of the contribution functions with perfect communication on the full network, written below. On the complete network, with perfect information, both agents receive all of the information that is generated.

$$\psi_1^{K_2}(t_1,t_2)=\alpha_1 t_1+(\gamma_{12}+\gamma_{21})t_1t_2+\alpha_2 t_2 \tag{4.9}$$

$$\psi_2^{K_2}(t_1,t_2)=\alpha_2 t_2+(\gamma_{12}+\gamma_{21})t_1t_2+\alpha_1 t_1 \tag{4.10}$$

Now, suppose that agent 1 receives information from agent 2, but that agent 2 does not receive information from agent 1. This gives rise to a one-way network, and the contribution functions here are

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 \tag{4.11}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 \tag{4.12}$$

The effect of this network on agent 2's contribution function is immediate. Since agent 2 does not interact with and does not receive information from agent 1, the terms γ_{21}, β_{212}, and β_{21} are not included in agent 2's contribution function. Agent 1's contribution function is almost the same as in the full network, except for the missing externality term, β_{121}. This is because agent 2 is not generating any information from their interaction with agent 1, so this information cannot reach agent 1.

Assuming perfect communication the contribution functions become

$$\psi_1(t_1, t_2) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \alpha_2 t_2 \tag{4.13}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 \tag{4.14}$$

Lastly, on the empty network the contribution functions are

$$\psi_1(t_1, t_2) = \alpha_1 t_1 \tag{4.15}$$

$$\psi_2(t_1, t_2) = \alpha_2 t_2 \tag{4.16}$$

This case is trivial; no interdependent terms are included because the agents are not interacting. Moreover, the effects of perfect communication are nonexistent since there is no interaction between the agents and hence no possibility for communication.

As a starting point of analysis for 2×2 collaborations on networks, we use the simplifying assumption of symmetry, where $\alpha_1 = \alpha_2$, which is denoted by α, and $\gamma_{12} = \gamma_{21}$, denoted by $\frac{\gamma}{2}$. With this, the sum $\gamma_{12} + \gamma_{21}$ in each agent's contribution function simplifies to γ.

The question we will now answer concerns which networks maximize the information produced in the collaboration. On a full network, the information is produced from a bidirectional interaction, as is the case in a standard collaboration. On the one-way network, the interaction is asymmetric– only one agent interacts with the information received. This reflects situations where one agent writes a book or a paper, for example, and the other agent uses this to produce additional information. In terms of music, this is an instance of "overdubbing."[15] Lastly, the empty network means there is no interaction between the agents. This could be two scientists pursuing a problem alone, or two musicians creating solo-work.

To proceed it is important to define an external receiver of the information. To do this, define a third agent in the collaboration, the audience, who receives the information produced. This agent has no strategies, and instead receives only the externalities produced in the collaboration. In a scientific collaboration the audience can be interpreted to be those in the scientific community who read the books and papers produced by agents 1 and 2. In the production of music, the audience can be an actual audience in a live performance, or those who listen to the recorded music.

The audience has contribution function $\psi_A(t_1, t_2) = \beta_{A1}t_1 + \beta_{A2}t_2 + (\beta_{A12} + \beta_{A21})t_1t_2$, where β_{A1} is a function of α_1, the information produced by agent 1, β_{A2} is a function of α_2, the information produced by agent 2, and β_{A12} and β_{A21} are, respectively, the information agent 1 produces with agent 2, and the information agent 2 produces with agent 1.

The assumption of perfect communication between the information generated in the collabo-

[15] Overdubbing is described briefly in the introduction, chapter 1

ration and the audience, although idealized, is not unrealistic. In the case of scientific work, for example, the audience is assumed be comprised of those literate in the jargon, methods, and pertinent questions involved in the collaboration. This way, whatever information is generated by those collaborating is "perfectly" communicated to the audience. Similarly, in the case of music, the audience is assumed to be those who understand and appreciate the musical ideas created and expressed.

Assuming perfect communication between all agents, including the audience, yields identical contribution functions for the three agents on the full network. For the audience, perfect communication implies $\beta_{A1} = \alpha_1$, $\beta_{A2} = \alpha_2$, $\beta_{A12} = \gamma_{12}$, and $\beta_{A21} = \gamma_{21}$. The contribution functions are $\psi_i(t_1, t_2) = \alpha t_1 + \alpha t_2 + \gamma t_1 t_2$ for $i = 1, 2, A$.

On the wheel network where agent 1 interacts with the information generated by agent 2, the audience receives the information agent 1 produces individually and from their interaction with agent 2's information, and the information generated by agent 2. Here, the collaboration functions are

$$\psi_1(t_1, t_2) = \alpha t_1 + \alpha t_2 + \frac{\gamma}{2} t_1 t_2 \tag{4.17}$$

$$\psi_2(t_1, t_2) = \alpha t_2 \tag{4.18}$$

$$\psi_A(t_1, t_2) = \alpha t_1 + \alpha t_2 + \frac{\gamma}{2} t_1 t_2 \tag{4.19}$$

Finally, on the empty network, the audience receives two independent quantities of information. This represents, for example, two scientists working independently and generating distinct results. The contribution functions in this case are

$$\psi_1(t_1, t_2) = \alpha t_1 \tag{4.20}$$

$$\psi_2(t_1, t_2) = \alpha t_2 \tag{4.21}$$

$$\psi_A(t_1, t_2) = \alpha t_1 + \alpha t_2 \tag{4.22}$$

Let us pause here to note that in this thesis the possibility of mutual information is ignored. For example, if the interpretation is of two scientists who are working on a problem, then the result of the empty network is that the scientists work on the problem without interacting, and the audience receives the information generated by each scientist alone. It is possible that, despite no interaction, the scientists generate some of the same information, in which case the total information generated would be less than the sum of the individual quantities. For simplicity, this possibility is ignored in what follows.

