INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

THREE APPLICATIONS OF GAME THEORY

KIMBERLY F. KATZ

A Dissertation

in

Economics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

1996

Supervisor of Dissertation

Graduate Group Chairperson

UMI Number: 9712943

UMI Microform 9712943
Copyright 1997, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

ACKNOWLEDGMENTS

First, I wish to express my most sincere gratitude to Prof. Akihiko Matsui, whose inspired mentoring and steadfast encouragement have meant an enormous amount to me. No single individual has had a greater impact on either my professional development in general or on this thesis in particular.

I am also greatly indebted to my other committee members, Prof. Stephen Morris and Prof. Andrew Postlewaite. Many thought-provoking discussions with them have strongly influenced the way I think about both the topics in this thesis and about economics research in general; this thesis has benefitted enormously as a result. I am also grateful to Prof. Beth Allen (now at the University of Minnesota), and to Profs. Stephen Coate and George Mailath, all of whom have taken an interest in my success as a graduate student and who have lent a great deal to the advancement of this thesis by taking time to have numerous conversations with me regarding various portions of the work.

Thanks, too, to many friends who have so enhanced my graduate school experience. In particular, thanks first to Nancy Brooks, Michele McLennan, and Natalie Shapiro for making our first year of grad school so manageable; I suspect that never before nor ever since has Penn seen such a neurotic group of first year students. And to them and many others, including Anita Chaudhuri, Jill Constantine, Leah Gutierrez, Bob Hunt, Russell Lamb, and Sara Solnick for making graduate student life so stimulating, while also regularly reminding me that there are good and fun things to life, even while one is trying to write a thesis.

Finally, a special thanks to my parents for their consistent love and support. And my deepest thanks to Greg for his boundless love and truly unbelievable patience. His unwavering belief in my ability to complete this dissertation was, without question, the cornerstone of my actually doing so.

ABSTRACT

THREE APPLICATIONS OF GAME THEORY

KIMBERLY F. KATZ

AKIHIKO MATSUI

The methods of game theory are used to discuss three features pertinent to numerous economic interactions. Explicitly recognizing these features offers new insights into many issues previously studied in economics. In Chapter 1, I study the behavior of agents confronted with a variety of different games, but who cannot fully analyze each of these games individually. This requires first defining similarity of games. Because of the fundamental relationship between similarity and decision making, the only acceptable way to do this is to use definitions generated endogenously, much in the way beliefs are held and updated by decision makers. I show that taking explicit account of similarity in this way and imposing a mild bounded rationality constraint on agents rules out certain equilibria which a repeated game framework admits. In Chapter 2, it is established that interaction requires coordination on certain behavioral standards. Typically, this coordination comes at a cost. This cost is incurred by both the individual attempting to achieve coordination and other members of society. A random matching model in which agents exchange endowments is used. Agents choose a standard of behavior, knowing that the gains from trade are higher if two trading agents have chosen the same standard. Agents preferring a particular standard are considered to be a community. It is shown that in some cases, total welfare of a minority community decreases when a trade barrier is lifted between it and the majority community. An example is also offered of a case in which, where no dominant culture exists, members of both communities may be worse off when trade barriers are lifted. In Chapter 3, I study the problem of modeling collective negotiations as two-player bargaining games, when the players in the negotiating groups have differing preferences over the outcome of the negotiation. I show that it is generally necessary to represent the group by a player with preferences significantly more extreme than those of a player with median preferences, even if the group makes decisions using majority rule voting. How much more extreme this representative should be depends upon the parameters of the game.

Contents

1 Similar Games **1**

1.1 Introduction 1

1.2 Defining Similarity 5

1.3 The Model 12

1.4 Equilibrium in Similar Games 14

1.5 Restricted Categorization 17

1.5.1 $f^i : \mathcal{C}^i \to \triangle(\mathcal{S})$ 19

1.5.2 $\varphi^i : \mathcal{C}^i \to E$ 22

1.6 Examples 26

1.7 Concluding Remarks 30

1.8 References 31

2 When Trade Requires Coordination **33**

2.1 Introduction 33

2.2 The Model 37

2.3 Equilibrium and Dynamics 40

2.3.1 Autarky 40

2.3.2 Unification 41

2.3.3 Dynamics 43

2.4 Welfare Implications 46

2.4.1 Example 1: Uniform Distribution 47

2.4.2 Example 2: Two Mass Points 48
2.5 The Case of 3 Standards 50
2.6 Concluding Remarks 53
2.7 References 55

3 Representing Collective Bargaining Games by Two-Player Bargaining Models **57**
3.1 Introduction 57
3.2 The Model 60
3.3 Comparative Statics 62
3.3.1 α Large 62
3.3.2 α Small 64
3.4 Alternatives to Majority Rule 66
3.5 Additional Remarks on the Complete Information Case 68
3.6 Negotiator Choice Under Incomplete Information 71
3.7 Concluding Remarks 77
3.8 References 78

List of Figures

1.1 Games Which Have Approximately the Same Payoffs, but Dissimilar Evident Ways to Play 7

1.2 These Measures Are Not "Transitive" 8

1.3 π References 13

1.4 Intuition Regarding Lemma 1 19

1.5 Prisoner's Dilemma 27

2.1 Matching Technology: Autarky Case 39

2.2 Matching Technology: Unification Case 39

2.3 Equilibrium: Autarky Case 42

2.4 Equilibrium with Best Response Dynamics: Unification Case 44

2.5 Probabilities in Equilibrium: Two Mass Points Example 48

2.6 Payoff Matrix: The Case of 3 Standards 51

3.1 Equilibria Dependent Upon α. 65

3.2 Reaction Functions When $\alpha \sim U[\underline{\alpha}, 1.5]$ 75

Chapter 1

Similar Games

> In reality, all arguments from experience are founded on the similarity which we discover among natural objects, and by which we are induced to expect effects similar to those which we have found to follow from such objects. ... From causes which appear *similar* we expect similar effects. This is the sum of all our experimental conclusions.
>
> Hume, *An Enquiry Concerning Human Understanding*, 1748

1.1 Introduction

When people are faced with a problem which is new to them, they usually rely on experience with problems they think are similar for help in finding a solution. Moreover, the basic process of human reasoning involves the use of analogy. And yet, economists have barely begun to explore the impact which this use of analogy has on the economic decisions people make. In this chapter, I begin to explicitly study the effects of reasoning by analogy on strategic economic behavior, emphasizing the fact that the analogies one player draws may differ from those another draws.

As an example, I focus in the early sections of this chapter on the learning literature in game theory. Typically, authors studying learning begin by assuming that some group of players are engaged in a repeated game. From here, various issues such as whether or not convergence to an equilibrium strategy or outcome is possible and which equilibria are most likely to be played are investigated. But we know that people rarely find themselves engaged in such a repeated situation or game. Rather, they are faced with a variety of different situations through time. And yet, people still learn from their past experience. Thus, this example offers a framework which is both important and straightforward for studying such analogy, or "similarity."

It has been claimed that the repeated game scenario can be considered as an approximation for the behavior of agents across 'similar' games. In this chapter, I show that taking explicit account of similarity can rule out equilibria that the repeated game analysis admits. Moreover, one class of equilibria ruled out are those necessary to allow the purification interpretation of mixed strategy equilibria. Thus, I propose that the problem of analyzing behavior across different games, even if they are 'similar,' is distinct from that of analyzing behavior in repeated games.

A key point that I argue in this chapter is that for us to be able to consider behavior of agents across similar games, we must first think about why similarity is important. The primary reason that similarity is important is that it is integrally related to the process of decision making.[1] Only after we better understand this relationship between similarity and decision making, can we begin to define similarity.

First, let me point out that the seemingly simple task of determining whether or not some group of objects or situations are 'similar' is, in fact, a rather daunting one. This is because two objects can be assessed to be similar in one situation whereas they may be seen as strikingly different in another. In addition, two people may have different views as to whether or not two objects or situations are similar, based perhaps on their having had different experiences or having perceived the same experiences differently. To see this, consider the following two examples.

Imagine a person hiking through the desert on a hot summer afternoon. To this

[1] This can be thought of as an attempt to step back and think further about an agent's decision making process. For a strong argument in favor of making such an attempt, see, for example, Binmore (1987, 1988).

hiker, a cold glass of water and a cold glass of milk are likely to seem fairly similar; both would quench the hiker's thirst. However, now consider this hiker sitting at the breakfast table on a cool fall morning. Quite possibly, the person will view these two items as being very different in that they do not work equally well in a bowl of cereal. Here, the use of the term similar is clearly related to the context in which the items being compared will be used.

But now, consider another example. Suppose that a traveller from, say, the Middle East is visiting the United States for the first time and has not studied American practices. And now suppose that the traveller decides to purchase a pair of pants. He enters a small clothing shop where he sees a variety of pants hanging on a rack and a shopkeeper behind a counter. To him, this situation seems familiar; it is similar to occasions when he has gone to purchase pants before, in his native country. He chooses a pair and takes them to the shopkeeper, who tells him that the pants cost $30. He immediately counters with an offer of $10 for the pants, since he assumes that if the sales person suggested he pay $30, he must surely only really be hoping to get $15 for the pants. The shopkeeper is offended. And while the shopkeeper will sell the pants for no less than $30, the traveller feels that he will certainly be getting an unfair price if he pays the full $30. So no sale is made.

Clearly, the past experiences of the shopkeeper and of the traveller differ. Each views the situation as being similar to other situations, but of course, these situations differ. Quite possibly, if the shopkeeper had seen this situation as being similar to that in a Middle East shop, he might have first suggested a price of $60, planning on getting $30 in the end. Similarly, if the traveller had seen the situation as similar to that in other U.S. shops, he would have recognized U.S. pricing practices and he might have paid the $30. Here, the behavior of these agents is clearly based on how each views the situation as similar to other situations. And, in this case, these views depend upon their past experiences.

From examples such as these, I conclude that models in which the determination of similarity is exogenously and singularly made do not adequately capture the reason that we, as economists, are interested in studying similarity. Again, the reason we care about similarity is that it is involved in agents' decision making process. Further, if we assume that similarity and decision making are fundamentally related, then we must also recognize that the way different agents find similarity among a group of problems or games may differ.

To handle this, I call upon a highly regarded and intuitive theory titled "analogical reasoning" which has been developed extensively by psychologists and which has received more recent attention from computer scientists working on topics in artificial intelligence. Analogical reasoning is a mechanism for exploiting past experience and/or knowledge of others' past experience in problem solving. The theory asserts that a person, when faced with a problem he has not previously seen, will be reminded of past situations that to him seem similar to the present problem. This reminding experience then serves to retrieve behaviors that were successful in these earlier problem solving episodes and this successful past behavior is then adapted to meet the demands of the current problem.

The model I construct is built with this theory in mind. I take the view that agents are only boundedly rational in the sense that they cannot fully analyze each game they must play de novo, due perhaps to the complexity of the game, a limited amount of time, or any other such reason. Instead, they must rely on comparisons they can make with other games. One example is to think of agents relying on their own past experience as a guide. An agent, when given a game to play which he has not yet seen, will think back to games which he has played previously and which he sees as 'similar' in deciding how to play the new game. Thus, the way in which the agent finds similarity among various games governs the way he will approach games he finds himself called upon to play.

I maintain that the way in which an agent finds similarity among games is a part of the agent's beliefs. I assume that beliefs (here, the similarity assessments) are given by nature, but that these assessments are updated after each game played, much in the way that beliefs are updated using Bayes rule in traditional models. Here, however, I do not specify a specific updating rule, but rather assume simply that the assessments are updated in some way so as to optimize for the player. To model this, I let the players in the model "choose" their similarity assessments in each period. However, again, I stress that I do not intend this to suggest that agents can consciously choose their beliefs, but rather that these beliefs are evolving (unconsciously) over time.

In my model, agents are faced with a variety of different games.[2] Based on some private information, which can be thought of as consisting of their past experience, group these games into 'categories.'[3] Games which are in a given category for an agent represent those games which that agent sees as similar and the agent must play the same strategy when confronted with any of these games. I show that certain equilibria which are possible in a repeated game model cannot be equilibria of this type of model under some fairly mild and rather plausible assumptions.

The organization of the chapter is as follows. In Section 2, I focus on defining similarity. In Section 3, the formal model is laid out, and in Section 4, the equilibrium concept for this model, which is merely a reinterpretation of Nash equilibrium, is introduced. In Section 5, I impose some interesting bounded rationality constraints on players and derive some results from this. Finally, Sections 6 and 7 contain some alternative examples and some concluding remarks.

1.2 Defining Similarity

"Similarity" is implicitly a part of virtually every decision and transaction a person makes. Even the construction of language is essentially a similarity assessment. For example, consider the word "computer." This is a word constructed to describe a group of goods which are considered similar enough to be given a single label. Of

[2]In my formal model, which for the purposes of this chapter is a static one, this is represented as the possibility of seeing a variety of different games.

[3]Again, as the formal model is a static one, I do not assume anything about the information agents have. However, implicitly, I have something such as this in mind. And again, this choice is merely a representation of the presumably unconscious, but optimizing, updating of beliefs.

course, we can discuss computers as being more or less similar to each other. such as a laptop computer versus a desktop computer. By the same token, all computers fit under the broader umbrella, with a more vague implicit definition of similarity, of "machine." Thus, some aspects of similarity are, in a certain sense, exogenously defined for us by virtue of an established labelling system.[4]

However, when it comes to individuals making decisions over a variety of problems (or choosing strategies in a variety of games), such definitions of similarity are not so well defined. Further, similarity takes on a much more explicit role in such a case. Thus, before we can discuss behavior of economic agents who face various similar situations or games, we must first be able to better define the term 'similar,' as it is relevant for these agents and these situations. There have thus far been few attempts within economics to deal explicitly with this problem. and as yet, no fully satisfactory results have been attained.[5]

The most prevalent definitions of this type of similarity are those in which two games are defined to be similar if they have "approximately the same payoffs." Examples of this are definitions in which similarity of two games is determined based on the distance between the games' payoffs. More specifically, games are assumed to be drawn from some metric space, and if the games are sufficiently close according to some metric, they are said to be similar.[6] While these definitions can be appealing in that they often have nice mathematical properties and are relatively intuitive. they do have some drawbacks. One immediate problem, as pointed out in Kreps (1990), is that games such as the two in Figure 1.1 will, under such a criterion, likely be accepted as similar. However, they are likely to be played quite differently, as each has a different focal point equilibrium.

Another thing to observe about these definitions is that, while they allow us to

[4]Of course, at some point, this too was endogenously defined.

[5]Some noteworthy and very interesting examples, however, include Rubinstein (1988), Gilboa and Schmeidler (1992), and Fudenberg and Kreps (1990). Rubinstein and Gilboa and Schmeidler deal with the issue of similarity in the context of the single agent decision problem. Fudenberg and Kreps use a game theoretic setting.

[6]See Fudenberg and Kreps (1990) for examples.

			II	
		L	C	R
I	U	0,0	10.1,10.1	0,0
	M	10,10	0,0	0,0
	D	0,0	0,0	10,10

			II	
		L	C	R
I	U	0,0	10,10	0,0
	M	10.1,10.1	0,0	0,0
	D	0,0	0,0	10,10

Figure 1.1: Games Which Have Approximately the Same Payoffs, but Dissimilar Evident Ways to Play

compare two games, they lack a transitivity property which, in certain circumstances, would be valuable in defining similarity among larger sets of games. The reason for this is illustrated in Figure 1.2. Suppose that the space of all games under consideration is depicted by the box $\mathcal{G}$ and that a game in $\mathcal{G}$ is represented by a point in the box.

Now, let our definition of similarity say that all games in an ε-ball around Γ are similar to Γ. So, in diagram (a), all points in $B_\varepsilon(\Gamma)$ represent games which are similar to Γ. However, notice in diagram (b) that $\Gamma \sim \Gamma'$, $\Gamma' \sim \Gamma''$, but $\Gamma \not\sim \Gamma''$. What this means is that we are unable to define classes of 'similar games' within $\mathcal{G}$ unless we choose some fixed, specified reference point in $\mathcal{G}$ with which to compare the remaining elements of $\mathcal{G}$.[7] Further, if part of our motivation for studying similarity of games is that we believe agents use rules of thumb in many situations, this becomes a concern, as we tend to assume that agents use these rules of thumb, or act 'similarly,' in situations which they see as similar. This is especially relevant in the case of normal form games, where the set of actions is discrete. For example, in the previous example

[7]However, one related positive aspect of these definitions which should not be overlooked is that they offer us a clear way of asking questions such as 'how similar are the games Γ and Γ'?'

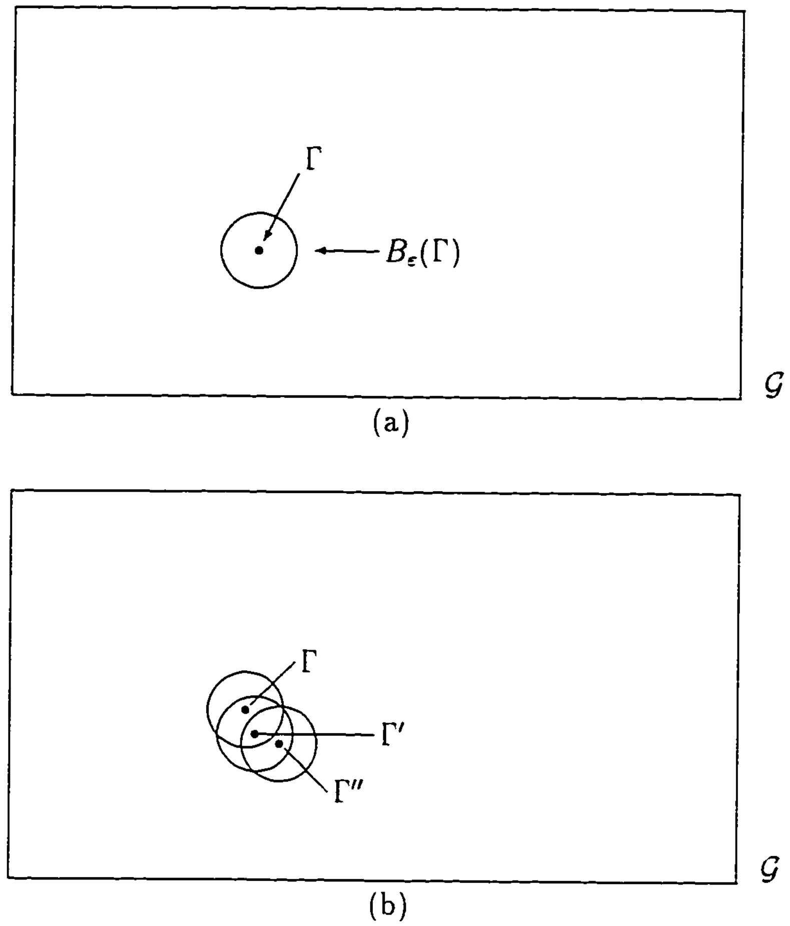


Figure 1.2: These Measures Are Not "Transitive"

of Figure 1.2, we may expect that in the lefthand game, player I will play U and player II will play C while in the righthand game, player I will play M and player II, L. Yet if we say that these two games are similar and that players will play some rule of thumb, or in this particular case take the same action, in games which are similar, clearly this leads to a contradiction.[8]

If we assume that players take this rule of thumb action in games which are similar such as those above, we would have to impose some sort of transitivity property on our similarity measure. Yet, if we did assume these measures to be transitive, we would essentially be saying that all games in $\mathcal{G}$ are similar and thus agents play the same in all of them. For most specifications of $\mathcal{G}$, this is nonsensical.

