The Pennsylvania State University

The Graduate School

Department of Economics

ESSAYS ON REPEATED GAMES

A Thesis in

Economics

by

Eliot Maenner

Submitted in Partial Fulfillment
of the Requirements
for the degree of

Doctor of Philosophy

December 2002

We approve the thesis of Eliot Maenner.

Date of Signature

Kalyan Chatterjee Distinguished Professor of Economics and Management Science Thesis Co-Adviser Co-Chair of Committee	
Vijay Krishna Professor of Economics Thesis Co-Adviser Co-Chair of Committee	
Tomas Sjöström Professor of Economics	
James Jordan Professor of Economics	
Anthony Kwasnica Assistant Professor of Management Science and Information Systems	
Robert Marshall Professor of Economics Department Head	

Abstract

The problem of equilibrium selection in repeated games is approached by incorporating explicit models of players' decision processes into the repeated game. The first essay, "Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma," incorporates boundedly rational players into a repeated game who adapt their strategies based on certain simple models they build of each other's strategies. In the learning process I analyze strategies are represented by finite-state automata and players have a preference for a simpler strategy to a more complex one, provided the two strategies yield the same utility payoff (that is, preferences are lexicographic). The process consists of an inference part, where a player constructs a minimally complex model of the other player based on the observed path of play, and an adaptation part, where a player chooses a best response to one of the inferences. I show, in the context of the infinitely repeated Prisoners' Dilemma, that the nature of the inference crucially affects the nature of the steady states of the dynamical system associated with the inference-adaptation process; an optimistic inference leads players to the unique subgame perfect equilibrium in stationary strategies while a cautious inference leads players to the subset of self-confirming equilibria with Nash outcome paths.

The second essay, "Negotiation in Repeated Games," incorporates an explicit noncooperative model of bargaining into a repeated game which is the process by which players switch between continuation payoffs. Players negotiate with each other through an alternating offers protocol while the repeated game is being played to attempt to determine future play. In each stage there are three substages in which the players play the stage game, a continuation payoff is potentially offered, and the offer is accepted or rejected. I show that when the discount factor is sufficiently near one that negotiation-compatible equilibria exist and all negotiation-compatible equilibria in the model are nearly efficient – the players will play a repeated game strategy that is efficient after the first stage.

Contents

List of Figures **vi**

1 Introduction **1**

1.1 Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma . . . 2

1.2 Negotiation in Repeated Games . . . 3

2 Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma **6**

2.1 Introduction . . . 6

2.1.1 The Dynamic Learning Process . . . 7

2.1.2 Simplicity . . . 8

2.1.3 Results . . . 10

2.1.4 Equilbrium Selection . . . 12

2.2 The Model . . . 14

2.2.1 Basic Definitions . . . 14

2.2.2 The Dynamic Learning Process . . . 16

2.2.3 Examples of the Dynamic Learning Process . . . 19

2.2.4 The Process as a Model of Bounded Rationality . . . 22

2.3 Steady States . . . 24

2.4 Convergence: Optimistic Inferences . . . 30

2.5 Convergence: Cautious Inferences . . . 42

2.6 Conclusion . . . 53

2.7 References . . . 53

3 Negotiation in Repeated Games **56**

3.1 Introduction 56

3.2 Model 61

3.2.1 Preliminaries 61

3.2.2 Bargaining in Repeated Games 62

3.3 Characterizations of the Solution 65

3.3.1 Negotiation-Compatible Equilibria 65

3.3.2 Efficient Renegotiation 68

3.4 Conclusion 82

3.5 References 82

List of Figures

2.1 Set of Payoffs in a Prisoners' Dilemma 11
2.2 The Prisoners' Dilemma ($g, l > 0$) 14
2.3 A State Table Representation of Tit-For-Tat 15

Chapter 1

Introduction

Equilibrium selection in repeated games can be achieved by incorporating explicit models of players' decision processes into the repeated game. A standard repeated game is a model of long-term competition in which players' interests are partially in conflict, yet selfish pursuit of one's own goals may result in undesirable outcomes without a workable plan for cooperation. A standard repeated game presupposes a particular model of the players' decision processes in which players optimize and information processing is not explicitly modeled beyond the fact that past actions in the long-term competition are observed. An explicit model of a decision process of an individual or a group not only includes what information individuals receive, it also includes a description of how individuals process information – either by themselves or as a group. The behavior of the players in a repeated game may differ considerably when an alternative player model is substituted in the environment for the rudimentary player model tacitly included in every standard repeated game.