The question of which networks maximize the information produced in the collaboration can be answered by looking at the contribution functions of the audience. Which regions in the parameter space together with which networks maximize the information received by the audience? Going forward, refer to the audience's function on the full network as ψ_A^F, on the one-way network as ψ_A^{OW}, and on the empty network as ψ_A^E.

There are four important cases to consider in terms of pure strategy Nash equilibrium structures in the symmetric case. Namely, these are

1. $|\alpha| > \gamma > 0$
2. $|\alpha| > -\gamma > 0$
3. $\gamma > |\alpha| > 0$
4. $-\gamma > |\alpha| > 0$

The networks that maximize the information received by the audience in each of the four cases above are described in the following theorem.

Theorem 4.19. *The full network maximizes the information received by the audience in cases 1, 3, and 4. For case 2, the empty network is optimal if and only if* $2|\alpha| > -\gamma$*, and the full network is optimal if and only if* $2|\alpha| < -\gamma$*.*

Proof. For case 1, because the magnitude of α is greater than the value of the positive γ, the strategy profile $(\operatorname{sgn}\alpha, \operatorname{sgn}\alpha)$ is a unique pure strategy Nash equilibrium. The fact that $\gamma > 0$ means that the agents both generate and communicate positive information when they coordinate. Without loss of generality, suppose $\alpha > 0$ so that the unique pure strategy Nash equilibrium is $(+1, +1)$. Then we have $\psi_A^C(+1, +1) = 2\alpha + \gamma$, $\psi_A^{OW}(+1, +1) = 2\alpha + \frac{\gamma}{2}$, and $\psi_A^E(+1, +1) = 2\alpha$. Since $\gamma > 0$, it is immediate that $\psi_A^C(+1, +1) > \psi_A^{OW}(+1, +1)$ and $\psi_A^C(+1, +1) > \psi_A^E(+1, +1)$. Hence, collaborations in the region of the parameter space described by case 1 are optimized on the full network.

Case 2 is similar to case 1 in the sense that the strategy profile $(\operatorname{sgn}\alpha, \operatorname{sgn}\alpha)$ is a unique pure strategy Nash equilibrium. The difference is that in this case $\gamma < 0$. Consequently, the agents play their preference but generate less information than they would have on their own. In this case, the information produced is optimized on the empty network. This is because, due to $\gamma < 0$, the audience's contribution function on the full and one-way networks have information taken away from what the individuals produce on their own. To demonstrate, suppose, without loss of generality, that $\alpha > 0$. The audience's contribution functions are $\psi_A^C(+1, +1) = 2\alpha + \gamma$, $\psi_A^{OW}(+1, +1) = 2\alpha + \frac{\gamma}{2}$, and $\psi_A^E(+1, +1) = 2\alpha$. Since $\gamma < 0$, the maximum must be $\psi_A^E(+1, +1)$. Consequently, collaborations with parameters described by case 2 are optimized on the empty network.

The pure strategy Nash equilibrium structure in case 3 is that of a coordination game with pure equilibria $(+1, +1)$ and $(-1, -1)$. Here, the sign of α determines which equilibrium is risk-dominant; $(+1,+1)$ when $\alpha > 0$, and $(-1,-1)$ when $\alpha < 0$. Using risk-dominance as an equilibrium refinement tool entails that the agents play the profile $(\operatorname{sgn}\alpha, \operatorname{sgn}\alpha)$. In this profile the α's and γ all contribute positive information. The same procedure as for the proof for case 1 can be used here, resulting in the full network being optimal.

Lastly, in case 4, the pure strategy Nash equilibria are $(+1,-1)$ and $(-1,+1)$. Suppose, without loss of generality, that $\alpha > 0$.Now, note that these equilibria do not survive the empty network. On the empty network the contribution functions lose all interdependent terms, like γ. Hence, on the empty network the agents both play $\operatorname{sgn}\alpha = +1$. The audience's contribution function in this case is $\psi_A^E(+1,+1) = 2\alpha$. On the one-way network, the agent receiving no information plays $\operatorname{sgn}\alpha = +1$. Although $-\gamma > |\alpha|$, it may be the case that $|\alpha| > -\frac{\gamma}{2}$. When this is so, the other agent also plays $\operatorname{sgn}\alpha = +1$, so that the unique pure strategy Nash equilibrium is $(+1,+1)$ and $\psi_A^{OW}(+1,+1) = 2\alpha + \frac{\gamma}{2}$. On the other hand, if $-\frac{\gamma}{2} > |\alpha|$, then, while the disconnected agent plays $\operatorname{sgn}\alpha = +1$, the other agent has pressures to mis-coordinate, hence they play $-\operatorname{sgn}\alpha = -1$. Here, supposing, without loss of generality, that the disconnected agent is agent 2, gives contribution function $\psi_A^{OW}(-1,+1) = -\frac{\gamma}{2}$. Finally, on the full network, the contribution function has the same value in both pure strategy Nash equilibria of the game, $(+1,-1)$ and $(-1,+1)$. This value is $\psi_A^F(+1,-1) = -\gamma$.