These two observations lead me to conclude that alternative definitions of similarity which allow boundaries to be drawn between various groups of games will be valuable for capturing our intuition about the way in which agents may view games. For normal form games, which we can think of as representable by their payoff vectors, we may, for example, want definitions which place linear boundaries between certain classes of games. Thus, the definitions I propose in this chapter allow such boundaries to be drawn.

Finally, and most importantly, is that these definitions, as they have been presented in the literature thus far, are exogenous in the sense that it is the modeler who defines *a priori* what it is for games to be similar. However, as discussed in Section 1, we are interested in studying similarity of games primarily because we believe similarity is related to the way in which agents make decisions. For example, we mentioned above that we see agents using rules of thumb in many situations. If we are to

[8]Of course, both of these examples bring up the issue of whether the labels we place on strategies have some meaning or are merely interchangeable entities. I will assume that the labels do play a role. That is, if a player plays a game such as the lefthand game of Figure 1.2 a number of times and sees a certain action, say L, taken by his opponent many times, then if he is suddenly confronted with the righthand game and believes that it is similar to the lefthand game, he *may* believe that his opponent will continue to take the action L. Assuming that the labels have no meaning would require that we assume players know that labels have no meaning, or put differently, that players have an additional layer of knowledge about the game. Again, in this chapter, I do not make this assumption.

capture this belief in our models, we should begin to consider models in which agents have endogenously generated assessments of the similarity of situations, where these assessments may be based on factors which are unique to the individual agents. By doing so, we can probe the role similarity plays in an agent's reasoning process.

To do this, I construct a model in which the agents interacting in the model build their own definitions of similarity and then use these definitions as learning tools for making strategy choices when faced with games which are unfamiliar to them. Towards this end, I propose a model which has the following rough characteristics. This description is a preview of the formal model outlined in Section 1.3.[9]

There is a large society of players who are repeatedly randomly matched and in each period, each matched pair is randomly assigned some game from a large set of games. I take the view that these players are unable to analyze each new game they are faced with de novo. Instead, the players each have a store of information, which can be thought of as including their own past experience, the past experiences of others, or both, and they use this information in a specific way when they are called upon to play a game.[10]

To capture the belief that players play similarly in games which they see as being similar, I assume that a player who is called to play a game will look into his information (or beliefs) and find a game or some group of games which he sees as being similar to the game he is now faced with. He will then employ a strategy which he feels was or would have been successful in these similar games. To model this, I implicitly assume that each period has two stages. In the first stage of the period, each player must in some way partition the set of potential games, where for simplicity I assume this to be the set of 2×2 normal form games. An element of the partition, which I will call a category, represents games that player believes to be 'similar.' Presumably,

[9]However, in the formal model I enumerate in Section 1.3, I will consider only a static, one-shot version of this dynamic model. However, underlying the one-shot game I set up, I have a dynamic story such as this one in mind.

[10]Note that the information set of each player need not and in general is not the same as that of other players. Again, in my formal model, I do not specify the origin of this information, but intuitively, it makes sense for our purposes to view it as having its origins in players' experiences.

the player can make use of any new information he acquired in the previous period. In addition, the player must associate a single strategy with each of these categories. This strategy will represent the way he estimates, at that point in time, is the best way to play games in that category. In the second stage of the period, the player is randomly matched with an opponent and the pair is randomly assigned a game to play. At this point, the player must look to the categorization he made earlier in the period, find the category into which the current game falls, and play according to the strategy he associated with that category when he made the categorization. Essentially, this association of a single strategy with a grouping of games is much like the idea of players choosing rules of thumb for games which they see as similar.

An equilibrium of such a model will involve not only the strategies being played in games, but the definitions of similarity which the players maintain. In this way, I do not exogenously define similarity of games at all, nor in the end can I pinpoint a unique 'correct' definition of 'similar games.' This should not be seen as a negative result, as what I can do, using my model, is analyze whether or not a particular definition of similarity is viable in equilibrium. Again, this corresponds to an analysis of whether players have 'correct' beliefs, but in this case, there is not a single 'truth' for all players, but rather players may in equilibrium have different views as to which games are similar.

Note also that I avoid the other problems which I associated earlier in this section with the distance-between-payoffs definitions by allowing players to define boundaries between games if they wish to do so. In my general model, players also have the option of categorizing games as similar using a metric approach and in this sense, these definitions can be seen as a special case of my model. However, for many classes of games, definitions of this type are not an equilibrium.

1.3 The Model

Let $\mathcal{G} = [0,1]^8$ represent the set of 2×2 normal form games and let an element of $\mathcal{G}$ be denoted by Γ.[11] Assume that there is a continuum of players, denoted by $I = [0,1]$. Players are randomly pairwise matched and for each pair, a game is randomly selected from $\mathcal{G}$. I assume that both the distribution governing the random matching of players and that governing the selection of a game for each pair of players are uniform. My result does not depend on this assumption, but it reduces the amount of notation I will need to introduce.

I will say that if players i and j are matched and assigned a game from $\mathcal{G}$, this game will be labelled as Γ when I am referring to the game as seen from player i's perspective and as Γ^T when referring to it as seen from player j's perspective.[12] These two processes (that of the matching and that of the game assignment) are assumed to be independent, though this is not crucial to my results.

Define the set of categories held by player $i \in I$, denoted $\mathcal{C}^i$, to be a partition of $\mathcal{G}$. A category is then defined to be an element of $\mathcal{C}^i$. I represent category κ of player i as C^i_κ and let $|\mathcal{C}^i|$ denote the cardinality of $\mathcal{C}^i$. Thus, if $|\Theta|$ is finite, $\mathcal{C}^i = \{C^i_1, C^i_2, \ldots, C^i_{|\mathcal{C}^i|}\}$. Games which are grouped together in a category are to be thought of as being similar.

The strategy set, which is the same for all players, is denoted by $\mathcal{S} = \{s_1, s_2\}$. The payoff to player i from playing s_i against the strategy s_j will be denoted by $\pi^i(s_i, s_j)$. For simplicity, I will refer to the specific payoffs of a game $\Gamma \in \mathcal{G}$ as $\{\pi^i_1, \pi^j_1, \ldots, \pi^i_4, \pi^j_4\}$, where these π's correspond to those in Figure 1.3. Then, an example of a category which player i could define is the following. Let $C^i_\kappa = \{\Gamma' \in \mathcal{G} \mid \pi^i_1 > \pi^i_3 \wedge \pi^i_2 > \pi^i_4$ in $\Gamma'\}$, where $\wedge$ represents the operator "and." Then C^i_κ represents a category which

[11] The assumption that $\mathcal{G} = [0,1]^8$ is a restriction in that it does not allow players to consider factors such as context in assessing games. The results in Section 1.4 do not depend in any way on this restriction, but I make the restriction nonetheless to avoid the need for additional and cumbersome notation. I will return to this discussion later in the chapter.

[12] Note that there is no restriction here to symmetric games and therefore that, in general, Γ and Γ^T will appear as different games to the players.

		j: s_1	s_2
i	s_1	π^i_1, π^j_1	π^i_2, π^j_2
	s_2	π^i_3, π^j_3	π^i_4, π^j_4

Figure 1.3: π References

contains all of the games in $\mathcal{G}$ in which s_1 strictly dominates s_2 for player i.

Finally, associated with each player i is a function f^i where initially, I will assume that $f^i : \mathcal{C}^i \to \triangle(\mathcal{S})$ where $\triangle(\mathcal{S})$ denotes the strategy simplex.[13] The strategy which corresponds to C^i_κ under f^i will be denoted σ^i_κ. When player i is selected to play Γ, he must determine into which element of $\mathcal{C}^i$ the game Γ falls and must then play the strategy which is associated with that category under f^i.

To capture the fact that I want a model in which a player's definition of similarity is endogenous, I assume that players choose their own partitions of $\mathcal{G}$ and their own corresponding functions, in the same sense that in standard game theoretic models, players choose their strategies.[14]

To most easily handle this, I assume that players are playing the following two stage metagame. In the first stage, each player i must choose a partition $\mathcal{C}^i$ of $\mathcal{G}$ and then a function f^i which maps each of the categories in $\mathcal{C}^i$ into a strategy. In the second stage of the game, each player is randomly matched with an opponent j and then each pair is randomly assigned a game Γ from $\mathcal{G}$. If $\Gamma \in C^i_\kappa$, then player i must play $\sigma^i_\kappa = f^i(C^i_\kappa)$. If his opponent, player j, has categorized such that $\Gamma^T \in C^j_\lambda$, then

[13]In Section 1.5.2, I will relax this and assume the there is a correspondence $\varphi^i : \mathcal{C}^i \to E$, where E is the space of continuous functions from $\mathcal{G}$ to [0,1], where [0,1] represents $\triangle(\mathcal{S})$, associated with each player.

[14]Recall that this is merely a representation of the updating of the agents' 'beliefs' and is not intended to suggest that agents realistically choose their beliefs.

player i can expect the payoff $\pi^i(\sigma^i_\kappa, \sigma^j_\lambda)$ and player j should expect $\pi^j(\sigma^j_\lambda, \sigma^i_\kappa)$.

For this exercise, it is sufficient to assume that the metagame is a one shot game, but extensions to a dynamic setting are immediate. To interpret this model in terms of the learning literature, we could simply assume that players play this metagame repeatedly. Each period would involve two stages and so forth. Alternatively, though considerably more difficultly, we could place some restrictions on the information which players could use when categorizing each period (such as their own experience) and build a dynamic model from here.

1.4 Equilibrium in Similar Games

In this section, I adapt the notion of a Nash equilibrium to the model just outlined. I redefine the equilibrium concept in terms of this model and examine some of its properties.

An equilibrium in this model involves both the set of categories which a player has defined and the function he has designated to map each of those categories into a strategy. Therefore, I will look at equilibrium categorizations of the players, where a categorization for player i is a pair $(\mathcal{C}^i, f^i)$. The set of categorizations for a society (i.e. the set of categorizations, one for each player) will be referred to as a configuration. To denote this, I define $\Psi = \{(\mathcal{C}^i, f^i)\}_{i \in I}$.

I do not at this point impose any restrictions on the way agents may partition $\mathcal{G}$. The first thing to observe is that for similarity to have some bite, it is necessary to impose some restrictions on players' behavior. By allowing players complete freedom in categorizing, I allow them to choose the singleton partition and a function which designates a Nash equilibrium strategy for each game. This means that any equilibrium of any game in $\mathcal{G}$ can be replicated. It is therefore not surprising that the results here look like the standard Nash equilibrium results. In Section 1.5, I study the implications of imposing restrictions on players' choices of categorizations.

I begin with some definitions. Write $E(\pi^i(\sigma_{\kappa_i}, \sigma_{\kappa_j}))$ for player i's expected payoff when he plays σ_{κ_i} against the mixed strategy of his opponent. Recall that his opponent is also randomly determined. Given $\Gamma \in \mathcal{G}$ and $\Psi = \{(C^i, f^i)\}_{i\in I}$, define $C(\Gamma) = \{C^i_{\kappa_i}\}_{i\in I}$ such that $\Gamma \in C^i_{\kappa_i} \in C^i$ for all $i \in I$. Let $f(C(\Gamma)) = \{\sigma^i_{\kappa_i}\}_{i\in I}$.

Definition 1 *Ψ is an <u>equilibrium configuration</u> if $\forall i$, $\forall (C', f')$, $\forall \Gamma \equiv (\pi^1, \pi^2)$, $E(\pi^i(\sigma^i_{\kappa_i}, \sigma^j_{\kappa_j})) \geq E(\pi^i(\sigma'_{\kappa_i}, \sigma^j_{\kappa_j})$ where $\sigma'_{\kappa_i} = f'(C'_{\kappa_i})$, $\Gamma \in C'_{\kappa}$ and $\sigma^j_{\kappa_j} = f^j(C^j_{\kappa_j})$, $\Gamma^T \in C^j_{\kappa_j}$.*

Notice that this concept of equilibrium is an ex-post concept which is strong in the sense that it requires equilibrium to hold for all games in $\mathcal{G}$. However, in the metagame being played, each game is played with probability zero. Thus, it may seem that an ex-ante type of equilibrium is more intuitively appealing. For this reason, I make the following definition. For a measurable configuration, Ψ, let $E\Pi^i(\cdot)$ be player i's expected payoff before he is assigned an opponent or a game in the two-stage one-shot game delineated in Section 1.3.

Definition 2 *A measurable configuration $\Psi = \{(C^i, f^i)\}_{i\in I}$ is an <u>ex-ante equilibrium configuration</u> if $\forall i$, $\forall (C', f')$, $E\Pi^i(\Psi) \geq E\Pi^i(\Psi_{-i}, (C', f'))$.*

In this chapter, and primarily for simplicity, I will use the ex-post concept of equilibrium, or in other words, I will discuss results assuming that Ψ is an equilibrium configuration. All of these results hold for almost all games if I were to assume instead that Ψ is an ex-ante equilibrium configuration. Such extensions are, in most cases, easy to see, and thus, I will discuss them explicitly only in a few cases.

I will say that a configuration of categories is <u>symmetric</u> if under Ψ, $C^i = C^j$ for all $i, j \in I$, or that a configuration is symmetric if all players have the same partition of $\mathcal{G}$.

Finally, I make the following definition.

Definition 3 *Given (C^*, f^*), let $(C^i, f^i) = (C^*, f^*)\ \forall i$. And let $\sigma^*_\kappa = f^*(C^*_\kappa)\ \forall \kappa$. Suppose $\Gamma \in C^*_\kappa$ and $\Gamma^T \in C^*_{\tilde{\kappa}}$. Then (C^*, f^*) is a <u>symmetric equilibrium categorization</u> if*

$\forall (C', f'), \forall \Gamma, \pi^i(\sigma^*_{\kappa}, \sigma^*_{\bar{\kappa}}) \geq \pi^i(\sigma'_{\kappa'}, \sigma^*_{\bar{\kappa}})$ whenever $\Gamma \in C^*_{\kappa} \bigcap C'_{\kappa'}$ and $\Gamma^T \in C^*_{\bar{\kappa}}$ where $\sigma'_{\kappa'} = f'(C'_{\kappa'})$. The set $\{(C^*, f^*)\}_{i \in I}$ will be called a <u>*symmetric equilibrium configuration.*</u>

Similarly, we can think of ex-ante symmetric equilibrium categorizations and ex-ante symmetric equilibrium configurations.

Since my definitions of equilibrium are essentially a reformulation of the basic notion of Nash equilibrium, many things immediately follow.

An equilibrium configuration exists.

A trivial equilibrium is constructed by considering the singleton partition, i.e. $C^i = \{\{\Gamma\}\}_{\Gamma \in \mathcal{G}}\ \forall i \in I$, and letting players take Nash equilibrium strategies for each game. Note that this implies that every subset of $\mathcal{G}$ has an equilibrium configuration as well. Further, since we know that there exists a symmetric mixed strategy equilibrium for every 2×2 normal form game, a symmetric equilibrium configuration exists for $\mathcal{G}$ as well as for every subset of $\mathcal{G}$.

Given a game $\Gamma \in \mathcal{G}$, and let E be the projection of Nash equilibrium strategies onto player i's component. Suppose $\Psi = \{(C^i, f^i)\}_{i \in I}$ is an equilibrium configuration of $\mathcal{G}$. If $\Gamma \in C^i_{\kappa_i}$, then $f^i(C^i_{\kappa_i}) = \sigma^i_{\kappa_i} \in E$. $\forall \Gamma \in C^i_{\kappa_i}$. $\forall C^i_{\kappa_i} \in C^i$. $\forall i \in I$.

This follows immediately from the definition of an equilibrium configuration. Thus. it is similarly straightforward to show the following.

Let $\Psi = \{(C^i, f^i)\}_{i \in I}$ be an ex-ante equilibrium configuration of $\mathcal{G}$. Then for almost all $\Gamma \in C^i_{\kappa_i}$, then $f^i(C^i_{\kappa_i}) = \sigma^i_{\kappa_i} \in NE(\Gamma)$, $\forall \Gamma \in C^i_{\kappa_i}$, $\forall C^i_{\kappa_i} \in C^i$, $\forall i \in I$.

For this reason, all of the results in the remainder of the chapter, which are discussed in terms of the equilibrum configuration. easily extend to the case of the ex-ante equilibrium configuration.

Finally, we can make this observation about equilibrium configurations of $\mathcal{G}$ before imposing restrictions. Consider some particular game $\Gamma \in \mathcal{G}$. Let $\sigma_1(i) = \Pr$ (i plays s_1) in Γ. Then, $\sigma_1 : [0,1] \to [0,1]$. It is assumed throughout this chapter that σ_1 is Lebesgue-measurable. Thus, $\int_0^1 \sigma_1(i)di = \Pr(s_1)$. I also assume throughout that group behavior, represented by $f : [0,1] \times [0,1]^8 \to \Delta(\mathcal{S})$ is measurable.

The following states that generically, every equilibrium configuration either involves all players playing the same strategies for all games or can be represented by some equivalent such configuration using nontrivial mixed strategies.

Let $\hat{\mathcal{G}} = \mathcal{G} \backslash \{\Gamma \in \mathcal{G}$ such that $\pi_1^i = \pi_3^i$ or $\pi_2^i = \pi_4^i\}$. Suppose that $\sigma(i)$ is the strategy that player i plays in Γ and $\tau(i)$ is the strategy that he plays in Γ^T. Then $(E(\sigma), E(\tau))$ is an equilibrium of Γ.

To see this, consider any game $\Gamma \in \mathcal{G}$ and assume that $\Gamma \in C^i_\kappa$. Let $\sigma = f^i(C^i_\kappa)$. Suppose that $f^j(C^j_{\kappa_j}) = \sigma_j$ for each $j \in I$ $(j \neq i)$ where $\Gamma^T \in C^j_{\kappa_j}$. If Ψ is an equilibrium configuration, then $\forall \sigma' \in \Delta(S), \forall i$, $E[\pi^i(\sigma, \sigma_j)] \geq E[\pi^i(\sigma', \sigma_j)]$. And we know, since π is linear in σ_j, $E[\pi^i(\sigma, \sigma_j)] = \pi^i(\sigma, E(\sigma_j))$. This implies that $\pi^i(\sigma, E(\sigma_j)) \geq \pi^i(\sigma', E(\sigma_j))$ $\forall i, \forall \sigma' \in \Delta(S)$. Finally, $\Gamma \in \hat{\mathcal{G}} \Rightarrow E(\sigma_j) = \sigma^{-i}$ for some $\sigma^{-i} \in \Delta(S)$.

1.5 Restricted Categorization

As mentioned earlier, the above statements hold when there are no restrictions placed on the way players may categorize games. Thus, the conclusions look like those of standard Nash equilibrium analysis. However, theories such as the rules of thumb theory imply that players cannot, or at least do not, perfectly discriminate between games. To explore the impact of this, we must place some restriction on the types of categorizations players may use.

The following assumption provides one such restriction. Pick a game Γ and let $C^i(\Gamma)$ be the category in C^i which contains Γ. And define the join of the players' partitions, which I will denote $\mathcal{J}$, to be the partition generated by

$$J(\Gamma) = \bigcap_{i \in I} C^i(\Gamma) \; \forall \Gamma \in \mathcal{G}.$$

Assumption 1 *The partition $\mathcal{J}$ is countable.*

First, notice that this assumption implies that for all players $i \in I$, the partition C^i is countable.

One restriction which would generate this feature of the join is to assume that players have only finitely many partitions to choose from, where each of these partitions is countable.

There are a number of ways in which Assumption 1 can be interpreted. For example, we can think of players as being bounded by language to defining only countably many categories. That is, if players are constrained to some plausible spoken language to define categories and certainly, languages have countably many words and thus word combinations, then players can define only countably many categories using this language.

One interpretation that lends itself to players' individual partitions is that players have limited memories. Thus, it is not feasible for a player to retain an uncountable partition in this memory or even simply not possible for a player to recall the categories from such a partition when needed, as I am assuming they do in this model.

Again, one partition which this restriction rules out is the singleton partition which was referred to in Section 1.4. However, under this restriction, players still have enormous flexibility in choosing which games to group together.

i \ j	s_1	s_2
s_1	2,3	3,2
s_2	3,2	2,3

Γ

i \ j	s_1	s_2
s_1	2,3 + ε	2,3
s_2	3,2	2,3

Γ^{ε}

Figure 1.4: Intuition Regarding Lemma 1

1.5.1 $f^i : \mathcal{C}^i \to \Delta(\mathcal{S})$

The strongest assumption being made at this point is that $f^i : \mathcal{C}^i \to \Delta(\mathcal{S})$. This says that a player must play identically in every game he sees as similar.[15] The combination of this restriction and Assumption 1 imply some facts worth noting.

Lemma 1 *Let $\Psi = \{(\mathcal{C}^i, f^i)\}_{i \in I}$ be a configuration which satisfies Assumption 1. Then every player is playing non-optimally for almost all games in $\mathcal{G}$ which have no pure strategy Nash equilibrium.*

The intuition behind this observation is quite simple. First, Assumption 1 implies that almost all games in $\mathcal{G}$ are in an element of $\mathcal{J}$ which has positive measure. If we pick any element of $\mathcal{J}$, all players in I view the games in that element of $\mathcal{J}$ to be similar. In other words, every player i has a partition $\mathcal{C}^i$ such that every game in this set is in some equivalence class defined by every player. Now, consider the games in Figure 1.4.

Label the lefthand game as Γ. This game clearly has no pure strategy Nash equilibrium. Suppose that Γ is in some category, say C^i_Γ, which has positive measure and which all players have defined. And suppose that an equilibrium is being played for

[15]Note that a complementary interpretation of this model is to assume that players cannot perfectly distinguish between games and that the partitions here are somewhat akin to information partitions. However, as I am assuming that players choose their partitions, I will continue to discuss the model in light of the above discussions.

this game. One example would be the symmetric equilibrium in which all players are playing s_1 with probability $\frac{1}{2}$ and s_2 with probability $\frac{1}{2}$. But the fact that C_Γ^i has positive measure means that there is some other game, such as Γ^ϵ which is also in C_Γ^i. Yet the strategy configuration in which all players are playing s_1 and s_2 with equal probabilities is not an equilibrium of Γ^ϵ. Rather, for a corresponding symmetric equilibrium, it would have to be the case that the i players were playing s_1 with probability $\frac{1}{2+\epsilon}$ and s_2 with probability $1 - \frac{1}{2+\epsilon}$. However, since players must play identically for all games which they see as similar, this is not possible. The formal proof of this lemma follows this intuition closely.

Proof. First, pick a game $\Gamma \in \mathcal{G}$ which has no pure strategy Nash equilibrium. And pick a strategy profile s which is an equilibrium of Γ. Then it is easy to show that the set of games in $\mathcal{G}$ for which s is an equilibrium is a set of measure zero. (It is a set of hyperplanes in $[0,1]^8$.

Now, let μ denote a measure on $[0,1]^8$. And let $\mathcal{P} = \{P_1, P_2, \ldots\}$ be a countable partition of $[0,1]^8$. And let $\mathcal{S} = \{S_1, S_2, \ldots\}$ be the subset of $\mathcal{P}$ such that $\mu(S_i) = 0$ for all i and let $\mathcal{T} = \{T_1, T_2, \ldots\}$ be the subset $\mathcal{P}$ such that $\mu(T_j) > 0$ for all j; and let $\mathcal{S}$ and $\mathcal{T}$ be such that $\mathcal{S} \bigcup \mathcal{T} = \mathcal{P}$.

We know that $\mu(\mathcal{S}) = \mu(\bigcup_i S_i) \leq \sum_i \mu(S_i) = 0$ by countable subadditivity of μ.

Thus, if $\mathcal{S} \bigcup \mathcal{T} = \mathcal{P}$, then we know that $\mathcal{T}$ must be nonempty, since $\mu(\mathcal{G}) > 0$. This implies that almost all games in $\mathcal{G}$ are in sets which have positive measure.

Since the set of games in $[0,1]^8$ which have no pure strategy equilibrium has positive measure as well, we also know that if this set is partitioned into countably many elements, almost all of these games must be in sets which have positive measure.

Therefore, for almost all games, if we pick a game Γ which has no pure strategy equilibrium, and we pick an equilibrium (mixed) strategy s for this game, then there must be some other game in the same category as this game for which s is not an equilibrium strategy, since the set of games for which s is an equilibrium is a set of measure zero while the category containing Γ is a set of positive measure.

One implication of this lemma is immediate.

Proposition 1 *Let $\Psi = \{(C^i, f^i)\}_{i \in I}$ be a configuration which satisfies Assumption 1. Then Ψ is not an equilibrium configuration.*

Proof. We know that players must be playing a Nash equilibrium for every game in $\mathcal{G}$, yet Lemma 1 tells us that under Assumption 1, this cannot be the case. Thus, Ψ cannot be an equilibrium configuration.

Note, however, note that there are ways to recapture an equilibrium in this model. The most intuitive of these is adding in certain forms of costs to categorization. In this way, we may bring back an equilibrium. One simple example is that we can require a player to pay some small cost c for each element of the partition he defines. That is, player i would pay $c|C^i|$. In this case, there would come a point in the player's refinement process where the benefits to further refining his partition are outweighed by the gain in cost associated with doing so. However, it is important to recognize that with costs as discussed here, the particular equilibria we can obtain will be a direct function of the costs we choose. Further pursuit of this direction is left to future research. In addition to the cost approach, the concept of ϵ-equilibrium used in place of the Nash equilibrium concept I use here could, properly formulated, restore equilibrium.

There is one further thing which the previous proposition implies. Let $\tilde{\mathcal{G}}$ represent the set of games in $\mathcal{G}$ which have at least one symmetric pure strategy equilibrium. Then, by the same process of reasoning, we can rule out players using mixed strategies for these games in equilibrium.

Proposition 2 *Let Ψ be a category symmetric equilibrium configuration of $\tilde{\mathcal{G}}$ in which all categories in C^i satisfy Assumption 1. Then Ψ is a symmetric equilibrium configuration and $\sigma^i_\kappa \in \mathcal{S}$ $\forall i$, $\forall \kappa$ such that $C^i_\kappa \in C^i$.*

The proof of this observation follows the same logic of the proof of Lemma 1. Thus, it will be omitted here.

1.5.2 $\varphi^i : \mathcal{C}^i \to E$

Consider the following generalization of the above model. Now, instead of assuming that $f^i : \mathcal{C}^i \to \Delta(\mathcal{S})$, I allow players to choose a correspondence $\varphi^i : \mathcal{C}^i \to E$ where E is the space of continuous functions from $\mathcal{G}$ to [0,1], and where this [0,1] represents $\Delta(\mathcal{S})$. Instead of forcing players to select a specific element of $\Delta(\mathcal{S})$, I allow them to associate a function, which must be continuous in payoffs, to each category.

This extension gives players the ability to adapt their behavior for differences they see between similar situations. However, players must still act "similarly" in games they deem to be similar. In this sense, I am able to directly handle the problem I mentioned in Section 1.2 regarding transitivity of the similarity measures. Players can implicitly characterize some games as being more similar than others, as was the case with these measures, but they still have the ability to draw boundaries where appropriate. This is important because clearly, equilibrium strategies are not continuous across all games.

As an example of a strategy choice which is now allowed but previously was not, suppose that player i chooses a partition which has a category containing only games which have no pure strategy equilibrium. Under the original assumption on f^i, player i would have to choose a single element of the strategy simplex to designate his play in games in this category. Now, however, he may choose a strategy which, for example, dictates that he play s_1 with probability p^* and s_2 with probability $1 - p^*$, where $p^* = \frac{\pi_4^j - \pi_3^j}{\pi_1^j - \pi_2^j - \pi_3^j + \pi_4^j}$ where j refers to his opponent. Note that the constant strategies which were required before are still allowed under this more general specification.

Also, notice that we can construct an equilibrium configuration in which every category has a nonempty interior. The problem in doing this before came primarily when we needed to categorize games with no symmetric pure strategy equilibrium, but allowing these more general strategy functions takes care of this problem.

One example of a symmetric equilibrium categorization which satisfies Assumption 1 is the following. For simplicity, the categorization is for the set of generic games.

or more specifically, games which do not involve ties in payoffs.

$$
\begin{aligned}
C_1^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i > \pi_3^i \;\wedge\; \pi_2^i > \pi_4^i \text{ in } \Gamma'\} \\
C_2^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i < \pi_3^i \;\wedge\; \pi_2^i < \pi_4^i \text{ in } \Gamma'\} \\
C_3^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i > \pi_3^i \;\wedge\; \pi_2^i < \pi_4^i \;\wedge\; \pi_1^j > \pi_2^j \;\wedge\; \pi_3^j > \pi_4^j \text{ in } \Gamma'\} \\
C_4^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i < \pi_3^i \;\wedge\; \pi_2^i > \pi_4^i \;\wedge\; \pi_1^j > \pi_2^j \;\wedge\; \pi_3^j > \pi_4^j \text{ in } \Gamma'\} \\
C_5^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i > \pi_3^i \;\wedge\; \pi_2^i < \pi_4^i \;\wedge\; \pi_1^j < \pi_2^j \;\wedge\; \pi_3^j < \pi_4^j \text{ in } \Gamma'\} \\
C_6^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i < \pi_3^i \;\wedge\; \pi_2^i > \pi_4^i \;\wedge\; \pi_1^j < \pi_2^j \;\wedge\; \pi_3^j < \pi_4^j \text{ in } \Gamma'\} \\
C_7^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i > \pi_3^i \;\wedge\; \pi_2^i < \pi_4^i \;\wedge\; \pi_1^j > \pi_2^j \;\wedge\; \pi_3^j < \pi_4^j \text{ in } \Gamma'\} \\
C_8^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i < \pi_3^i \;\wedge\; \pi_2^i > \pi_4^i \;\wedge\; \pi_1^j < \pi_2^j \;\wedge\; \pi_3^j > \pi_4^j \text{ in } \Gamma'\} \\
C_9^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i > \pi_3^i \;\wedge\; \pi_2^i < \pi_4^i \;\wedge\; \pi_1^j < \pi_2^j \;\wedge\; \pi_3^j > \pi_4^j \text{ in } \Gamma'\} \\
C_{10}^* &= \{\Gamma' \in \bar{\mathcal{G}} \mid \pi_1^i < \pi_3^i \;\wedge\; \pi_2^i > \pi_4^i \;\wedge\; \pi_1^j > \pi_2^j \;\wedge\; \pi_3^j < \pi_4^j \text{ in } \Gamma'\}
\end{aligned}
$$

and where $f^*(C_1^*) = f^*(C_3^*) = f^*(C_6^*) = f^*(C_7^*) = s_1$; $f^*(C_2^*) = f^*(C_4^*) = f^*(C_5^*) = s_2$; and $f^*(C_8^*) = f^*(C_9^*) = f^*(C_{10}^*) = s_1$ with probability p^* and s_2 with probability $1 - p^*$, where $p^* = \frac{\pi_3^j - \pi_4^j}{\pi_2^j + \pi_3^j - \pi_1^j - \pi_4^j}$ for player i when his opponent is player j.

Notice that this is just a way of saying, without using game theoretic terminology, that in one possible equilibrium, players recognize dominance solvable games (those in C_1^* through C_6^*); coordination and battle-of-the-sexes games (C_7^*); games of 'chicken' (C_8^*); and games in which there are no pure strategy Nash equilibria (C_9^* and C_{10}^*).

Proposition 3 *Let Ψ be an equilibrium configuration of $\mathcal{G}$ which satisfies Assumption 1. Then Ψ must be such that there are players playing nontrivial mixed strategies for almost all games in $\mathcal{G}$ which do not have a pure strategy equilibrium.*

Equilibria which are strict purifications of mixed strategy equilibria cannot be part of equilibrium for almost all games in the set of games with no pure strategy equilibrium. Note that there are many types of partitions which would imply that this result apply for other games in $\mathcal{G}$ which do not have a symmetric equilibrium.

The intuition behind this result is much like that behind Proposition 1. Consider again, the games in Figure 1.4. Suppose that Γ is in the interior of C_Γ^i for all i. The purification equilibrium for the game Γ involves $\frac{1}{2}$ of the i-population playing s_1 and $\frac{1}{2}$ playing s_2. However, the equivalent equilibrium for the game Γ^ε involves $\frac{1}{2+\varepsilon}$ of the i-population playing s_1 and $1-\frac{1}{2+\varepsilon}$ playing s_2. If players can use only pure strategies, then clearly, at least one player, say player k, had to be playing s_1 in Γ but s_2 in Γ^ε, which would imply a discontinuity in the function $\varphi^k(C_\Gamma^k)$.

Proof. The proof of this proposition relies on the fact that a strategy correspondence which has a jump from s_1 directly to s_2 as we change the game is not continuous.

Pick a game Γ in $\mathcal{G}^{mix}$ which is in the interior of some category, say C_Γ^i in C^i. We know from Assumption 1 that such a game exists, and further, that almost all games in $\mathcal{G}^{mix}$ are in a category which has positive measure. Suppose that the equilibrium for this game is the purified equilibrium, with some fraction ρ of the population playing s_1 and the remaining fraction $1-\rho$ playing s_2. It must be the case that

$$\rho\pi_1^j + (1-\rho)\pi_3^j = \rho\pi_2^j + (1-\rho)\pi_4^j,$$

or equivalently,

$$\rho = \frac{\pi_4^j - \pi_3^j}{\pi_1^j - \pi_2^j - \pi_3^j + \pi_4^j}.$$

But as in earlier proofs, we know that for all values of ε which have absolute value less than some δ, the games $\Gamma^\varepsilon = \{\pi_1^i, \pi_1^j + \varepsilon, \pi_2^i, \pi_2^j, \ldots, \pi_4^i, \pi_4^j\}$ are also in C_Γ^i. For the equivalent purified equilibrium of this game, it must be the case that

$$\rho = \frac{\pi_4^j - \pi_3^j}{\pi_1^j + \varepsilon - \pi_2^j - \pi_3^j + \pi_4^j}.$$

This would imply that at least one player, say player k, played s_1 in Γ, but s_2 in Γ^ε, no matter how small ε. This implies that φ^k could not be continuous. Thus, it must be that equilibrium involves players playing nontrivial mixed strategies for these games.

Again, this says that in the case where all agents "view the world in the same way," no equilibrium configurations can contain purifications of mixed strategy equilibria, at least for almost all games.[16] This result is interesting on two counts. First, we know that people have been concerned about the notion of mixed strategies for a long time and that the concept of purification was motivated by people's distrust of mixed strategies.[17] Yet this proposition points out that if the world in which agents are making decisions is complex, purification of strategies must also be approached with caution. Purified mixed strategies in such a world are subject to concerns as are the mixed strategies themselves.

More broadly, this result points out that there are equilibria which players playing a single game repeatedly may be able to 'learn' but which, when players are playing a variety of different games over time, may be too complex. That is, the set of equilibria in a model such as mine is smaller than that obtainable in a repeated game situation. Specifically, there are equilibria which involve players playing different strategies for the same game which can be obtained in the repeated game model, but which cannot be supported here. In fact, it should be noted that the purification equilibrium is simply one example of an equilibrium configuration which combining this model with Assumption 1 rules out. Rather, there many (uncountably many, in fact) which are ruled out by this assumption. It is easy to construct examples. Thus, when we assume players are playing a single game repeatedly while we actually believe they are playing a variety of different games, we are losing significant complexity.

[16]One point worth reiterating here is that I have not allowed for the possibility of players considering the context in which they are playing some game. Further, I am restricting the discussion to a single, fixed society. If I allow for either the possibility that players can condition on some context in which they are playing a given game or the possibility that players from two different societies are being matched and can recognize this, there could be equilibria where this does not hold. One possible example (mentioned in Chapter 6 of Kreps (1990)) is that of social conventions such as the deferral in certain cultures of students to their professors. In battle-of-the-sexes games, for example, if a student can condition on the fact that he is matched with a professor, then coordination on the equilibrium which gives the professor his highest payoff is possible.

[17]The seminal reference on this topic is Harsanyi (1973).

Finally, let me point out that the restriction of $\mathcal{G}$ to the set of 2×2 games was purely for simplicity, and that the results hold for any set of $n \times k$ games, as well as for sets of games which contain various-dimensioned finite strategy normal form games. All that is required is that we assume all players to be categorizing over the same set of games. Note as well that the basic ideas outlined here can be readily extended to handle situations other than normal form games. Some of these are discussed in the following section.

1.6 Examples

There are many situations which models based of this chapter's main idea can help us to better understand. Some involve a relatively direct application of the above model, while others require a bit more abstraction. Here, I mention two of these. In addition, there are other types of issues which such a model can help us address, and I note one here as well.