To summarize, the possible values of the audience's contribution function in case 4, with the assumption that $\alpha > 0$, are $\psi_A^E(+1,+1) = 2\alpha$, $\psi_A^{OW}(+1,+1) = 2\alpha + \frac{\gamma}{2}$ when $|\alpha| > -\frac{\gamma}{2}$ and $\psi_A^{OW}(-1,+1) = -\frac{\gamma}{2}$ when $|\alpha| < -\frac{\gamma}{2}$, and $\psi_A^F(+1,-1) = -\gamma$. From this it immediately follows that the empty network is optimal if and only if $2|\alpha| > -\gamma$, and that the full network is optimal if and only if $2|\alpha| < -\gamma$. $\square$

One immediate consequence of Theorem 4.19 is that, in all cases but case 2, when $2|\alpha| > -\gamma$, the full network maximizes production. When $2|\alpha| > -\gamma$ in case 2, the empty network is optimal because the agents are both inclined to play their preference, which is the same due to symmetry, and, when collaborating, playing the same preference generates negative information because of $\gamma < 0$. Case 1 and 3 both involve the agents playing the same preference, which is supported by $\gamma > 0$, thereby making the collaboration fruitful. Case 4 is interesting because the agents must mis-coordinate, but in doing so cannot both play the same preference. Because mis-coordinating produces enough information to override this fact, it is better for the agents to collaborate and mis-coordinate, than to produce only α on the empty network.

Another consequence is that, with the assumption of symmetry, the one-way network does not maximize the information produced in any of the cases. This makes sense because the one-way network implies asymmetry, which cannot happen when symmetry is assumed between the parameters.

An exhaustive analysis of the complete parameter space, without the assumption of symmetry, will better illuminate the intricacies involved, and demonstrate those situations wherein the one-way network optimizes the information produced. Alas, this analysis is in progress and is omitted from the thesis. The intention of this section is to introduce the notion of the audience as the receiver of information and to initiate this line of research. Additional future work involves extending the notion of the audience to the case of 3 agents, and synthesizing it with the results offered in the next subsection, where the effects of networks are studied on some of the Nash equilibrium structures possible in a $2 \times 2 \times 2$ collaboration.

4.7.2 $2 \times 2 \times 2$

Up to isomorphism, there are 16 directed networks on 3 vertices. Focus is given only to the complete network, the wheel network, and the star network. The complete network is a standard $2 \times 2 \times 2$ collaboration, which was discussed earlier in the chapter. Consider the wheel and star networks in Figure 4.14 below.

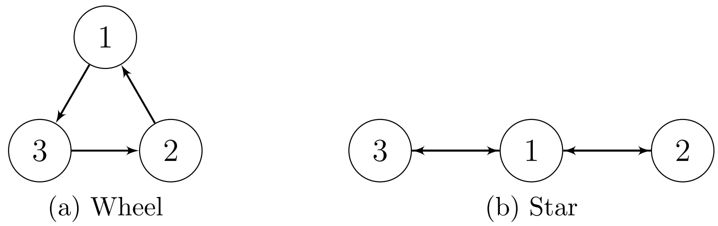


Figure 4.14: Wheel and Star Networks

The wheel network in Figure 4.14 has contribution functions

$$\psi_1(t_1,t_2,t_3) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \beta_{12} t_2 + \beta_{123} t_2 t_3$$
$$\psi_2(t_1,t_2,t_3) = \alpha_2 t_2 + \gamma_{23} t_2 t_3 + \beta_{23} t_3 + \beta_{231} t_1 t_3$$
$$\psi_3(t_1,t_2,t_3) = \alpha_3 t_3 + \gamma_{31} t_1 t_3 + \beta_{31} t_1 + \beta_{312} t_1 t_2$$

Agent 1 produces information α_1 independently of agent 2 and information γ_{12} with agent 2's incoming stream of information. The externality terms received by agent 1 are β_{12} and β_{123}. The first term, β_{12} is the information agent 2 produces independently of the other agents. The second term, β_{123} is the information agent 2 produces with agent 3. Although agent 1 is not connected to agent 3, because they receive the information that agent 2 produces, any information that agent 2 produces for their interaction with agent 3 gets passed to agent 1. The contribution functions of agents 2 and 3 are similar, with the proper indices replaced.