Example. Perhaps the best known example which this theory can directly and systematically help us better understand involves repeated play of the prisoner's dilemma game such as that in Figure 1.5. There is a body of experimental evidence which suggests the following. When a population of players (most of whom have not previously seen the formal game) are asked to choose a strategy in a prisoner's dilemma game, they often choose to cooperate. When asked to play this game repeatedly, the pattern that is often observed is one in which players begin by cooperating, but eventually, defection starts to occur and ultimately, takes hold. Such a phenomenon can result within the framework of the above model. To see this, consider a dynamic version of the model. There are a variety of ways we can set up such a model and obtain the same result. For example, if we assume that players define categories based only on their own past experience, then a player who had primarily encountered situations (or games) in which cooperation was preferable (as is presumably the case for many of life's daily encounters), it is quite possible and even likely, that this player would

		j	
		s_1	s_2
i	s_1	4,4	0,5
	s_2	5,0	1,1

Figure 1.5: Prisoner's Dilemma

initially believe the prisoner's dilemma to be similar to some such game. Thus, when first confronted with the prisoner's dilemma, the player would cooperate. However, after playing such a game a few times, the player will likely realize that this game is indeed not similar to his previous experiences, and at this point, he should choose to recategorize, placing this game in a separate category and assigning it the strategy "defect."

Alternatively, we could model this situation as one in which players can only recategorize periodically, as opposed to after each experience. In such a case, again, if a player's categorization at the time he is confronted with the first repetition of the prisoner's dilemma suggests that this player choose to cooperate when asked to play such a game, we would see cooperation early on. As players are allowed to recategorize, we would expect to see more and more defections.

Example. Thus far, I have focused on categorizations over payoffs. However, we can also consider circumstances in which players categorize over not only (or not even) payoffs, but also (or alternatively) over other characteristics of a game or situation as well. Examples include categorizations or similarity assessments over an opponent's type, over one's own type, over the context in which the game is being played or the decision is being made, and over the rules of the game. Such extensions are virtually immediate. The following example is just one of many which emphasize the importance of considering similarity as something endogenously defined.

Consider this bargaining model which I take from Rosenthal (1991). There is a continuum of bargainers who are randomly matched in pairs on a regular basis to play one of a set of bargaining games. Each player selects a bargaining effort level and whoever has the higher effort level in a matched pair wins the prize. Players know that the set of games they can be given to play contains K different games, each of which is characterized by (W_k, T_k), and they know the distribution from which games will be drawn. Game k will be randomly chosen with probability p_k. If players i and j are matched and they have chosen effort levels e_i and e_j, respectively, then payoffs are as follows. If $e_i > e_j$, then i's payoff is $W_k - e_i$ and j's payoff is $-e_j$. If $e_i = e_j$, then each player receives $T_k - e_i$. ($W_k > T_k > 0$ for all k.) It is assumed that players choose their effort level once and for all at the beginning of time, and that they must exert this effort level regardless of the game which they are given in any particular period. In other words, this is the rule-of-thumb which players choose.

Rosenthal shows that in this metagame, there is a unique population equilibrium which is the c.d.f. of the uniform distribution on $[0, W]$, where $W = \sum_k p_k W_k$.

In his discussion, Rosenthal proposes that one manageable modification of the model is to allow for the possibility that individuals maintain more than one effort level and subdivide the universe of bargaining games into subsets according to which rule is to be used.

Using my model, I can show that, if players are restricted to maintaining fewer effort levels than there are games, i.e. if the number of effort levels a player can define is less than K, then in equilibrium, it *cannot* be the case that all players choose the same partition of this set of games.[18] To see this, consider the following very simple example. Suppose that $K = 3$ with $W_1 < W_2 < W_3$ and that players can choose at most 2 effort levels. Notice first that in Rosenthal's equilibrium, the expected payoff to

[18] I am assuming that $W_k \neq W_l$ for $k \neq l$.

every player is zero. I can easily show that if every player partitions the set of games in the same way, then any one player can give himself a strictly positive (expected) payoff by altering his partition.

Using my example, consider the case where every player has the partition $\{\{W_1, W_2\}, \{W_3\}\}$. Then if there is an equilibrium, it must be described by the c.d.f. of the uniform distribution on $[0, \frac{W_1+W_2}{2}]$ for games 1 and 2, and the c.d.f. of the uniform distribution on $[0, W_3]$ for game 3. And again, each player expects a payoff of zero.

But consider the following option for any one player, say i, in the population. Choose the partition $\{\{W_1, W_3\}, \{W_2\}\}$. And select an effort level of zero for the games $\{W_1, W_3\}$, but an effort level of $\frac{W_1+W_2}{2} + \varepsilon$ for some small $\varepsilon > 0$ for the game W_2. Then player i's expected payoff from the overall game is $\frac{1}{3}[W_2 - (\frac{W_1+W_2}{2} + \varepsilon)]$ which is clearly strictly greater than zero. This exercise can be easily carried out for any partitioning of the games to show that in equilibrium, it cannot be the case that all players have chosen the same partition.

This example illustrates that allowing players to endogenously choose partitions over the space of games (or situations) will, in some cases, lead to very interesting results in which players in equilibrium will not all "see the world in the same way."

Additional Interpretations. There are a variety of questions which this model can help us address. One such question is that of optimal partitioning, in general. Other attempts to deal with this issue can be found in the literature; one interesting example is Dow (1991). In this work, Dow is concerned with the decision problem of an agent with limited memory, searching to find a low price, whose memory is represented by a partition of the set of possible past prices. To represent the fact that his memory is limited, it is assumed that the number of elements of the partition is limited. He goes on to characterize the optimal partition for a specific simple example.

The model I construct here can be easily adapted to deal with such issues. It also allows us to attack questions regarding the optimal partitioning of more complicated situations and games. The bargaining example above offers one such example.

Finally, using this type of model, we can consider game theoretic situations in which one player has, say, a better memory than does his opponent. Similarly, we can also discuss situations in which one player has more computing ability or simply more information than does his opponent.

1.7 Concluding Remarks

The results and examples in this chapter have been cited to illustrate the importance of explicitly considering endogenously defined definitions of similarity in our economic modeling. This is particularly true if we believe that agents are either boundedly rational or able to handle only limited complexity, taking as given that agents face a very complicated world.

The results in this chapter first suggest that we can improve upon the repeated game framework, which has been used thus far in the learning literature, in cases where we believe that agents are actually playing a variety of different games over time, even if these games are 'similar'. There are behavior patterns which can be learned if a single game is played repeatedly, but which may not be so easily learned when a variety of different games are being played.

In addition, we can explain interesting phenomena supported by experimental evidence, such as those discussed in the previous section. Further, this model allows us to understand that in cases where players choose rules of thumb for their behavior, we can find cases where in equilibrium, players will necessarily choose different partitions over the space of games or situations. This avenue has not, to the best of my knowledge, been pursued previously.

In general, further research into the impact of endogenous definitions of similarity, as defined in this chapter, is clearly warranted.

1.8 References

[1] Binmore, K. (1987 and 1988), "Modeling Rational Players, Parts I and II" *Economics and Philosophy* , 3 and 4.

[2] Carbonell, J. (1983), "Learning by Analogy: Formulating and Generalizing Plans from Past Experience," in Michalski, R., Carbonell, J. and Mitchell, T., ed., *Machine Learning: An Artificial Intelligence Approach* , (Morgan Kaufmann Publishers, Inc.).

[3] Cherniak, C. (1986), *Minimal Rationality* (Cambridge, MA: The M.I.T. Press).

[4] Dow, J. (1991), "Search Decisions with Limited Memory," *The Review of Economic Studies* , 58 (1).

[5] Fudenberg, D. and Kreps, D. (1990), "A Theory of Learning, Experimentation, and Equilibrium in Games," (mimeo, Stanford University Graduate School of Business).

[6] Gilboa, I. and Schmeidler, D. (1992), "Case-Based Decision Theory" (mimeo, Northwestern University).

[7] Harsanyi, J. (1973), "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *International Journal of Game Theory*, 2.

[8] Hartigan, J.A. (1986), "Similarity Probability" (mimeo, Yale University).

[9] Hume, D. (reprinted 1988), *An Enquiry Concerning Human Understanding* (Buffalo, NY: Prometheus Books).

[10] Kreps, D. (1990), *Game Theory and Economic Modelling* (Oxford: Oxford University Press).

[11] Rosenthal, R. (1991), "Bargaining Rules of Thumb" (mimeo, Boston University).

[12] Rubinstein, A. (1988), "Similarity and Decision-making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)", *Journal of Economic Theory*, 46.

Chapter 2

When Trade Requires Coordination*

> I am the son of immigrant parents. ... I had to learn American culture on my own, and it is difficult to explain to someone who has not gone through it what this means. ... In short, I was a cultural orphan (like many others) and "speak the culture with an accent."
>
> Isaac Asimov, exerpted from *Yours, Isaac Asimov*

2.1 Introduction

Interaction requires coordination, and the interactions which arise in economic activities are certainly no exception. Trade, for example, requires coordination on a variety of conventions such as language. Corporate mergers often require that the merging companies coordinate on a single corporate culture. Network externalities, a feature of computer and other communications networks, force agents to coordinate one or at most a small number of operating systems. Rural to urban migration is the result of the need for people to coordinate on a geographical location for many of the jobs available in a post agricultural economy. Women often feel that to succeed in fields predominantly inhabited by males, they must adopt numerous attitudes and behavi-

*This chapter is the result of joint work with Akihiko Matsui.

ors traditionally associated with men. Immigrants frequently find that to obtain the better jobs in their new country, they must assimilate into the predominant culture of this country, or phrased differently, they must coordinate their customs with those of their new compatriots. And members of racial or ethnic groups which have historically interacted largely with only members from their own community find that a similar predicament applies to them.

The costs of achieving such coordination can be nontrivial and are, in many cases, very large. For instance, learning a foreign language can be expensive and time consuming. The expenses of moving one's home from one location to another are apparent to anyone who has moved. And changing one's behavior patterns to behavior patterns which are less natural or instinctive can be uncomfortable or even traumatic. (Evidence of such costs can be found in Rubin (1995), Cose (1993), and de Beauvoir (1952) among others.)

In addition, the need for coordination in interactions presents us with a situation involving strategic complementarity. In such a situation, the optimal strategy of an agent depends positively upon the strategies of other agents, or in other words, the more people that choose a behavior, the higher is the payoff to all people who choose this behavior. The literature on strategic complementarity is extensive, including Cooper and John (1988), Farrell and Saloner (1985), Matsuyama (1991), and Murphy, Shleifer, and Vishny (1989). Due to strategic complementarity, members of a minority group may have the incentive to adopt behaviors of a majority group even if they inherently prefer their own behavior patterns, since adopting these alternative behaviors allows them the chance to interact with a larger group of people.

This strategic complementarity, in effect, implies a "cost" to members of a community when some members of the community decide to coordinate with the members of another group. To see this in terms of our earlier examples, consider first the case of someone from a rural community leaving permanently for the city. When this person leaves, the remaining members of the community may be left worse off. A parent may have one person fewer to help manage the family farm and business owners in town now have one consumer fewer. Similarly, when a member of an immigrant or ethnic group assimilates into another group in society, the group members who remain are left, in effect, with one member fewer to interact with. This may reduce their economic community's size and/or it may reduce the group's political power by rendering it a smaller group. Numerous such costs are possible. These costs are discussed, implicitly, in Rubin (1995) and Wilson (1980).

By formally modeling the coordination component of interaction, rather than relying solely on anecdotal evidence, we are able to more clearly and precisely understand the relationship between this need for coordination and the standards of behavior which people choose. Further, we can explicitly study the welfare effects of this relationship. In so doing, we find that in some cases, when a group's opportunities for trade (or interaction in general) expand to include the possibility of trading with members of another group, but with coordination required, a community can end up worse off than before these added trading opportunities arose.

The model we suggest has the following characteristics. We begin by assuming that the gains from trade are higher if the two agents trading use the same standard of behavior. Each agent chooses such a standard from an exogenously given set of standards before he is randomly matched with a trading partner. Once matched with a trading partner, if the agents have chosen the same standard, then they may trade and will naturally do so if a beneficial trade exists. If they have not chosen the same standard, then trade may be possible, but agents incur a significant cost to doing so due to the lack of coordination. We normalize this payoff to zero in the model.

All members of society belong to one of two distinct communities, with members of one community preferring one standard while members of the other community prefer another. Further, the costs associated with adopting a less preferred standard differ among agents within a community.

As a reference point, we begin with the situation in which agents trade only with members of their own community. In this situation we have in mind that there is effectively a barrier between the two communities in the sense that a member of one community simply does not come into contact with a member of the other community. We assume that each community is in the stable equilibrium in which all members choose the standard preferred within that community. We then examine what happens when this "barrier" is lifted, expanding the trading opportunities for everyone.

Using this framework, we find that in some cases, when we account for these costs associated with coordination, total welfare of a minority community will decrease when the barrier between the communities is lifted. In these cases, some members of the minority community choose to coordinate their behavior with that being used by the members of the majority community. The remaining members of the minority community find the cost of doing this to be too high, and thus they continue to use their preferred standard. While the members who have switched their standards of behavior may be better off than they were in the autarky case (though not all necessarily will be), all community members who continue to use the initial standard are worse off due to the strategic complementarity we mentioned earlier. This loss may outweigh the gains achieved by those who switched to the majority standard. This logic differs from that found in Hart (1975) and in the literature on customs unions, both of which also examine cases in which a welfare loss can result upon the lifting of a trade barrier.[1] Further, this result is in sharp contrast to the most traditional results which state that when two countries lift a trade barrier, the welfare of both countries will at least weakly increase.

Further, we find that even in some cases where total welfare increases, agents in a minority community may choose to change their standard of behavior, or "assimilate," even though they are worse off assimilated into this larger community than they were before the expansion of trade opportunities ever took place. Again, the driving force here is the strategic complementarity of the situation.

[1] Hart (1975) examines a case involving incomplete markets and shows that the addition of an asset which allows trade between some of the markets, but not all of them, can lead to a decrease in welfare. In the literature on customs unions, it is shown that when a country has two trading partners, one with high costs and the other with low costs, and this country forms a customs union with the higher cost partner, while imposing a high tariff on the low cost partner, then a decrease in welfare is also possible.

We also look briefly at an extension of this model, in which agents may choose from a set of three standards. We offer an example of a case in which, beginning from the autarkic situation in which all agents use their most preferred standard, all agents in both communities change their standard to the standard which is not the most preferred standard of either community when the barrier between the communities is lifted. In this example, both communities, and more specifically all members in both communities, are left worse off.

The remainder of the chapter is organized as follows. The formal model is laid out in Section 2, while a description of equilibrium and dynamics can be found in Section 3. An analysis of the welfare implications which result from our accounting for the cost of coordination makes up Section 4. The extension of the model in which agents choose from three standards is discussed in Section 5, and we conclude in Section 6.

2.2 The Model

We consider an exchange economy consisting of two types of agents, type x and type y. The sizes of the type x group and that of the type y group are equal. There are two commodities, x and y. A type x agent is endowed with 2 units of good x, while a type y agent is endowed with 2 units of good y. Each type x agent is indexed by a number in $(0, 1)$, as is each type y agent. Agents get zero utility from consuming only one good, while they get positive utility from consuming a unit each of x and y. In this world, trade takes place in the form of a one-for-one swap of goods. When agents trade, they need to coordinate on a behavioral standard, either L or R, for the benefit to trade to be positive.

Each agent belongs to one of two communities, A or B. Agents in A strictly prefer coordination on standard L, while agents in B strictly prefer coordination on R. However, coordination on some standard is preferred to miscoordination by all agents. If an x-endowed agent i, from community A, uses standard L and trades with another agent using L , the agent receives a utility level which we normalize to

1, while if he uses standard R and trades with another agent using R, he receives utility $\mu_R(i) \in (0,1)$. Similarly, if an x-endowed agent j from B uses standard R and trades with another agent using R, the agent receives utility 1, while if he uses standard L and trades with another agent using L, he receives utility $\mu_L(j) \in (0,1)$. Symmetrically, a y-endowed agent i in A receives utility 1 if he uses standard L and trades with another agent using L, while he receives utility $\nu_R(i) \in (0,1)$ if he chooses standard R and trades with another agent using R. A y-endowed agent j in B receives utility $\nu_L(j) \in (0,1)$ if he uses standard L and trades with another agent using L, while he receives 1 if he uses R and trades with another agent also using R.[2]

We denote the fraction of the x-type agents who are in community A by n_x and the fraction of the y-type agents who are in community A by n_y. Thus, the fraction of the total population which is in community A is $\frac{n_x+n_y}{2}$. We arrange x-type players uniformly on the line [0,1] such that the following is true. For $i \leq n_x$, $0 < \mu_R(i) < 1$ and $\mu_R(i)$ is monotonically increasing. For $i > n_x$, $0 < \mu_L(i) < 1$ and $\mu_L(i)$ is monotonically decreasing. This places players in community A on the first portion of the line and those in B after them. Also, note that this is equivalent to placing the players who have the most trouble changing their standard of behavior at the outer ends of the line, while placing those who would find it least difficult to make this change nearest the members of the community to which they do not belong. In the same manner, we arrange y-type players on the line [0,1] such that the following holds. For $j \leq n_y$, $0 < \mu_R(j) < 1$ and $\mu_R(j)$ is monotonically increasing. For $j > n_y$, $0 < \mu_L(j) < 1$ and $\mu_L(j)$ is monotonically decreasing.

There is no centralized market where agents can meet to exchange commodities. Rather, agents are randomly matched into pairs. We examine two cases. We will refer to the first of these cases as the "autarky" case, and it will serve as a benchmark for analyzing the results we obtain in the second case. In the autarky case, agents

[2]Note that by defining μ and ν as we do, our model can also be interpreted to address situations in which agents from one community can interact profitably with members of the other community, but they may not be able to fully reap the benefit of the interaction in the way that the agents using their most preferred standard can.

	A	B
A	$\frac{n_x+n_y}{2}$	0
B	0	$\frac{2-n_x-n_y}{2}$

Figure 2.1: Matching Technology: Autarky Case

	A	B
A	$\frac{n_x+n_y}{2}$	$\frac{2-n_x-n_y}{2}$
B	$\frac{n_x+n_y}{2}$	$\frac{2-n_x-n_y}{2}$

Figure 2.2: Matching Technology: Unification Case

are matched only with agents from their own community according to the matching technology enumerated in Figure 2.1. In other words, an agent in community A meets a trading partner with probability $\frac{n_x+n_y}{2}$, while with probability $\frac{2-n_x-n_y}{2}$, the agent meets no one. Similarly, an agent in community B meets a trading partner with probability $\frac{2-n_x-n_y}{2}$ and meets no one with probability $\frac{n_x+n_y}{2}$.

The second case we examine is the "unification" case. Here, agents may meet trading partners from either community, according to the matching technology of Figure 2.2. In this case, an agent will meet a trading partner with probability 1. Recall, however, that this is not equivalent to saying that an agent will trade with probability 1. Agents in a pair must use the same standard of behavior and have different commodities for trade to actually occur.

2.3 Equilibrium and Dynamics

We consider two situations. The first is that of an autarky, which serves as a benchmark for the second situation in which there is no barrier in the sense of uniform random matching.

2.3.1 Autarky

If there is a barrier between the two communities such that no agent can meet an agent from the other community, we can analyze each community separately. We consider community A. The analysis for community B is symmetric. First, any pure strategy equilibrium, i.e. a situation from which no one has an incentive to deviate, can be characterized by two numbers, $m_x \in [0, n_x]$ and $m_y \in [0, n_y]$, where agent i of type x (resp. type y) takes standard L if and only if $i < m_x$ (resp. $i < m_y$). [3] Indeed, if $i < j$, then $\mu_R(i) \leq \mu_R(j)$ and $\nu_R(i) \leq \nu_R(j)$ and therefore, if player i takes R in an equilibrium, player j of the same type weakly prefers R to L as well.

The agent i of type x obtains $\frac{m_y}{2}$ if he takes L and $\frac{1}{2}(n_y - m_y)\mu_R(m_x)$ if he takes R. Therefore, his incentive conditions are given by

$$m_y > (n_y - m_y)\mu_R(i) \qquad \text{if } i < m_x, \tag{2.1}$$

$$m_y < (n_y - m_y)\mu_R(i) \qquad \text{if } i > m_x. \tag{2.2}$$

Inequalities (2.1) and (2.2) give the incentive curve for type x agents, i.e. the curve on which no agent of type x has an incentive to deviate. Similarly, for an agent i of type y, we have

$$m_x > (n_x - m_x)\nu_R(i) \qquad \text{if } i < m_y \tag{2.3}$$

$$m_x < (n_x - m_x)\nu_R(i) \qquad \text{if } i > m_y \tag{2.4}$$

[3] This is unique up to permutation among those with the same μ's. Also, we ignore the action taken by the agent at the threshold, i.e., agent m_x.

Inequalities (2.3) and (2.4) jointly give the incentive curve for type y agents. The intersections of these two curves determine the equilibria of this community. Since $\mu_R(m_x)$ and $\nu_R(m_y)$ are functions of m_x and m_y, respectively, we can draw equilibrium conditions on a (m_x, m_y)-plane. Two examples are given in Figure 2.3. The first illustrates the case in which μ and ν are distributed uniformly, and the second where the distributions are made up of two mass points. As is always the case, there are multiple equilibria, including one at (0,0) and another at (n_x, n_y). Yet, if we use the theory of evolution, we can identify $(m_x, m_y) = (n_x, n_y)$ as the unique "stable" equilibrium. More precisely, it is the unique stochastically stable equilibrium according to Foster and Young (1990), the unique long run equilibrium according to Kandori, Mailath and Rob (1993), and the unique globally absorbing state according to Matsui and Matsuyama (1995). Further, the equilibrium (n_x, n_y) is stable in this sense for all distributions of μ and ν. When we discuss dynamics after the barrier is removed, we assume, therefore, (n_x, n_y) as the initial condition in community A. By the same token, we assume that (n_x, n_y) is the initial condition in community B. This will simplify our dynamics as well.

2.3.2 Unification

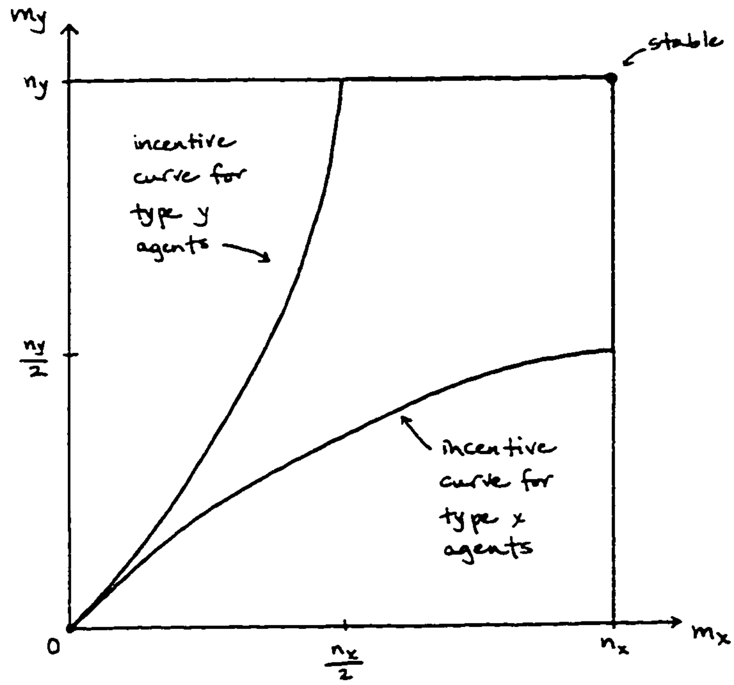
The analysis of the case with no barrier is similar to that of the autarky. As before, an equilibrium is essentially characterized by two thresholds, m_x and m_y. Now, the x-type agent at m_x obtains $m_y\mu_L(m_x)$ if he takes L, and $(1 - m_y)\mu_R(m_x)$ if he takes R. Therefore, the first condition becomes

$$m_y\mu_L(i) > (1 - m_y)\mu_R(i) \qquad \text{if } i < m_x \tag{2.5}$$

$$m_y\mu_L(i) < (1 - m_y)\mu_R(i) \qquad \text{if } i > m_x. \tag{2.6}$$

Similarly, the second condition is given by

$$m_x\nu_L(i) > (1 - m_x)\nu_R(i) \qquad \text{if } i < m_y \tag{2.7}$$



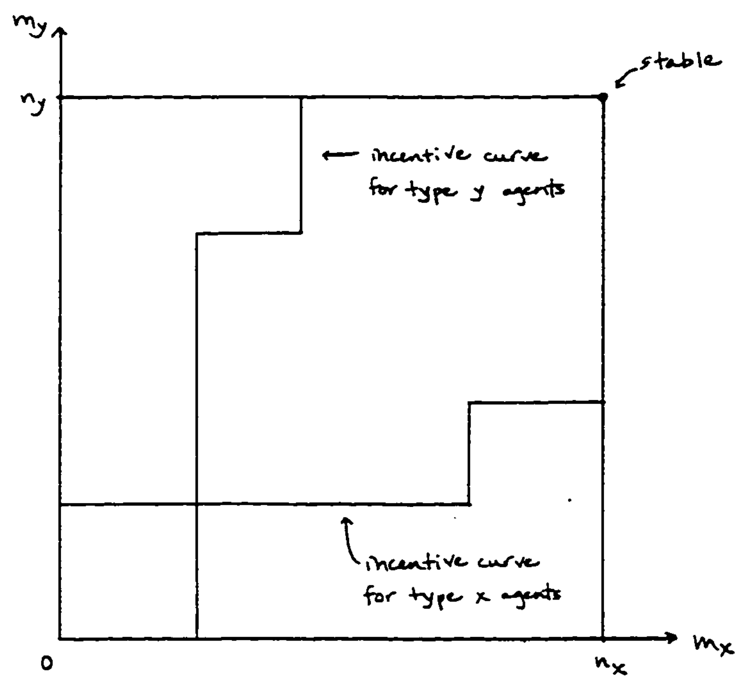


Figure 2.3: Equilibrium: Autarky Case

$$m_x \nu_L(i) < (1 - m_x)\nu_R(i) \qquad \text{if } i > m_y. \tag{2.8}$$

As before, (2.5) through (2.8) jointly determine the incentive curves and hence equilibrium.

2.3.3 Dynamics

We use best response dynamics to select equilibrium, which makes our results conclusive. A *best response dynamic* is a dynamic in which agents gradually adjust their actions to a best response action to the current strategy profile. Such a slow adjustment process is appropriate to our problem, since cultural adjustment traits change only slowly. (See, for example, Cavalli-Sforza and Feldman (1981).) Thus, we assume that time is continuous, as is the dynamic path. For the sake of simplicity of the analysis, we assume further that if many agents have incentive to switch their actions, those who have greater incentive than others switch first. This enables us to characterize the state of the dynamical system by two thresholds m_x and m_y if the initial condition is also expressed by two thresholds. If (m_x, m_y) satisfies

$$m_y \mu_L(m_x + \varepsilon) > (1 - m_y)\mu_R(m_x + \varepsilon)$$

for some $\varepsilon > 0$, then type x agent $m_x + \varepsilon$, who is now taking R, prefers L to R, and m_x increases. Similar conditions are applied to other cases, which enables us to draw a phase diagram. Examples are given in Figure 2.4. The stable equilibrium shown in Figure 2.4a represents what we refer to as a *completely assimilated equilibrium*, (0,0), meaning that in this equilibrium, all members of one community have coordinated with, or assimilated into, the other community. We refer to the stable equilibria of Figures 2.4b and 2.4c as *partially assimilated equilibria*, which are given by (m_x, m_y) with $(m_x, m_y) \in [0, n_x] \times [0, n_y] \setminus \{(0,0), (n_x, n_y)\}$. Here, some members of one community have coordinated with the members of the other community, but not all members have done so.

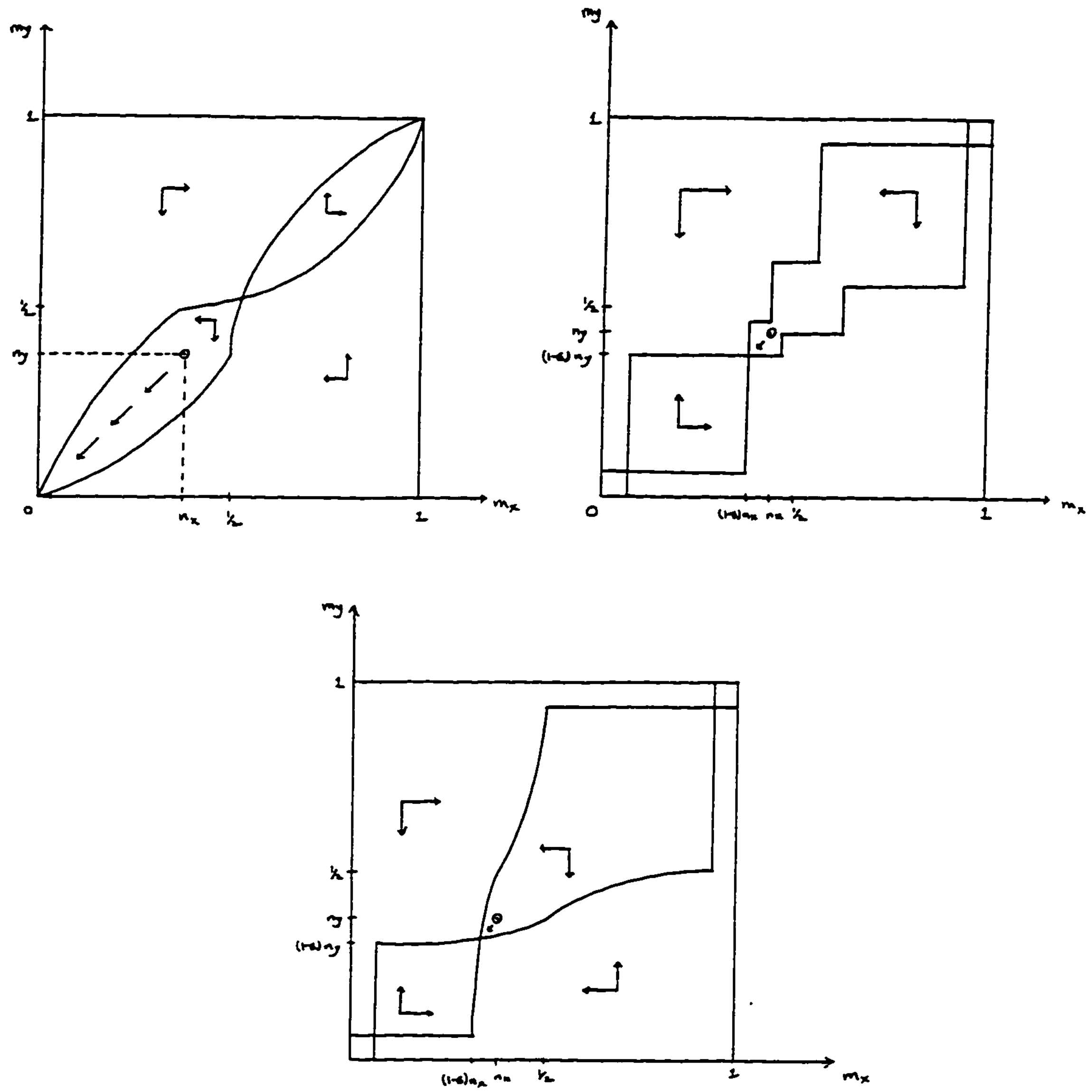


Figure 2.4: Equilibrium with Best Response Dynamics: Unification Case

We fix the initial condition at (n_x, n_y). Then in each case, there is the unique equilibrium which is accessible from the initial condition under best response dynamics irrespective of the relative speed of adjustment for two types of agents.

We now classify situations into two cases. In the first case, the equilibrium which is uniquely accessible from (n_x, n_y) is a completely assimilated equilibrium. In the second, this uniquely accessible equilibrium is a partially assimilated equilibrium.

If the two incentive curves obtained from inequalities (2.5) through (2.8) coincide in the box $[0, n_x] \times [0, n_y]$, i.e., if μ_R and ν_R satisfy

$$\frac{x}{1-x} = \nu_R\left(\frac{\mu_R(x)}{1+\mu_R(x)}\right), \quad \forall x \in (0, n_x), \tag{2.9}$$

then there are a continuum of equilibria. Such a pair of incentive curves plays an important role in determining whether the equilibrium which is uniquely accessible from (n_x, n_y) is a completely assimilated equilibrium or a partially assimilated equilibrium. Let a pair of functions μ_R^* and ν_R^* satisfy (2.9).[4] Consider another pair μ_R and ν_R. If we have

$$\mu_R(i) > \mu_R^*(i), \ \forall i \in (0, n_x),$$

and

$$\nu_R(i) > \nu_R^*(i), \ \forall i \in (0, n_y),$$

then (0,0) is the unique equilibrium which is accessible from (n_x, n_y) since the only intersection of the two curves below (n_x, n_y) is (0,0). That is, if agents in community A do not dislike R too much, they completely assimilate into the larger community. On the other hand, if there exists an interval $(\underline{m}, \bar{m}) \subset (0, n_x)$ with positive length such that $\mu_R(i) < \mu_R^*(i)$ for all $i \in I$, and if $\nu_R(i) \leq \nu_R^*(i)$ for all $i \in (0, n_y)$, then the two curves intersect at (m_x, m_y) for some $m_x \in [\bar{m}, n_x)$. Therefore, the system either stays at the initial situation or converges to a partially assimilated equilibrium.

[4]There are infinitely many such pairs.

2.4 Welfare Implications

We now turn to the welfare implications of our accounting for the costs of coordination. We study two explicit examples.

In all of our examples, we restrict our attention to the situation where n_x and n_y are less than $\frac{1}{2}$. Thus, in the stable equilibrium which is accessible from the initial condition (n_x, n_y) given by the "autarky case," it is members of community A who will switch from their most preferred standard if anyone does.[5] Therefore, we need only refer to the distributions of μ_R and ν_R; the distributions of μ_L and of ν_L do not factor. Also for convenience, we normalize the size of the population endowed with x to be 1, and thus the size of the population endowed with y to be 1 as well.

We first examine the case in which μ_R and ν_R are uniformly distributed on (0.1) across the (respective) agents in A. We then turn to essentially the other end of the spectrum by examining the case in which μ_R and ν_R each take on one of two values, $\underline{\mu}_R$ and $\bar{\mu}_R$ and respectively $\underline{\nu}_R$ and $\bar{\nu}_R$, and a positive mass of players incurs each level of cost.

We use a simple welfare function in which the utility of each member in the community is given equal weight. Thus, we define total welfare for community A to be the sum of the total welfare of A's type x agents, W_x, and the total welfare of the type y agents, W_y as follows.

$$W_x = \int_0^{n_x} Eu_i \, di \tag{2.10}$$

$$W_y = \int_0^{n_y} Eu_i \, di. \tag{2.11}$$

where Eu_i is the expected utility of an agent i of the respective type. Using these functions, we find that community A's total welfare in the autarky case is equal to

[5] Note that this is a sufficient condition for members of A to be the only agents who switch in the abovementioned equilibria, but not a necessary condition.

$\frac{1}{2}n_x n_y + \frac{1}{2}n_x n_y = n_x n_y$. Again, this will serve as our benchmark level of welfare for the discussions which follow.

2.4.1 Example 1: Uniform Distribution

Let $\mu_R(i)$ and $\nu_R(i)$ be distributed uniformly on (0,1). Then $\mu_R(i) = \frac{i}{n_x}$ and similarly, $\nu_R(i) = \frac{i}{n_y}$. We know from section 2.3 that the stable equilibrium accessible from our initial (autarky) condition (n_x, n_y) is the equilibrium in which all members of A adopt their less preferred standard, L, while members of B continue to use L.

We make our calculations for type x agents only. The results are symmetric for the type y agents. Total welfare for the type x members of community A at the equilibrium point is expressed here by

$$W_{x,u} = \frac{1}{2n_x}\int_0^{n_x} i \; di, \tag{2.12}$$

which yields a level of welfare in this equilibrium equal to $\frac{n_x}{4}$. Given our restriction that n_x and n_y are less than $\frac{1}{2}$, $W_{x,u}$ is clearly greater than this group's level of welfare, $\frac{1}{2}n_x n_y$, in the autarky case. Thus, the expansion of trade opportunities has in this case increased the community's total welfare using this measure of welfare.

This does not mean, however, that all members of the community are better off. And we can readily see that they are not. More specifically, all type x members of A with cost multiplier $\mu_R(i) < n_y$ are now strictly worse off than they were before the lifting of the barrier. Why, then, do they change their standard? They change their standard because of the negative externality imposed upon them when the members of their community with low switching costs change standards. As a result of this externality, they find themselves in a position where clinging to their preferred standard leaves them even worse off than they will be if they change their standard. This is a condition which does not arise when only standard transaction costs are involved in trade.

	L	R
L	$\frac{n_x+n_y-\varepsilon(n_x+n_y)}{2}$	$\frac{2-n_x-n_y-\varepsilon(n_x+n_y)}{2}$
R	$\frac{n_x+n_y-\varepsilon(n_x+n_y)}{2}$	$\frac{2-n_x-n_y+\varepsilon(n_x+n_y)}{2}$

Figure 2.5: Probabilities in Equilibrium: Two Mass Points Example

2.4.2 Example 2: Two Mass Points

Now we turn to the example in which all members of the community have one of a small set of cost multipliers. We assume that a fraction ε of the x-endowed members of A have cost $\mu_R(i) = \bar{\mu}_R$ while the remaining $1 - \varepsilon$ of this population have cost $\mu_R(i) = \underline{\mu}_R$, where $\bar{\mu}_R > \underline{\mu}_R$. Similarly, a fraction ε of the y-endowed members of A have cost $\nu_R(i) = \bar{\nu}_R$ while the remaining agents of this type have cost $\nu_R(i) = \underline{\nu}_R$, where $\bar{\nu}_R > \underline{\nu}_R$.