Under perfect communication these become

$$\psi_1(t_1,t_2,t_3) = \alpha_1 t_1 + \gamma_{12} t_1 t_2 + \alpha_2 t_2 + \gamma_{23} t_2 t_3$$
$$\psi_2(t_1,t_2,t_3) = \alpha_2 t_2 + \gamma_{23} t_2 t_3 + \alpha_3 t_3 + \gamma_{31} t_1 t_3$$
$$\psi_3(t_1,t_2,t_3) = \alpha_3 t_3 + \gamma_{31} t_1 t_3 + \alpha_1 t_1 + \gamma_{12} t_1 t_2$$

The star network in Figure 4.14 has contribution functions

$$\psi_1(t_1,t_2,t_3) = \alpha_1 t_1 + (\gamma_{12}+\beta_{121})t_1t_2 + (\gamma_{13}+\beta_{131})t_1t_3 + \delta_{123}t_1t_2t_3 + \beta_{12}t_2 + \beta_{13}t_3$$

$$\psi_2(t_1,t_2,t_3) = \alpha_2 t_2 + (\gamma_{21}+\beta_{212})t_1t_2 + \beta_{2123}t_1t_2t_3 + \beta_{21}t_1 + \beta_{213}t_1t_3$$

$$\psi_3(t_1,t_2,t_3) = \alpha_3 t_3 + (\gamma_{31}+\beta_{313})t_1t_3 + \beta_{3123}t_1t_2t_3 + \beta_{31}t_1 + \beta_{312}t_1t_2$$

The agent in the center, agent 1, has almost the same contribution function as in the fully connected network. What is missing are the β_{123}, β_{132}, β_{1231}, and β_{1312} terms. Because agents 2 and 3 are not connected, they do not produce information with each other, hence β_{123} and β_{132} are missing. In addition, because they are not connected, they do not produce information with the entirety of the group, which explains the missing β_{1231} and β_{1312} terms in agent 1's contribution function.

The contribution functions for agents 2 and 3 are almost as they would be if these agents were engaged in a 2×2 collaboration with agent 1, with the additional traces of the three agent reality that is communicated through agent 1. Both agents 2 and 3 receive information that agent 1 produces with the other agent; agent 2 receives β_{213}, the information agent 1 generates with agent 3, and agent 3 receives β_{312}, the information that agent 1 generates with agent 2. In addition, both agents receive information that is dependent on all of their strategies, namely β_{2123} for agent 2, and β_{3123} for agent 3. Although agents 2 and 3 are not connected to the whole group, agent 1 is, and so the information agent 1 generates from this holistic interaction gets relayed to agents 2 and 3.

Under perfect communication these become

$$\psi_1(t_1,t_2,t_3) = \alpha_1 t_1 + (\gamma_{12}+\gamma_{21})t_1t_2 + (\gamma_{13}+\gamma_{31})t_1t_3 + \delta_{123}t_1t_2t_3 + \alpha_2 t_2 + \alpha_3 t_3$$

$$\psi_2(t_1,t_2,t_3) = \alpha_2 t_2 + (\gamma_{21}+\gamma_{12})t_1t_2 + \delta_{123}t_1t_2t_3 + \alpha_1 t_1 + \gamma_{13}t_1t_3$$

$$\psi_3(t_1,t_2,t_3) = \alpha_3 t_3 + (\gamma_{31}+\gamma_{13})t_1t_3 + \delta_{123}t_1t_2t_3 + \alpha_1 t_1 + \gamma_{12}t_1t_2$$

Assuming symmetry gives, for the wheel network,

$$\psi_1(t_1,t_2,t_3) = \alpha t_1 + \frac{\gamma}{2}t_1t_2 + \alpha t_2 + \frac{\gamma}{2}t_2t_3 \tag{4.23}$$

$$\psi_2(t_1,t_2,t_3) = \alpha t_2 + \frac{\gamma}{2}t_2t_3 + \alpha t_3 + \frac{\gamma}{2}t_1t_3 \tag{4.24}$$

$$\psi_3(t_1,t_2,t_3) = \alpha t_3 + \frac{\gamma}{2}t_1t_3 + \alpha t_1 + \frac{\gamma}{2}t_1t_2 \tag{4.25}$$

For the star network,

$$\psi_1(t_1,t_2,t_3) = \alpha t_1 + \gamma t_1t_2 + \gamma t_1t_3 + \frac{\delta}{3}t_1t_2t_3 + \alpha t_2 + \alpha t_3 \tag{4.26}$$

$$\psi_2(t_1,t_2,t_3) = \alpha t_2 + \gamma t_1t_2 + \frac{\delta}{3}t_1t_2t_3 + \alpha t_1 + \frac{\gamma}{2}t_1t_3 \tag{4.27}$$

$$\psi_3(t_1,t_2,t_3) = \alpha t_3 + \gamma t_1t_3 + \frac{\delta}{3}t_1t_2t_3 + \alpha t_1 + \frac{\gamma}{2}t_1t_2 \tag{4.28}$$

The Nash terms of the contribution function for agent i are those that come with t_i. This is because the Nash terms must produce changes when agent i deviates unilaterally. In the contribution functions of the star network given above, although agents 2 and 3 are disconnected from each other, from agent 1's transmission of $\delta t_1t_2t_3$, this 3-agent interaction is included in the Nash terms of agents 2 and 3's contribution functions. An immediate and important implication of this is that the influence of δ reaches all agents in the star network, hinting that interactions dependent on all three agents are possible on the star network. For the wheel network, this is not so. Since no agent is interacting with all agents, the information $\frac{\delta}{3}$ is never generated and hence never transmitted.