We assume that $\underline{\mu}_R$, $\bar{\mu}_R$, $\underline{\nu}_R$, and $\bar{\nu}_R$ are such that, in equilibrium, players with costs equal to either $\bar{\mu}_R$ or $\bar{\nu}_R$ will switch to using standard R, while players with costs $\underline{\mu}_R$ or $\underline{\nu}_R$ will continue to use L. Specifically, this means that $\bar{\mu}_R > \frac{n_y}{1-n_y}$ and $\bar{\nu}_R > \frac{n_x}{1-n_x}$, while $\underline{\mu}_R < \frac{(1-\varepsilon)n_y}{1-(1-\varepsilon)n_y}$ and $\underline{\nu}_R < \frac{(1-\varepsilon)n_x}{1-(1-\varepsilon)n_x}$. From these conditions, we see that for a given $\bar{\mu}_R$ and $\underline{\mu}_R$, there is a range of n_y for which a partial equilibrium exists. An equivalent statement can be made for $\bar{\nu}_R$ and $\underline{\nu}_R$. When these conditions are satisfied, we can characterize the probabilities, in equilibrium, that an agent will meet a trading partner using L or R by those in Figure 2.5.

Given this, the total welfare of A's type x agents in equilibrium is now described by

$$W_{x,m} = \frac{1}{2}\int_0^{(1-\varepsilon)n_x}(1-\varepsilon)n_y)\;di \;+\; \frac{\bar{\mu}_R}{2}\int_{(1-\varepsilon)n_x}^{n_x}(1-(1-\varepsilon)n_y)\;di \tag{2.13}$$

which yields a welfare level equal to

$$W_{x,m} = \frac{1}{2}(1-\varepsilon)^2 n_x n_y \;+\; \frac{\varepsilon}{2}\{\bar{\mu}_R[1-(1-\varepsilon)n_y]n_x\}.$$

This expression, while somewhat messy, is readily interpretable. The first term in the expression represents the utility level of the community members who continue to use L. This mass of members, in the autarky case, would have received welfare level $\int_0^{(1-\varepsilon)n_x}\frac{n_y}{2}\;di \;=\; \frac{1}{2}(1-\varepsilon)n_x n_y \;>\; \frac{1}{2}(1-\varepsilon)^2 n_x n_y$. These members have suffered a welfare loss. This is, of course, the direct result of the negative externality imposed upon them when the $\bar{\mu}_R$ and $\bar{\nu}_R$ members of their community switch to R.

On the other hand, for their incentive constraint to have been satisfied, the agents with $\bar{\mu}_R$ and $\bar{\nu}_R$ must have experienced a welfare gain. This is easy to verify. The second term in the above expression represents the new level of welfare which these agents receive. Previously, again referring to the autarky case, they received $\int_{(1-\varepsilon)n_x}^{n_x}\frac{n_y}{2}\;di \;=\; \frac{1}{2}\varepsilon n_x n_y$ which, given our initial restrictions on $\bar{\mu}_R$ and $\bar{\nu}_R$ is strictly less than their new level of welfare.

We now ask whether or not the welfare gain experienced by the agents using R outweighs the welfare loss incurred by those continuing to use L. We will look specifically at the case where $\bar{\mu}_R = \bar{\nu}_R = 1$, since if the inequality holds under this condition, it will certainly hold in the case where $\bar{\mu}_R$ and $\bar{\nu}_R$ are less than 1, as in this case, the welfare gain experienced by the gaining agents is diminished. As before, we look first at the net change in welfare which the type x agents experience, and then we can examine, separately, the type y agents. If the type x agents have incurred a net welfare loss, the following inequality will hold.

$$\varepsilon(1-\varepsilon)n_x n_y > \varepsilon n_x[1-(1-\varepsilon)n_y] - \varepsilon n_x n_y.$$

which gives us the condition

$$n_y > \frac{1}{3-2\varepsilon}$$

Since the relevant inequality for the y-type agents is symmetric, we can conclude that the following condition will also hold if the y -type agents experience a net welfare loss.

$$n_x > \frac{1}{3-2\varepsilon}$$

Again, these conditions are sufficient but if $\bar{\mu}_R$ and/or $\bar{\nu}_R$ are strictly less than one, then weaker conditions will suffice. Either way, these conditions tell us immediately that for ranges of n_x and n_y, a net welfare loss may result with the expansion of trade opportunities when the costs of coordination are accounted for.

2.5 The Case of 3 Standards

We now consider the situation in which agents may choose from among three behavioral standards, namely L, C and R. We look at a fairly specific example. Let the payoff matrix for type x agents be that in Figure 2.6. We assume that a corresponding matrix applies to the type y agents. However, as before, we focus our welfare analysis on the type x agents since the analysis for the y agents is symmetric.

We assume that the communities A and B are equal in size, or more specifically that $n_x = n_y = \frac{1}{2}$. And we retain the assumption that the standard L is the most preferred standard by members of community A while R is most preferred by members of B. Thus, we retain the normalization that for type x members of A, $\mu_L = 1$, while

	L	C	R
L	μ_L	0	0
C	γ	μ_c	γ
R	0	0	μ_R

Figure 2.6: Payoff Matrix: The Case of 3 Standards

for members of B, $\mu_R = 1$. Further, we can retain the initial condition at which all members of A use L and all members of B use R. Note, however, that m_x and m_y as previous defined are no longer sufficient to characterize equilibrium here because of the addition of the third standard.

As mentioned in the introduction to the chapter, the welfare results in this case do not rely upon there being heterogeneity among agents with respect to μ_L and μ_R, though the results do hold for appropriate parameter values when heterogeneity is present. Thus, for ease of enumeration, we will assume that agents within a community are homogeneous in this regard. Further, we can assume that all agents in both communities earn the payoff γ when using C and trading with someone using L or R, while earning the payoff μ_C from using C and trading with someone using C.

We assume that $\gamma > max\{n_y, 1 - n_y\}$ and that if agents from A and B switch to standard C from their preferred standards, they do so at the same rate.[6] We use the remainder of this section to show that if

$$\gamma > \frac{1}{2} > \mu_C > \frac{1}{2}\gamma - \frac{1}{4}, \tag{2.14}$$

[6]One assumption which we could make that would make our assumption regarding agents switching at the same rate most intuitively appealing is the assumption that μ_R for agents in A is equal to μ_L for agents in B. However, as this assumption is in itself not necessary, we do not make it.

then the lifting of a trade barrier between A and B leads to a welfare loss for every individual in both A and B.

Given our assumptions, the stable equilibrium accessible from our initial condition, using best response dynamics, is the equilibrium in which all members of both A and B choose standard C. To see this, we first consider the decision of an agent at the initial point when the barrier is lifted. If the agent is a member of A, then taking L offers an expected payoff of $\frac{1}{2}n_y = \frac{1}{4}$, taking C offers $\frac{1}{2}\gamma$, and taking R offers $\frac{1}{2}\mu_R(1-n_y) = \frac{1}{4}\mu_R$. The agent's best response is clearly to choose C, given our assumption regarding the value of γ. The same argument holds for members of B. Thus, we expect some agents to switch to C.

Now, since agents from A and B switch to C at the same rate, we can say that at some fixed point in time, a fraction c of the agents in both groups are using C. Thus, in evaluating his options, an agent in A sees that his expected payoff equals $\frac{1}{2}(\frac{1}{2}-c)$ if he chooses L, $\frac{1}{2}[\gamma(1-2c)+\mu_C(2c)]$ if he chooses C, and $\frac{1}{2}\mu_R(\frac{1}{2}-c)$ if he chooses R. Since choosing R is clearly a dominated strategy, we need only assess the comparison between his choosing L and C. Doing so, we find that if the following equation holds, then an agent will still prefer C to L if the following inequality holds.

$$\gamma(1-2c) + \mu_C(2c) - (\frac{1}{2} - c) > 0$$

This equation will hold for all $c \in [0,1]$ if it holds for c=1. Thus, we find that if

$$\mu_C > \frac{1}{2}\gamma - \frac{1}{4}$$

then in equilibrium, all agents will stay with the choice C. In equilibrium, the payoff expected by every agent equals $\frac{1}{2}\mu_C$. If $\mu_C < \frac{1}{2} = n_y = 1 - n_y$, then the expected payoff to every agent is lower than it was at the initial condition. Therefore, we say that all agents in both communities experience a welfare loss upon the lifting of the trade barrier between the communities.

2.6 Concluding Remarks

We have highlighted the importance of explicitly considering the need for coordination in interactions when modeling economic behaviors where such coordination is required. We have shown that when we account for the costs of such coordination, there are cases in which total welfare of a minority community decreases when a trade barrier between the two communities is lifted. In addition, we offered an example which illustrates that in a situation where no dominant culture exists, every member of both communities may ultimately be worse off upon the lifting of a trade barrier.

Several remarks are in order. First, we have not in this work considered important intergenerational issues which are pertinent in any discussion of assimilation. It is often argued that one of the most serious problems associated with assimilation is the gap which arises between generations. Parents become alienated from their children and cannot pass on the wisdom they have inherited from generations of people that came before them. Children who wish to assimilate must learn the new culture on their own. They often remain second class citizens in the new society. This effect may persist, in some cases becoming intensified and while in others, becoming weaker. In cases where this effect becomes larger, the rate of economic growth may be higher for members of a dominant group in society than for those coming from a minority group. In addition to this problem, we have assumed that people make their choices myopically. We have not considered the case in which people take into account future generations when making their own decisions regarding assimilation.

Our next remark is related to our first. We do not presently deal with situations in which discrimination makes it essentially impossible for one group to coordinate with, or assimilate into, another group. This problem arises most commonly in cases when a group has some recognizable traits which cannot be changed, even by choice, such as gender or skin color. As the Folk Theorem has shown us, discrimination is sustainable in equilibrium even if the only difference between people is their "names." In such cases, it may be that members of one community would like to coordinate with the members of another group, but when they take the appropriate behaviors which would seemingly allow them to do so, they effectively end up as a group unto themselves, forced to interact primarily within the newly formed, third group.

As a final remark, different situations present different problems. For example, in the case of computer networks, standardization may imply the need for complete coordination. On the other hand, culture cannot be described by a single trait. (See Cavalli-Sforza and Feldman.) Adopting one trait but not another may have effects which we do not capture with the model in this chapter. Thus, we must more carefully examine the contents of such traits when we apply our analysis to specific problems. One typical question which must be addressed in this vein is the question or which traits can be changed and at what cost.

Finally, in the future, careful applied work is needed if we are to better understand the nature of the problems suggested by the present analysis. All of these issues are left to future research.

2.7 References

[1] Asimov, Stanley, ed. (1995), *Yours, Isaac Asimov: A Lifetime of Letters* (New York: Doubleday).

[2] Cavalli-Sforza, L. L. and M.W. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton, NJ: Princeton University Press).

[3] Cooper, Russell and Andrew John (1988), "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics* 103, pp. 441-463.

[4] Cose, Ellis (1993), *The Rage of a Privileged Class* (New York: HarperCollins Publishers).

[5] de Beauvoir, Simone de (1952), *The Second Sex* (New York: Alfred A. Knopf, Inc.), pp. 679-715.

[6] Farrell, Joseph and Garth Saloner (1985), "Standardization, Compatibility, and Innovation," *Rand Journal of Economics* 16 (Spring 1985), pp. 70-83.

[7] Foster, Dean and H. Peyton Young (1990), "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology,* vol. 38, pp. 219-232.

[8] Hart, Oliver (1975), "On the Optimality of Equilibrium When the Market Structure is Incomplete," *Journal of Economic Theory,* pp. 418-443.

[9] Kandori, M., G. Mailath and R. Rob (1993), "Learning, Mutation, and Long run Equilibria in Games," *Econometrica* 61: pp. 29-56.

[10] Matsui, Akihiko and Kiminori Matsuyama (1995), "An Approach to Equilibrium Selection," *Journal of Economic Theory* 65: pp. 415-34.

[11] Matsui, Akihiko and Masahiro Okuno-Fujiwara (1994), "Evolution and Interaction of Cultures," mimeo.

[12] Matsuyama, Kiminori (1991), "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *Quarterly Journal of Economics* 106 (May 1991), pp. 617-650.

[13] Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1989), "Industrialization and the Big Push," *Journal of Political Economy* 97 (October 1989), pp. 1003-1026.

[14] Rubin, Barry (1995), *Assimilation and Its Discontents* (New York: Times Books).

[15] Wilson, William Julius (1980), *The Declining Significance of Race: Blacks and Changing American Institutions* (Chicago: The University of Chicago Press).

Chapter 3

Representing Collective Bargaining Games by Two-Player Bargaining Models

3.1 Introduction

Bargaining is an important economic reality, and as such, economists have constructed numerous models of bargaining in an attempt to gain insight into the topic. These models range from the early axiomatic bargaining model of Nash (1950) through the vast array of strategic models beginning with Nash (1953) and including the first dynamic models of Stähl (1972) and Rubinstein (1982).

One feature common to virtually all of these bargaining models is that they involve two players, negotiating against each other, to decide how to divide some pie between them.[1] However, most of the important negotiations we are attempting to gain insight into by building these models are collective negotiations. That is, most of the negotiations involve groups of players negotiating against other groups of players

[1] In a few additional studies, attempts have been made to model situations in which $n > 2$ people are negotiating to divide the pie into n pieces. For references and a summary of some of the results, see Osborne and Rubinstein (1990), Chapter 3.

to divide the pie. Examples include everything from labor negotiations, in which a group of workers is negotiating against a team of management, to international treaty negotiations, in which entire nations of people are negotiating against other entire nations of people (or at least some governing body such as the United States Congress is the effective player in the game). To date, there has been an implicit assumption that our two-player models can be viewed as representations of collective bargaining situations as well as simply models of two-player negotiations.[2]

I accept the premise that we can represent collective bargaining situations by two-player bargaining models. However, we cannot do so blindly. My aim in writing this chapter is to better understand what exactly it means to represent a collective bargaining situation by studying two-player bargaining models. In so doing, I find that intuition alone does not allow us to intelligently propose outcomes for collective negotiations; instead, it will generally lead us astray. In the process, I also find some interesting results which enable a better understanding of phenomena we often see occurring in real bargaining situations.

The reason that we must be careful when proposing results for collective negotiations based on two-player bargaining models is the diversity of preferences among the members of negotiating groups. If groups consist of identical members, then the two-player representation for collective negotiations involving these groups is trivially applicable. However, members of groups involved in a negotiation are rarely identical in their preferences over the outcome of the negotiation. Often, there is commonality in what they aim to achieve in the negotiation, but generally, there are significant differences as well. For example, in a labor negotiation, members of a negotiating labor union all prefer higher wages to lower ones. However, the union members are likely to differ in their preferences over which nonwage benefits to sacrifice for higher wages, and likely in their preferences over how soon an agreement must be achieved as well.

[2] An important exception is Raiffa (1982). In this work, the nature of some of the problems specific to collective negotiations are pointed out, and interesting examples are discussed. Thus, this work serves as excellent background reading to this chapter.

In many important international negotiations, these differences can be even more extreme. When the United States negotiated with Canada and Mexico over NAFTA (or for that matter, any time the United States is involved in trade negotiations with foreign trading partners), members of Congress, who must ultimately agree to the terms of a treaty, all want terms which are most favorable to the United States. However, no one could argue over the claim that they differ in their preferences over how to achieve such an outcome.[3] Negotiations in the Middle East between Israel and its Arab neighbors are perhaps among the most well known examples of such differences within the negotiating groups.

Thus, if we set out to model these collective negotiations as two-player games, a crucial question which immediately arises is that of who these "two players" are. What preferences does a player representing a diverse group have? Without the answer to this question, we cannot begin to discuss collective bargaining using the two-player framework, as all of our results for bargaining games depend critically on the preferences of the bargaining players.

Suppose, as an example, that one of the groups in our collective negotiation votes under majority rule. That is, assume that any agreement which results from the negotiation must be agreed to by at least fifty percent of the group's members. Intuition might lead us to believe, then, that the median voter of the group would be the correct choice for our "representative" player, i.e. the player whose preferences should be used in modeling the negotiation as a two-player game.

This intuition is wrong. If the group members are allowed to choose, by majority vote, a representative from among themselves to negotiate on behalf of the group, assuming that when the group member actually bargains, he does so according to his own preferences, then, in most games, the group will not choose the median voter to represent them. Rather, they will choose a more extreme member of the group. How

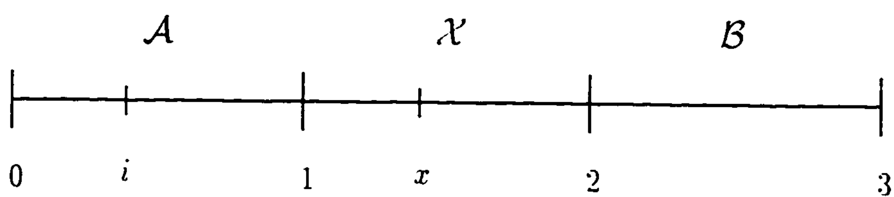
[3]The reasons for the differing preferences of members within a group are not our issue here. Thus, even if these differening "preferences" are the direct result of different beliefs over the parameters of the situation, they are considered here to simply be differing preferences. For the purposes of our exercise, this representation is most staightforward and qualitatively equivalent.

much more extreme depends on a number of things, most significant of which is the actual parameters of the bargaining game. The reasons for this will become clear as we study specific cases later in the chapter. However, it is worth pointing out that this result does, in fact, correspond with much of what we see in many real negotiations. Often, groups send someone more extreme than the middle to negotiate on their behalf. In particular, we certainly see this in the types of collective negotiations mentioned earlier, i.e. wage negotiations and international treaty negotiations of virtually every kind.

The chapter is organized as follows. In Section 2, I describe the basic model. Section 3 contains comparative statics results for the basic model in which players are completely informed about the parameters of the game they are playing and about their opponents. Sections 4 and 5 then discuss additional issues pertinent to this complete information case. In Section 6, I examine a case in which players are incompletely informed. Section 7 contains concluding remarks.

3.2 The Model

Two groups, denoted $\mathcal{A}$ and $\mathcal{B}$, are to bargain with each other over the location at which something is to be placed. Each group is made up of a continuum of expected payoff maximizing members located along a line segment. Group $\mathcal{A}$ is distributed uniformly on [0,1] while group $\mathcal{B}$ is distributed uniformly on [2,3]. I refer to the player located at i as player i. An offer is a point (or "location") on $\mathcal{X}$= [1,2]. The following diagram should help the reader visualize the setup.



Each player would like the agreed upon location to be as close to himself as possible. The payoff to player i if the agreement point is x is described by $\alpha - |x - i|$, where $\alpha \geq 0$ is a known constant. Disagreement results in a payoff of 0 to all players. Note that I do not restrict $\alpha - |x - i|$ to being greater than or equal to zero; it can be negative and thus, for some values of α, agreement can lead to a payoff worse than that caused by disagreement for some players.

The game is a two-stage game. In the first stage, each group chooses a negotiator to represent their group. In the second stage of the game, the chosen negotiators meet and negotiate over x based on their own preferences. The outcome of this negotiation determines all players' payoffs. Below, I lay out these two stages in detail.