Let us make all of this precise by investigating the situations of interest– coordination, division of labor, and the legislator game– on the wheel and star networks. These results are summarized in Theorems 4.20, 4.21, and 4.22 below. For most proofs, without loss of generality, the contribution functions are written in terms of the wheel and star network in Figure 4.14. A short discussion is provided for each theorem, but major points are reserved for after the theorems, where the results can be visualized in Figures 4.15, 4.16, and 4.17.

Theorem 4.20. *A $2 \times 2 \times 2$ symmetric identical play collaboration has the form of a pure coordination game*

- *on the full network if and only if $|\alpha + \delta| < 2\gamma$*
- *on the wheel network if and only if $|\alpha| < \frac{\gamma}{2}$*
- *on the star network if and only if $|\alpha + \frac{\delta}{3}| < \gamma$*

Proof. The proof for the full network follows by taking $\alpha_1 = \alpha_2 = \alpha_3$, $\gamma_{12} = \gamma_{21} = \gamma_{13} = \gamma_{31} = \gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$ in Theorem 4.11.

For the wheel network, consider agent i's contribution function as in (4.23) - (4.25), namely $\psi_i(t_1, t_2, t_3) = \alpha t_i + \frac{\gamma}{2} t_i t_j + \alpha t_j + \frac{\gamma}{j} t_j t_j$ where agent i receives information from agent j, and j from k. The Nash component of i's contribution function is $\psi_i^N(t_1, t_2, t_3) = \alpha t_i + \frac{\gamma}{2} t_i t_j$ because these are the only terms that are affected by agent i's unilateral deviations. The Nash function is positive in $(+1, +1, +1)$ and $(-1, -1, -1)$, and hence Nash equilibria, if and only if $\frac{\gamma}{2} > |\alpha_i|$. Since i is arbitrary, this holds for $i = 1, 2, 3$. All other strategy profiles are a unilateral deviation away from $(+1, +1, +1)$ or $(-1, -1, -1)$ and must have at least one negative Nash term. Consequently, it cannot be an equilibrium. Therefore, $(+1, +1, +1)$ and $(-1, -1, -1)$ are the only pure strategy Nash equilibria if and only if $\frac{\gamma}{2} > |\alpha_i|$ for $i = 1, 2, 3$.

For the star network, similar computations to the above give, for the center agent $|\alpha+\frac{\delta}{3}| < 2\gamma$, and for the agents at the extremes $|\alpha+\frac{\delta}{3}| < \gamma$. The second inequality implies the first, making the first inequality redundant. Hence, the collaboration on the star network has the form of a coordination game if and only if $|\alpha + \frac{\delta}{3}| < \gamma$. □

Theorem 4.20 shows that there are regions in the parameter space that give rise to a global coordination game in all three networks of interest.

Theorem 4.21. *A* $2\times2\times2$ *symmetric identical play collaboration has the form of a division of labor game with pure strategy Nash equilibria* $(+1,-1,-1)$, $(-1,+1,-1)$, *and* $(-1,-1,+1)$

- *on the full network if and only if* $|\alpha+\delta| < -2\gamma$ *and* $-\alpha+\delta > 0$
- *impossible on the wheel network*
- *on the star network if and only if* $|\alpha+\frac{\delta}{3}| < -\gamma$ *and* $-\alpha+\frac{\delta}{3} > |\gamma|$

Alternatively, it has pure Nash equilibria $(+1,+1,-1)$, $(+1,-1,+1)$, *and* $(-1,+1,+1)$

- *on the full network if and only if* $|\alpha+\delta| < -2\gamma$ *and* $-\alpha+\delta < 0$
- *impossible on the wheel network*
- *on the star network if and only if* $|\alpha+\frac{\delta}{3}| < -\gamma$ *and* $\alpha-\frac{\delta}{3} > |\gamma|$

Proof. The proof for the full network follows by taking $\alpha_1 = \alpha_2 = \alpha_3$, $\gamma_{12} = \gamma_{21} = \gamma_{13} = \gamma_{31} = \gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$ in Theorem 4.12.

For the wheel network, agent 1's Nash terms in the profiles $(+1,+1,-1)$ and $(-1,+1,+1)$ of case 1 are $\psi_1^N(+1,+1,-1) = \alpha+\frac{\gamma}{2}$ and $\psi_1^N(-1,+1,+1) = -\alpha-\frac{\gamma}{2}$. Because it is impossible for both of these to be positive, it is impossible to have the desired equilibria in the wheel network. A similar demonstration is easily constructed for case 2 of the theorem.

On the star network, both cases of the theorem require that the strategy profiles $(+1,+1,+1)$ and $(-1,-1,-1)$ have negative Nash terms for all agents. For the center agent this is true if and only if $|\alpha+\frac{\delta}{3}| < -2\gamma$, and for the agents at the extremes if and only if $|\alpha+\frac{\delta}{3}| < -\gamma$. Because the second inequality implies the first, both conditions being true can expressed just with $|\alpha+\frac{\delta}{3}| < -\gamma$. Then, so far, all other profiles where exactly two agents are coordinating have one positive Nash term– the agent who is mis-coordinating.

The first case of the theorem is where the two agents coordinate on -1, and the second case is where they coordinate on $+1$. No matter what, there will be mis-coordination with exactly one agent. For the first case, the center agent chooses to coordinate on -1 if and only if $-\alpha+\frac{\delta}{3} > 0$ and the agents at the extremes do the same if and only if $-\alpha+\gamma+\frac{\delta}{3}$ and $-\alpha-\gamma+\frac{\delta}{3}$, which together imply $-\alpha+\frac{\delta}{3} > |\gamma|$. When the inequality is true for the agents at the extremes, it is automatically true for the center agent. Hence, for case 1 of the theorem, $|\alpha+\frac{\delta}{3}| < -\gamma$ and $-\alpha+\frac{\delta}{3} > |\gamma|$ if and only if the game has the pure strategy Nash equilibria $(+1,-1,-1)$, $(-1,+1,-1)$, and $(-1,-1,+1)$. Case 2 follows from similar computations. □

Theorem 4.21 shows that it is impossible to have the division of labor game on the wheel network. Intuitively, this makes sense; the Nash structure of the division of labor game is dependent on δ, the emergent Nash parameter in $2 \times 2 \times 2$ games. Because no agent in the wheel network is playing with two other agents, this parameter cannot be in any agent's contribution function. The same reasoning holds for Theorem 4.22 below, and the impossibility of this Nash structure on the wheel network. On the star network, the proof of Theorem 4.21 reveals that the inequalities required for the agents at the extremes are stronger than those for the center agent, for whom the Nash terms are indistinguishable from those of the full network. A consequence of this is that there are regions in the parameter space where the center agent is bound to the structure of division of labor, but the agents at the extreme are not.

Theorem 4.22. *A $2 \times 2 \times 2$ symmetric identical play collaboration has the form of a legislator game with pure strategy Nash equilibria $(+1, +1, +1)$, $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$*

- *on the full network if and only if $\alpha + \delta > 2|\gamma|$ and $-\alpha + \delta > 0$*
- *impossible on the wheel network*
- *on the star network if and only if $\alpha + \frac{\delta}{3} > 2|\gamma|$ and $-\alpha + \frac{\delta}{3} > |\gamma|$*

Alternatively, it has pure Nash equilibria $(-1, -1, -1)$, $(+1, +1, -1)$, $(+1, -1, +1)$, and $(-1, +1, +1)$

- *on the full network if and only if $\alpha + \delta < -2|\gamma|$ and $\alpha - \delta > 0$*
- *impossible on the wheel network*
- *on the star network if and only if $\alpha + \frac{\delta}{3} < -2|\gamma|$ and $\alpha - \frac{\delta}{3} > |\gamma|$*

Proof. The proof for the full network follows by taking $\alpha_1 = \alpha_2 = \alpha_3$, $\gamma_{12} = \gamma_{21} = \gamma_{13} = \gamma_{31} = \gamma_{23} = \gamma_{32}$, and $\delta_{123} = \delta_{231} = \delta_{312}$ in Theorem 4.14.

The proof of the impossibility of the legislator structure on the wheel network is the same as the proof for the wheel network in Theorem 4.21.

For the star network, we can borrow the second half of the star network's proof for Theorem 4.21, concluding with the inequality $-\alpha + \frac{\delta}{3} > |\gamma|$ for case 1 of the theorem. This inequality is true if and only if strategy profiles $(+1, -1, -1)$, $(-1, +1, -1)$, and $(-1, -1, +1)$ are pure strategy Nash equilibria. For the legislator structure described by case 1, the additional equilibrium $(+1, +1, +1)$ is needed. The Nash terms in this profile are positive for all agents if and only if $\alpha + \frac{\delta}{3} > -2\gamma$ and $\alpha + \frac{\delta}{3} > -\gamma$. Although the Nash terms being negative in the profile $(-1,-1,-1)$ follows from $-\alpha + \frac{\delta}{3} > |\gamma|$, looking closer reveals that $\alpha + \frac{\delta}{3} > 2\gamma$ and $\alpha + \frac{\delta}{3} > \gamma$. Hence, $\frac{\delta}{3} > 2|\gamma|$ and $\alpha + \frac{\delta}{3} > |\gamma|$. It is clear that the first inequality implies the second. Therefore, the collaboration on the star network has the form of the legislator game in case 1 of the theorem if and only if $-\alpha + \frac{\delta}{3} > |\gamma|$ and $\frac{\delta}{3} > 2|\gamma|$. The proof of case 2 follows the same demonstrations. □

The major difference between the division of labor in Theorem 4.21 and the legislator structure in Theorem 4.22, is that one of the full coordination profiles is also an equilibrium in the legislator structure. This changes the inequality $|\alpha + \frac{\delta}{3}| < -\gamma$ of case 1 of Theorem 4.21 to the inequality $\alpha + \frac{\delta}{3} > 2|\gamma|$ of case 1 of Theorem 4.22. An immediate insight from this is that the relationship between α and $\frac{\delta}{3}$ has to change drastically, but stay within the confines of $-\alpha + \frac{\delta}{3} > |\gamma|$ (the common inequality in both theorems), for the collaboration to transition from having the form of a division of labor game to the form of a legislator game.