Again, in the first stage of the game, groups choose negotiators to act on their behalf. Group $\mathcal{A}$ chooses a negotiator a_n while group $\mathcal{B}$ simultaneously chooses b_n. I assume that a_n must be in $\mathcal{A}$, i.e. in $[0,1]$, and symmetrically, that b_n must be in $\mathcal{B}$. That is, groups are choosing a member of their group to act as their representative. Both groups make this choice according to a majority rule vote. More specifically, agreement of fifty percent of players is required to appoint a particular group member as the group's negotiator.

I use the term *median voter* to refer to the voter, or group member, who is located halfway into the group. In this model, this equates to the players located at $a = 0.5$ and $b = 2.5$.

In the second stage of the game, the negotiators a_n and b_n will meet privately to bargain in an alternating offers framework. For simplicity, a_n will make the first offer. These negotiators negotiate based upon their own preferences, or equivalently, based upon their own locations. They have a common discount rate $\delta \in [0,1]$, which is a discount rate also common to all members of both $\mathcal{A}$ and $\mathcal{B}$. If they reach agreement on x, each player i receives a payoff $\alpha - |x - i|$, even if for a particular player, this payoff is negative. If the negotiators do not agree, all players accept the disagreement payoff of zero. No player can opt out of receiving the negotiation-determined payoff. I will focus on equilibria which are subgame perfect.

3.3 Comparative Statics

We now turn to our primary question, that of which group member should be considered the representative player for the group. Because the results vary as the value of α changes, I break the discussion into two sections.

3.3.1 α Large

Recall that the payoff to player i when the agreement point equals x is $\alpha - |x - i|$. I begin by considering the situation in which α is "very large," by which I mean that α is large enough such that every player receives a positive payoff for every possible agreement point. In our model, this is equivalent to saying that $\alpha \geq 2$. Later in this section, I consider the case in which α is still "large," but not very large, i.e. $\alpha \in [1.5, 2)$.

When $\alpha \geq 2$, in equilibrium, in the second stage of the game, agreement will be reached immediately. The negotiator making the first offer, in this case a_n, will offer a location x^* such that the payoff π^* to the B's negotiator, b_n, is

$$\pi^* = \max\left\{\frac{\delta(2\alpha + a_n - b_n)}{1+\delta},\ \alpha - (b_n - 1)\right\} \tag{3.1}$$

Thus, a_n will offer

$$x^* = \max\left\{\frac{\delta(2\alpha + a_n - b_n)}{1+\delta} - \alpha + b_n,\ 1\right\} \tag{3.2}$$

Player b_n will accept this initial offer, but will reject any offer in which x is lower.

Now, we can step back and assess whether a group fares better having a middle-of-the-pack negotiator or a more or less extreme one. Note that a more "extreme" negotiator in this model is a player located nearer the endpoints of the [0,3] line segment. These players are more difficult to satisfy in the negotiation stage of the game. Consider the following.

$$\frac{dx^*}{db_n} = 1 - \frac{\delta}{1+\delta} \tag{3.3}$$

Clearly, $\frac{dx^*}{db_n} \geq 0$ for all $\delta \geq 0$. Thus, as b_n increases, x^* increases. The higher the value of b_n, the better off are the members of $\mathcal{B}$. The same question can be addressed with respect to group $\mathcal{A}$, who is making the first offer. Here,

$$\frac{dx^*}{da_n} = \frac{\delta}{1+\delta} \tag{3.4}$$

$\frac{dx^*}{da_n} \geq 0$ for all $\delta \geq 0$. Thus, as a_n increases, x^* increases. As the members of $\mathcal{A}$ want x^* to be as small as possible, this means that the lower the value of a_n, the better off are the members of $\mathcal{A}$. So, regardless of whether a group is making the first offer or the second, it is in the interest of every member of that group to choose an extreme negotiator and hence, if the group votes on a negotiator, its members would (unanimously) agree to be represented by their most extreme player. This holds for all $\alpha \geq 2$. To summarize,

Theorem 1 *$\forall\alpha \geq 2$, $a_n^* = 0$ and $b_n^* = 3$ is the unique equilibrium in the first stage of the two-stage game defined above. In the second stage, agreement will be reached immediately on* $x^* = \max\left\{\frac{\delta(2\alpha+a_n-b_n)}{1+\delta} - \alpha + b_n,\ 1\right\}$.

Thus, in a simple two-player model of this situation, if groups can choose their representatives, the only way to capture this result is to assume that the two players are the most extreme players in their groups, i.e. the players located at $a = 0$ and $b = 3$.

If $\alpha < 2$, it is no longer the case that all players in both groups would prefer any agreement to no agreement at all. However, when α is less than but still close to 2, or more specifically, when $\alpha \in [1.5, 2)$, agreement can still be reached even if both groups are represented by their most extreme member. For α in this range, there are still *some* agreement points which all players would prefer to no agreement, and in equilibrium, one of these will be chosen. Here, the fact that α is lower effectively shrinks the set of feasible equilibrium offers from [1,2] to $[b_n - \alpha, a_n + \alpha]$, or in our case specifically, where a_n=0 and b_n=3, $[3 - \alpha, \alpha] \subset [1,2]$. The calculation of an equilibrium, however, is otherwise the same as the calculation in the case of $\alpha > 2$; with a_n making the first offer,

$$x^* = \max\left\{\frac{\delta(2\alpha + a_n - b_n)}{1+\delta} - \alpha + b_n, 3 - \alpha\right\}$$

and it is still the case that the more extreme the group's negotiator, the more favorable the equilibrium outcome will be for that group. Further, this will still be a unanimous choice.

3.3.2 α Small

We now turn to the case in which α is small, or $\alpha < 1.5$. When $\alpha \in (1, 1.5)$, agreement will no longer be reached in equilibrium if both groups choose their most extremely located members as representatives. The most extremely located players in each group cannot both, simultaneously, receive a payoff greater than the disagreement point from agreement on any point in $\mathcal{X}$. Yet, it will nevertheless be the case that a majority of the group's members would prefer certain agreements to disagreement.

In this case, in fact, there are a continuum of equilibria. More specifically,

Theorem 2 *For $\alpha \in (1, 1.5)$, there exist a continuum of equilibria. These equilibria are characterized by the first-stage choices*

$$b_n^* - a_n^* = 2\alpha$$
$$\text{such that } a_n^* \leq 0.5 \text{ and } b_n^* \geq 2.5.$$

In the second stage, agreement will be reached immediately on $x^* = a_n^* + \alpha = b_n^* - \alpha$.

Figure 3.1 is provided to enable us to visualize the relationships between equilibrium choices of negotiators and α and between equilibrium offers and α more clearly.

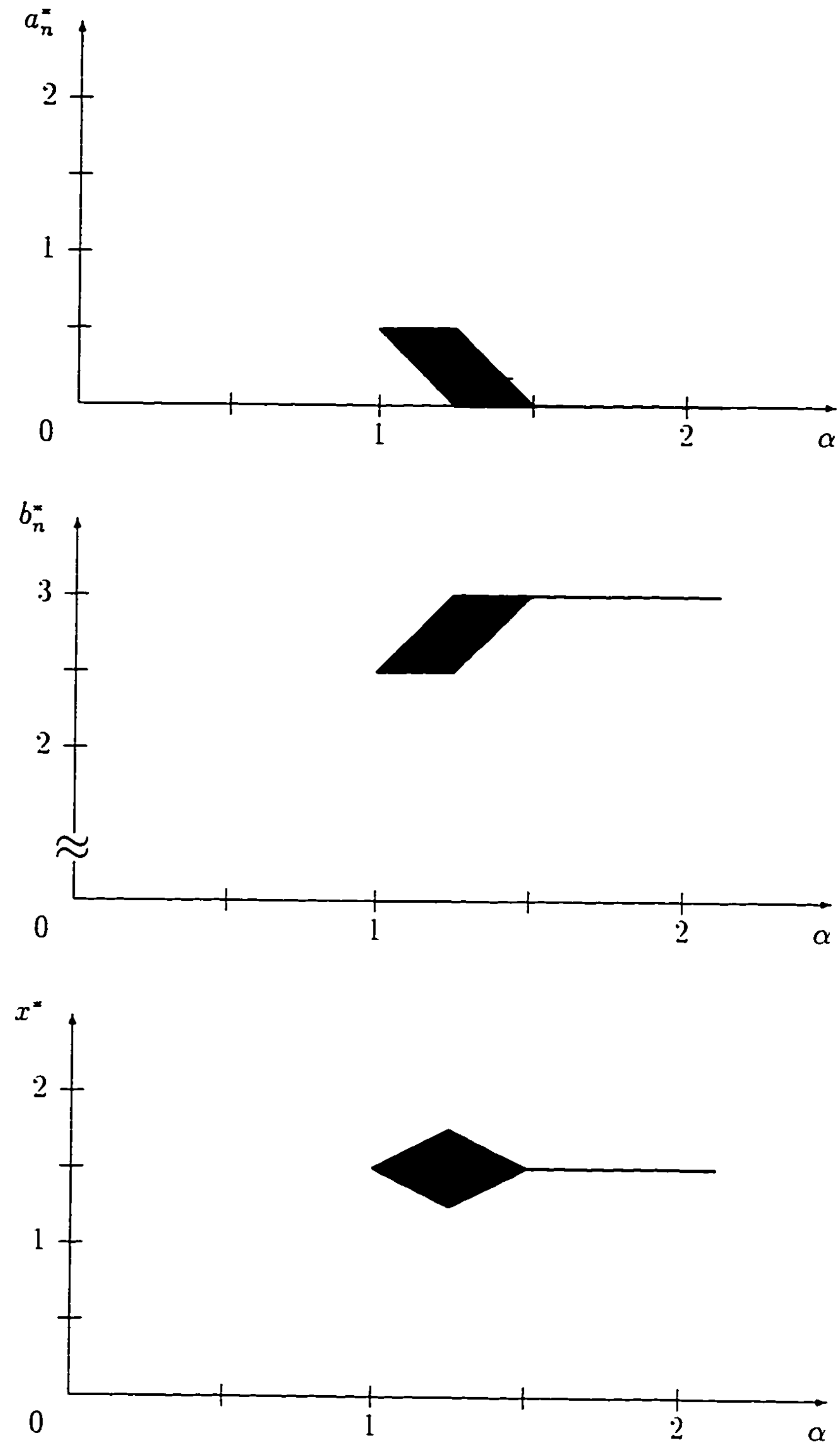


Figure 3.1: Equilibria Dependent Upon α.

This result is straightforward to verify. Note that in any of these equilibria in the case where α is small, there are players who are receiving a negative payoff as a result of the agreement on x. Thus, the choice of a negotiator in these cases is not a unanimous one. Rather, some of the most extremely located players would have voted to choose a more extreme negotiator, even if this resulted in no agreement in the second stage of the game.

This result brings to mind our general intuition that in games of complete information in which players bargain in an alternating offers framework, we have a unique equilibrium outcome. The implication of this result is that when we have a collective negotiation between groups whose members have differing preferences over the outcome of the negotiation, we may not be able to point to a unique equilibrium when considering the broader game in which these groups must first choose a negotiator. An alternative way to view this is that while, in a particular negotiation between two players (negotiators), we do have a unique equilibrium outcome. we have a continuum of two-player bargaining games which act as equilibrium representations of the collective bargaining situation we are interested in.

We can also see from Figure 3.1 that when $\alpha = 1$, there is again a unique equilibrium in which $a_n^* = 0.5$ and $b_n^* = 2.5$, while $x^* = 1.5$. I do not spend time on the case in which $\alpha < 1$, since in this case, no agreement is possible in the second stage of the game when at least fifty percent of players within each group much approve a negotiator choice in the first stage of the game. Even the median voters themselves, acting as representatives of their groups, could not reach agreement on any x, but rather would prefer disagreement.

3.4 Alternatives to Majority Rule

In the above discussion, it was assumed that groups vote using majority rule. I now turn to consider the effects that this voting rule has on our results and hence, the impact that other possible voting rules would have on our results.

I restrict my attention to the case in which α is small. As pointed out above, when α is large, groups unanimously choose their negotiator to be the most extreme player, regardless of the voting rule that the group uses to make its decisions. However, when α is small, i.e. when $\alpha \in [1, 1.5)$, there is disagreement among group members as to the choice of a negotiator. Even though a majority of the group's members can agree on a negotiator in equilibrium, not all members are voting in favor of this negotiator.

Let a group's voting rule be defined to be a point in [.5,1]. This point represents the fraction of players in the group who must vote in favor of a player for that player to be confirmed as the group's representative. Thus, a voting rule of 0.5 corresponds to majority rule, while a voting rule of 1.0 corresponds to the unanimity rule. Specifically, I denote group $\mathcal{A}$'s and group $\mathcal{B}$'s voting rules as $v_{\mathcal{A}}$ and $v_{\mathcal{B}}$, respectively.

The reader can verify that if a player located at i votes in favor of choosing a particular group member as the group's representative, then all players in i's group less extreme than i will vote in favor of this representative as well. Thus, in our discussion in Section 3, we were able assume that if a choice was confirmed by the median voter, it would be confirmed. Now, we can say the same of some critical voter in the group. Here, the term critical voter refers to the voter, or group member, who is located at the position $v_{\mathcal{I}}$ "into" group $\mathcal{I}$, where moving into the group is equivalent to moving away from $\mathcal{X} = [1,2]$. For example, if $\mathcal{A}$'s voting rule is $\frac{2}{3}$, then $\mathcal{A}$'s critical voter is located at $a = \frac{1}{3}$. And of course, a symmetric statement can be made about group $\mathcal{B}$.

Consider, for the moment, fixing group $\mathcal{B}$'s voting rule at $v_{\mathcal{B}} = 0.5$, or majority rule. We said previously that if $v_{\mathcal{A}} = 0.5$ also, then there are a continuum of equilibria which can be described as in Theorem 2. That is, all choices of negotiators by the two groups which satisfy $b_n^* - a_n^* = 2\alpha$ where $a_n^* \leq 0.5$ and $b_n^* \geq 2.5$ are equilibria of the first stage of our bargaining game.

Now, instead, suppose that group $\mathcal{A}$ has a voting rule $v_{\mathcal{A}} = \frac{3}{4}$. Now, agreement is possible only for $\alpha \geq 1.125$. That is, $\mathcal{A}$'s critical voter, located at $a = \frac{1}{4}$ cannot reach any agreement with $\mathcal{B}$'s critical voter $b = 2\frac{1}{2}$ for $\alpha < 1.125$. However, for $\alpha \in$

[1.125, 1.5), the set of equilibria, while still a continuum (for $\alpha > 1.125$), is reduced in size. Equilibria in the first stage of the game still satisfy the condition $b_n^* - a_n^* = 2\alpha$ but now must satisfy the conditions $a_n^* \leq \frac{1}{4}$ and $b_n^* \geq 2\frac{1}{2}$. The corresponding second stage equilibrium outcomes remain the same. This not only reduces the set of equilibria for games with α in this range, but does so in favor of group $\mathcal{A}$. Those equilibria which were least favorable to $\mathcal{A}$ are those that have been eliminated.

Thus, for a given value of α, holding v_b fixed at $v_b = 0.5$, group $\mathcal{A}$ fares best with its critical voter located at the position $\max\{2.5 - 2\alpha, 0\}$ or stated in terms of its voting rule, with

$$v_a = \max\{1 - (2.5 - 2\alpha),\ 1\} = \max\{2\alpha - 1.5,\ 1\}.$$

Not only does this reduce the set of equilibria for all $\alpha > 1$, but it in fact implies a unique equilibrium for all $\alpha \geq 1$. This unique equilibrium is equal to the equilibrium most favorable to $\mathcal{A}$ of the set of equilibria described by Theorem 2.

Again, however, recall that we fixed $v_{\mathcal{B}}$ in the above discussion. If we instead consider $v_{\mathcal{A}}$ and $v_{\mathcal{B}}$ to be variables which are also chosen by $\mathcal{A}$ and $\mathcal{B}$, respectively, we then revert to a situation in which there are a continuum of equilibria. Whereas previously, our equilibrium condition stated that the distance between the two groups' chosen negotiators would, in equilibrium, equal 2α, in this game the distance between the group's critical voters would equal 2α and the negotiator would be chosen to be the group's critical voter.

3.5 Additional Remarks on the Complete Information Case

The model presented above is, of course, a simplification of real negotiation procedures, chosen in part for its illustrative value. However, in many collective negotiations, while groups' chosen negotiators do get together privately to negotiate an agreement without the intervention of other group members, after these negotiators have come to an agreement between themselves, group members may have the opportunity to vote on whether or not to accept the agreement. For example, when the United States negotiates a trade agreement with another nation, a representative chosen by the U.S. government gets together with a representative chosen by the other nation's government. They negotiate secretly until both agree to terms for the treaty. However, before the treaty is actually enacted, the terms of the agreement are brought to the U.S. Congress for a vote. Only if Congress votes in favor of the terms is the treaty enacted; otherwise, the negotiators must begin to negotiate anew. To cover negotiations which occur in this pattern, only minor modifications to our model are required. I discuss the basic changes which are needed below, and express the equilibrium of this modified model. As can be expected, the equilibrium outcome is a function of not only a group's negotiator, but also explicitly of the group's voting rule. This in turn allows us to consider situations in which either one or both of the variables, negotiator and voting rule, are given exogenously. In the case of the U.S. Congress, for instance, it is often the President who chooses a negotiator to represent the U.S., while Congress must vote on the terms of the negotiated agreement based upon a predetermined voting rule.

Bargaining between the groups' negotiators takes place as before. However, now, once an agreement is reached by the negotiators, this agreement is presented to each group for a vote. If both groups vote to accept the agreement, the game ends with players receiving the payoff designated by this agreement. If, however, either group votes to reject this agreement, the negotiators resume bargaining, with the negotiator making the first offer alternating between a_n and b_n.

There are a number of possible assumptions which can be made regarding discount rates in such a game. One option is to assume that there is one discount rate which holds between rounds of the bargaining game taking place between the negotiators and that when agreement is reached, voting is instantaneous, so that this is the only discount factor taken into account. Alternatively, thinking about modeling real time events, one can assume that there are two separate discount rates, one which is faced between rounds of the negotiation, and another which is faced in the event of rejection of agreements between the actual negotiations themselves, with this second discount rate presumably being larger. For simplicity, I let there be a single discount factor.

The equilibrium outcome in this model is similar to that in our primary model. However, the groups' voting rules may now affect the outcome. In equilibrium, agreement will still be reached immediately. However, the location of x^* will depend on the relative locations of both the groups' critical voters and their chosen negotiators. I define a_v and b_v to be the locations of the critical voters in groups $\mathcal{A}$ and $\mathcal{B}$, respectively.