To better visualize the results of Theorems 4.20, 4.21, and 4.22, let us, without loss of generality, take $\alpha = 1$ so the regions can be plotted on the $\gamma\delta$-plane. Write σ as the permutation operator, where σ applied to a strategy profile represents the set containing all permutations of the profile. For example,

$$\sigma((+1,-1,-1)) = \{(+1,-1,-1),(-1,+1,-1),(-1,-1,+1)\}.$$

As stated in the theorems, the division of labor and legislator equilibria do not survive the wheel network. The plots also show that there are regions that induce full coordination on each particular network but not the others. In addition, there are regions that induce full coordination on the full network and star but not the wheel. Similarly, there are regions that induce coordination on the wheel network and the star, but not the full network. Finally, there are regions that induce coordination on all three networks.

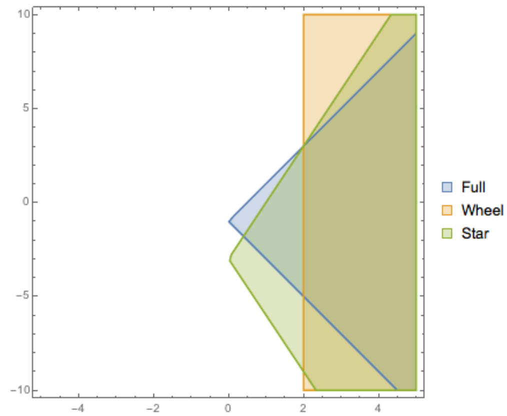
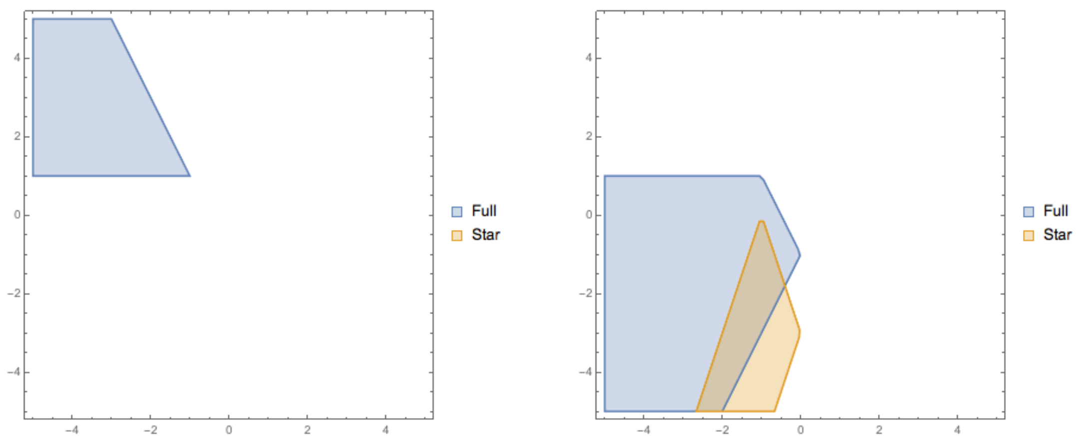
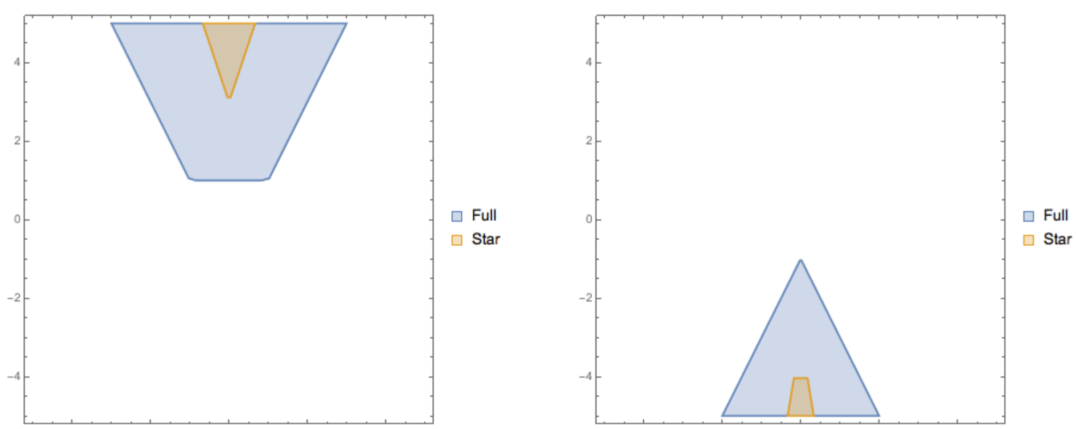


Figure 4.15: Full Coordination



(a) $\sigma((+1, -1, -1))$ (b) $\sigma((+1, +1, -1))$

Figure 4.16: Division of Labor



(a) $(+1, +1, +1), \sigma((+1, -1, -1))$ (b) $(-1, -1, -1), \sigma((+1, +1, -1))$

Figure 4.17: Legislator

For division of labor, when $\alpha > 0$ (because we are taking $\alpha = 1$), only the full network is able to produce the desired Nash structure of $(+1, -1, -1), (-1, +1, -1)$, and $(-1, -1, +1)$. The alternative however, where the pure Nash equilibria are $(+1, +1, -1), (+1, -1, +1)$, and $(-1, +1, +1)$, exists for the full network and star network. There are cases where the Nash structure appears in both networks, and other cases where it only appears in one but not the other.

For the legislator game, every region with a legislator Nash structure in the star network also induces a legislator structure in the full network. In other words, the region for the star network is a subset of the region for the full network. On the other hand, there are regions that give rise to a legislator structure in the full network but not the star network.

Significant work remains to fully analyze the collaborations and group creativity on networks, like in the collective improvisation experiments. The goal of the collaboration model is to open the first door in a new interpretation of using the structure of game theory to discuss information flow on networks of collaboration using the coordinate system developed in chapter 2, and how this information is communicated through the various networks of information flow. The takeaway from the results in this section is that certain parameter distributions that induce desired results in a complete network do not necessarily carry over to networks in which information flow is tampered with.

4.8 Information Flow and the Trio Model

In Section 4.3 the externality terms were not analyzed,and the interpretation of information rather than payoffs in a game was motivated. In Section 4.5 this notion of information generation and transmission was expanded through the coordinates of the collaboration.

An important feature of the trio game in Section 4.3 is the existence of the drummer, an agent who cannot transmit melodic and harmonic information. In the collaboration model, this can be reflected in the information the drummer generates alone and with the other agents. Remember that we are denoting the saxophone player as agent s, the bassist as agent b, and the drummer as agent d. Then, β_{sd} and β_{bd}, the information agents s and b receive from agent d, respectively, cannot contain melodic and harmonic information. Furthermore, β_{sdb} and β_{bds}, the information agent i receives from agent d's interaction with agent j, for $i, j = s, b$ and $i \neq j$, also cannot contain melodic and harmonic information. This presents interesting implications and possibilities for the model, but regrettably a proper analysis of these details are beyond the scope of this thesis, and this analysis is saved for future work.

We hope it is apparent that there is a rich world of research waiting to be explored in the collaboration model.

Bibliography

[1] V. Bala and S. Goyal. Learning from neighbours. *Review of Economic Studies*, 65(3):595–621, 1998.

[2] V. Bala and S. Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000.

[3] B. Bartlett. *Practical Recording Techniques: The step-by-step approach to professional audio recording*. Focal Press, 2013.

[4] L. E. Blume. The statistical mechanics of strategic interaction. *Games and economic behavior*, 5(3):387–424, 1993.

[5] O. Candogan, I. Menache, A. Ozdaglar, and P. A. Parrilo. Flows and decompositions of games: Harmonic and potential games. *Mathematics of Operations Research*, 36(3):474–503, 2011.

[6] J. C. Harsanyi, R. Selten, et al. A general theory of equilibrium selection in games. *MIT Press Books*, 1, 1988.

[7] D. Jessie and R. Kendall. Decomposing models of bounded rationality. Technical report, 2015.

[8] D. T. Jessie and D. G. Saari. *Symmetric Decomposition of Games: The me vs. we Tension*. Forthcoming.

[9] M. Kennedy, T. Rutherford-Johnson, and J. Kennedy. *The Oxford dictionary of music*. Oxford University Press, 2013.

[10] B. Milkowski. *Jaco: the extraordinary and tragic life of Jaco Pastorius*. Hal Leonard Corporation, 2005.

[11] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

[12] J. Newton and D. Sercombe. Agency, potential and contagion. *SSRN Electronic Journal*, 2017.

[13] J. Pastorius, D. Alias, M. Brecker, J. DeJohnette, P. Erskine, C. Findley, H. Hancock, H. Johnson, T. Scott, W. Shorter, and T. Thielemans. *Word of mouth.* Warner Bros, 1981.

[14] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[15] G. T. Toussaint. A comparison of rhythmic similarity measures. In *ISMIR*, 2004.

[16] G. T. Toussaint. The geometry of musical rhythm. In *Japanese Conference on Discrete and Computational Geometry*, pages 198–212. Springer, 2004.

[17] D. Tymoczko. The geometry of musical chords. *Science*, 313(5783):72–74, 2006.

[18] H. P. Young. The diffusion of innovations in social networks. *The economy as an evolving complex system III: Current perspectives and future directions*, 267:39, 2006.

[19] H. P. Young. The dynamics of social innovation. *Proceedings of the National Academy of Sciences*, 108:21285–21291, 2011.

[20] K. J. S. Zollman. The communication structure of epistemic communities. *Philosophy of Science*, 74(5):574–587, 2007.

[21] K. J. S. Zollman. The epistemic benefit of transient diversity. *Erkenntnis*, 72(1):17–35, 2009.

[22] K. J. S. Zollman. Social network structure and the achievement of consensus. *Politics, Philosophy & Economics*, 11(1):26–44, 2011.