If $b_v > b_n$, then a_n will offer x^* such that the payoff to player b_v, $\pi^*_{b_v}$ is

$$\pi^*_{b_v} = \begin{cases} \max\left\{\frac{\delta(2\alpha+a_v-b_v)}{1+\delta}, \alpha-(b_v-1)\right\} & \text{if } a_v \leq a_n \\ \max\left\{\frac{\delta(2\alpha+a_n-b_v)}{1+\delta}, \alpha-(b_v-1)\right\} & \text{if } a_v > a_n. \end{cases}$$

If, on the other hand, $b_v \leq b_n$, then a_n will offer x^* such that the payoff to player b_n equal to $\pi^*_{b_n}$ where

$$\pi^*_{b_n} = \begin{cases} \max\left\{\frac{\delta(2\alpha+a_v-b_n)}{1+\delta}, \alpha-(b_n-1)\right\} & \text{if } a_v \leq a_n \\ \max\left\{\frac{\delta(2\alpha+a_n-b_n)}{1+\delta}, \alpha-(b_n-1)\right\} & \text{if } a_v > a_n. \end{cases}$$

Thus, a_n will offer

$$x^* = \begin{cases} \max\left\{\frac{\delta(2\alpha+a_v-b_v)}{1+\delta} - \alpha + b_v, 1\right\} & \text{if } a_v \leq a_n \text{ and } b_v > b_n \\ \max\left\{\frac{\delta(2\alpha+a_n-b_v)}{1+\delta} - \alpha + b_v, 1\right\} & \text{if } a_v > a_n \text{ and } b_v > b_n \\ \max\left\{\frac{\delta(2\alpha+a_v-b_n)}{1+\delta} - \alpha + b_n, 1\right\} & \text{if } a_v \leq a_n \text{ and } b_v \leq b_n \\ \max\left\{\frac{\delta(2\alpha+a_n-b_n)}{1+\delta} - \alpha + b_n, 1\right\} & \text{if } a_v > a_n \text{ and } b_v \leq b_n. \end{cases}$$

It is interesting to note that U.S. Congressional rules were set up such that Congress follows a majority rule voting rule when voting on most domestic issues, but for ratification of most international treaties, requires a vote of 2/3 of the members of Congress. This work implies that this puts the U.S. in a stronger negotiating position in treaty negotiations than would a majority rule vote for international treaty ratification. However, it can lead in some instances to disagreements in some such negotiations despite the fact that a majority of members would prefer agreement.

It is straightforward to see that we can also discuss cases in which groups choose a negotiator to act on their behalf in addition to, or instead of, a voting rule. For example, in some situations, it is considerably less costly to choose a negotiator than it is to choose a voting rule, given some previously defined constraints on the situation. For example, to change a voting rule which applies to the U.S. Congress is virtually impossible, while choosing a negotiator to act on behalf of the U.S. for any given negotiation is relatively simple. A very modest restructuring of the model would allow for this.

3.6 Negotiator Choice Under Incomplete Information

In our discussions to this point, we have assumed that players have complete information regarding all parameters of the game and regarding the preferences of their opponents. In this section, I relax one of these assumptions to discuss collective bargaining under incomplete information. To do so, I assume that players have incomplete information regarding the value of α when they make their choice of a negotiator in the first stage of the game.[4] Only after groups have chosen their representatives is the true value of α realized, so that the negotiators know the realization of α when bargaining in the second stage of the game.

Specifically, I assume that when players must make their choices of a negotiator for

[4]It is possible to instead relax the assumption that players have complete information about the preferences of their opponents. However, alternating-offers bargaining models of two-sided incomplete information do not, in general, have a unique equilibrium even in the true two-player case. Thus, it is more difficult to focus on our issue of interest, i.e., the choice of a representative player for a group of players with diverse preferences.

their group, they do not know the value of α. Rather, they know that α will be chosen by nature after this first stage of the game, but before the second stage. Further, α will be drawn from a uniform distribution on some interval $[\,\underline{\alpha}, \overline{\alpha}\,]$, with $\underline{\alpha}$ and $\overline{\alpha}$ known. Thus, when the chosen negotiators bargain in the second stage of the game, they will know the realization of α.

For simplicity, I return to the assumption made in Sections 2 and 3 that both groups vote under majority rule in the first stage. I also return to the simplest (initial) basic model, in which groups choose a negotiator in the first round and these negotiators bargain privately in the second round; the agreement they reach in this second stage will determine the payoff to all players. No further voting takes place.

We know from equation 3.2 that in equilibrium in the second stage of the game, a_n will make the initial offer x^* when $b_n - a_n > 2\alpha$ and α is high enough that agreement can be reached in equilibrium, where

$$x^* = \max\left\{\frac{\delta(2\alpha + a_n - b_n)}{1+\delta} - \alpha + b_n, 1\right\}$$

This offer will be accepted by b_n, hence leading to immediate agreement. To simplify our calculations, I assume that $\delta \approx 1$. We can then rewrite this equilibrium offer as

$$\begin{aligned} x^* &= \max\left\{\frac{(2\alpha + a_n - b_n)}{2} - \alpha + b_n, 1\right\} \\ &= \frac{(a_n + b_n)}{2} \end{aligned}$$

Hence, we can write the payoff to player $a \in \mathcal{A}$ as a function $f_a(\alpha, a_n, b_n)$

$$f_a = \begin{cases} \alpha - \frac{1}{2}(a_n + b_n) + a & \text{if } b_n - a_n \leq 2\alpha \\ 0 & \text{otherwise} \end{cases}$$

and that to a player $b \in \mathcal{B}$ as

$$f_b = \begin{cases} \alpha + \frac{1}{2}(a_n + b_n) - b & \text{if } b_n - a_n \leq 2\alpha \\ 0 & \text{otherwise} \end{cases}$$

We now step back to consider the point at which players must choose a negotiator to represent their group. At this stage of the game, players are uncertain as to the value of α, and thus they must make their decisions based upon the payoff they expect to receive after α is realized and the second stage negotiation takes place. Letting $g_i(\alpha, a_n, b_n)$ be the expected payoff to player i, we have for a player $a \in \mathcal{A}$ and a player $\in \mathcal{B}$, respectively

$$g_a = \begin{cases} E(\alpha) - \frac{1}{2}(a_n + b_n) + a & \text{if } b_n - a_n \leq 2\alpha \\ 0 & \text{otherwise} \end{cases}$$

$$g_b = \begin{cases} E(\alpha) + \frac{1}{2}(a_n + b_n) - b & \text{if } b_n - a_n \leq 2\alpha \\ 0 & \text{otherwise} \end{cases}$$

Since groups $\mathcal{A}$ and $\mathcal{B}$ make their choices of a negotiator simultaneously in the first stage of the game, we can appropriately take our game to be a normal form between the groups' median voters.

Equilibria in this game now depend on the values of both $\underline{\alpha}$ and $\overline{\alpha}$. To assess these equilibria, I break the distributions over $[\,\underline{\alpha}, \overline{\alpha}\,]$ into three classes. First, if $\underline{\alpha} \geq 1.5$, then regardless of the realization of α, agreement will be reached with probability 1 in equilibrium. Hence, both groups will choose their most extreme member as their negotiator in the first stage. Agreement will be reached on $x^* = 1.5$ immediately. Further, if $\overline{\alpha} < 1$, agreement will never be possible, and thus I ignore this case.

The most interesting case thus arises when $\underline{\alpha} < 1.5$ and/or $\overline{\alpha} \geq 1$. Then, group members must consider tradeoffs between appearing strong in the second stage negotiation and thus decreasing the probability with which agreement can be reached and increasing the probability of agreement while thus decreasing their negotiator's bargaining power. Said another way, as a group's negotiator becomes more extreme, the better will be the agreement point x^* when agreement is possible, but the lower will be the probability that agreement is reachable at all.

To get a feel for equilibria in this model within the abovementioned class of distributions, I first consider the class of distributions represented by $\alpha \sim U[\underline{\alpha}, 1.5]$. In Figure 3.2, the reaction curves of $\mathcal{A}$'s median voter $a = 0.5$ and $\mathcal{B}$'s median voter $b = 2.5$ for five values of $\underline{\alpha}$ are plotted. Again, since we are looking for equilibria of a normal form game between the two groups' median voters, this is appropriate. Specifically, the five diagrams illustrate, in order, the reaction curves in the cases when α is distributed on [1,1.5], [1.2,1.5], [1.25, 1.5], [1.3,1.5], and [1.4,1.5]. This enables us to clearly see the changes which occur as we change the value of $\underline{\alpha}$.

The intersection of these two reaction curves determines the equilibrium negotiator choices a_n^* and b_n^*. Thus, we can see that within this class of distributions, there is a unique equilibrium when $\underline{\alpha} \leq 1.25$. This holds for all $\underline{\alpha} \geq 0$. Further, the equilibrium values of a_n^* and b_n^* remain constant at the symmetric values $a_n^* = 0.25$ and $b_n^* = 2.75$. However, for $\underline{\alpha} \in (1.25, 1.5)$, there are, instead, a continuum of equilibria when $\underline{\alpha} \in (1.25, 1.5)$.

More generally, we can characterize the equilibrium results as follows.

Theorem 3 *Let $\overline{\alpha} \in (1, 2]$. Then for all $\underline{\alpha} \geq 0$ such that $2\underline{\alpha} \leq \overline{\alpha} + 1$, there is a unique equilibrium in which a_n^* and b_n^* are characterized by*

$$\begin{aligned} a_n^* &= 1.5 - (\overline{\alpha} + 1)/2 \\ b_n^* &= 1.5 + (\overline{\alpha} + 1)/2 \end{aligned}$$

with $x^* = 1.5$ when the realized value of α is greater than or equal to $(b_n^* - a_n^*)/2$.

Thus, when $\underline{\alpha}$ is small enough relative to $\overline{\alpha}$, we have a result which is highly different from that in the complete information case for α anywhere in the range $[\underline{\alpha}, \overline{\alpha}]$.

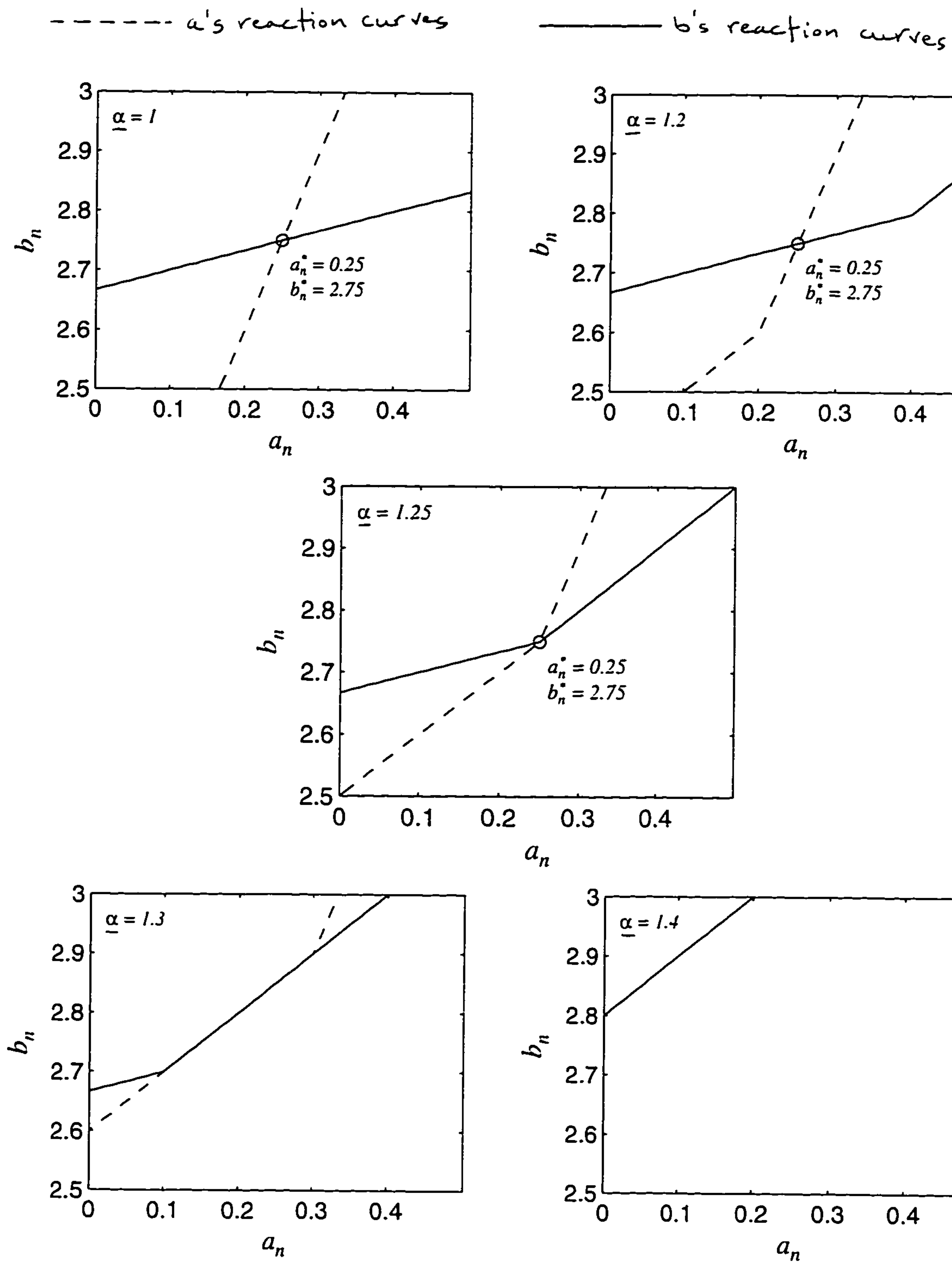


Figure 3.2: Reaction Functions When $\alpha \sim U[\underline{\alpha}, 1.5]$

Recall that under complete information, when $\alpha \in (1, 1.5)$, there are a continuum of equilibria, always. Here, with incomplete information, the level of uncertainty adds, intuitively, an increased awareness of the possibility that negotiators may be unable to reach agreement, even when agreement is desired by a majority of group members. It is worth noting that information regarding the expected value of α is insufficient for a characterization of the equilibria in such a game of incomplete information.

Note also that in this case, the median voters in both groups would receive a strictly higher expected payoff if both groups choose their median voter as their group's representative than that which they are receiving in equilibrium. Choosing the median voters as representatives increases the probability of reaching agreement while keeping the equilibrium agreement point the same. This is not, however, sustainable in equilibrium. The logic is that of the Prisoner's Dilemma game. Consider the median voter in $\mathcal{A}$. If the members of $\mathcal{B}$ choose their median voter to be the group's representative, this median voter in $\mathcal{A}$ has a strict preference to deviate from this "cooperative" outcome and choose a more extreme player as the representative for $\mathcal{A}$. The decrease in the probability of agreement is outweighed by the increase in payoff over the range of α for which such agreement is possible. A look back to Figure 3.2 will confirm this.

Of course, there is a range of outcomes to which Theorem 3 does not apply. Rather, we have the following.

Theorem 4 *Let $\overline{\alpha} \in (1, 2]$. Then for all $\underline{\alpha} < \overline{\alpha}$ such that $2\underline{\alpha} > \overline{\alpha} + 1$, there are a continuum of equilibria. These equilibria are an intermediate subset of the equilibria found in the complete information case for $\alpha = \underline{\alpha}$, characterized by $b_n^* - a_n^* = 2\underline{\alpha}$.*

In this case, $\underline{\alpha}$ is nearer $\overline{\alpha}$. Thus, we are, in a sense, moving closer to the complete information case. When $\underline{\alpha}$ is "close" but not "very close" to $\overline{\alpha}$, the set of equilibria under incomplete information is a strict subset of those found in the complete information game in which $\alpha = \underline{\alpha}$. Intuitively, players still find the uncertainty to be significant enough to make it worthwhile not to choose their most extreme players as negotiators.

However, when $\underline{\alpha}$ becomes "close enough" to $\overline{\alpha}$, the set of equilibria under incomplete information is equivalent to that under complete information. Returning again to Figure 3.2, and turning our attention specifically to the fifth and final diagram, we see that the equilibria of the game in which α is distributed uniformly on [1.4,1.5] are exactly those in the complete information game in which $\alpha = 1.4$. Here, the uncertainty is less severe, and hence, there are equilibria in which one group does choose their most extreme group member who could reach agreement with *someone* who is a candidate for the role of negotiator in the opponent group.

Thus, coming back to our main goal, it should be clear that moving between a collective bargaining game and its two-player representation is certainly possible, but not trivial. To choose a representative player for a group demands extreme care. There is no "intuitive" choice for a group representative in many such games.

3.7 Concluding Remarks

In this work, I have highlighted the need for paying careful attention to the collective nature of collective bargaining problems when using two-player bargaining models to represent them. When the members of a group involved in a collective negotiation have interests which vary along some dimension, the issue of determining which preferences should "represent" the group when the game is formulated as a two-player game cannot be decided based upon intuition alone. This problem arises both in games of complete information and games of incomplete information. In most cases, groups would prefer to be represented by someone with preferences more extreme than those of the member with median preferences. How much more extreme is a function of the parameters, rules, and informational constraints of the game being played.

Some comments are warranted. First, I have restricted my attention here to a very simple bargaining model. Yet, outcomes of bargaining models are highly dependent upon the rules of the game. Thus, for bargaining games with different rules, the results may differ significantly from those here. However, again, when attempting to model real collective negotiations, it is important to explicitly consider the collective nature of the situation in order to determine what the appropriate two-player representation of the game should be.

Also, I have assumed that preferences of group members differ along only one dimension. In reality, it is often the case that groups are negotiating along many dimensions simultaneously, and members of the groups have varying preferences over many of these dimensions. To study this, conceptually, we could discuss a model of bargaining in which players are located at various points in a 2-dimensional space and are negotiating to locate something in this space. It is unclear how this complexity would affect the results offered here.

Finally, the results I present here are static in the sense that players are choosing voting rules and/or negotiators for a single negotiation. However, in practice, for a variety of reasons, groups choose voting rules and/or negotiators only periodically, with many negotiations taking place between these choices. Thus, additional considerations such as developing reputations for future bargaining situations may arise. Considering these issues in a dynamic framework would allow analysis of a richer array of such problems which undoubtedly arise in true collective bargaining situations. Such exercises are left to future research.

3.8 References

[1] Baron, David P. and John A. Ferejohn (1989), "Bargaining in Legislatures." *American Political Science Review* 83:1181-1206.

[2] Chatterjee, Kalyan and William Samuelson (1983), "Bargaining under Incomplete Information." *Operations Research* 31: 835-851.

[3] Nash, J.F. (1950), "The Bargaining Problem," *Econometrica* 18: 155-162.

[4] Nash, J.F. (1953), "Two-person Cooperative Games." *Econometrica* 21:128-140.

[5] Osborne, Martin J. and Ariel Rubinstein (1990), *Bargaining and Markets*, San Diego: Academic Press.

[6] Raiffa, Howard (1982), *The Art and Science of Negotiation*, Cambridge: The Belknap Press of Harvard University Press.

[7] Rubinstein, Ariel (1982), "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50: 97-109.

[8] Stähl, I. (1972), *Bargaining Theory*. Economics Research Institute, Stockholm School of Economics.

[9] Watson, Joel (1995), "Alternating-Offer Bargaining with Two-Sided Incomplete Information," University of California, San Diego Discussion Paper 94-13R.