The behavioral predictions of standard repeated games are best summarized by the folk theorem: under minor technical assumptions, any feasible and individually rational payoff can be a perfect equilibrium payoff when players are sufficiently patient .[1] The reason for the multiplicity is straightforward – time permits effective punishments to be constructed. The indeterminism of this result stands in contrast to the relatively sharp predictions obtained when the stage game is repeated only once. Yet, static models are unsatisfactory in a dynamic world.

[1] Rubinstein (1979), Fudenberg and Maskin (1986).

In this thesis the problem of equilibrium selection in repeated games is approached by retaining the basic model of long-term competition and varying the player model of the individuals who are in this situation. I model the behavior of individuals whose decisions are restrained by complexity considerations, both in the strategies that they choose and in how they form inferences about each other's strategies. A manifestation of this behavior is that they react in cautious or optimistic ways to the same information and they adapt their strategies during the course of play. I also model individuals who attempt to compromise on future play by negotiating each other. It is by having players in a formal noncooperative game simulate their more complex counterparts in the world that I obtain sharper predictions in repeated games.

1.1 Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma

The model in Chapter 1, "Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma," incorporates boundedly rational players into a repeated game who adapt their strategies based on certain simple models they build of each other's strategies. In the learning process I analyze strategies are represented by finite-state automata and players have a preference for a simpler strategy to a more complex one, provided the two strategies yield the same utility payoff (that is, preferences are lexicographic). The process consists of an inference part, where a player constructs a minimally complex model of the other player based on the observed path of play, and an adaptation part, where a player chooses a best response to one of the inferences. I show, in the context of the infinitely repeated Prisoners' Dilemma, that the nature of the inference crucially affects the nature of the steady states of the dynamical system associated with the inference-adaptation process; an optimistic inference leads players to the unique subgame perfect equilibrium in stationary strategies while a cautious inference leads players to the subset of self-confirming equilibria with Nash outcome paths. I also demonstrate convergence to these steady states when the automata are restricted to be at most two states.

In the theory of repeated games the focus of research has been to show how dependence on history can lead to a large class of outcomes being sustainable as equilibria; for example, cooperation in the infinitely repeated Prisoners' Dilemma with players using strategies that pun-

ish each other for deviating from cooperation. In contrast, stationary (history-independent) strategies are used in many dynamic models in economics primarily for reasons of analytical tractability. When researchers impose this restriction they often justify it by an ad hoc equilibrium selection argument asserting that players have a predilection for simple strategies which have a minimal cost to implement. The learning model I analyze provides an explanation of simple behavior based on first principles. In particular, it addresses the question of when players will learn to coordinate on the unique stationary equilibrium of the infinitely repeated prisoners' dilemma.

The question of when restrictions to stationary strategies are justified cannot be divorced from the question of what kinds of behavior lead players to coordinate on stationary strategies. An approach employed by researchers in a variety of dynamic games to this equilibrium selection question models how the costs of complexity can affect players choices. This is a natural approach for two reasons. The first reason is that it is an attempt to formalize the ad hoc equilibrium selection arguments. The second reason is that in many dynamic games, including repeated games, it is often the case that there is a multiplicity of perfect equilibrium payoffs precisely because the players can use strategies that are history-dependent and complex. A common way to implement this approach is to incorporate a measure of complexity into the preferences of the players and then perform equilibrium analysis with the new preferences.[2]

The complexity of strategies also matters to the players in the model I analyze, but in addition I take a learning approach and model the complexity of the players' inference problem in an adaptive framework. In addition to choosing strategies, the players, in order to adapt their strategies, continually solve an inference problem in the course of play. It is this learning process which leads players to coordinate on particular strategies – a selection from the full set of perfect equilibria – even though they may not have began the game with these strategies.

1.2 Negotiation in Repeated Games

The essay in Chapter 2, "Negotiation in Repeated Games," incorporates an explicit noncooperative model of bargaining into a repeated game which is the process by which players switch

[2] For example, see Chatterjee and Sabourian (2000).

between continuation payoffs. Players, locked into long-term competition, negotiate with each other while the repeated game is being played to attempt to determine future play. The bargaining process is modeled by an alternating offers protocol.[3] In each stage of the game there are three substages in which the players play the stage game, a continuation payoff is potentially offered, and the offer is accepted or rejected.

There are four obstacles to using an alternating offer in a repeated game context: (1) the set of continuation payoffs is not a fixed pie, nor do negotiations stop after one acceptance, (2) continuation payoffs are not legally enforced, (3) negotiations are potentially just cheap talk, and (4) equilibrium payoff sets for a particular discount factor may lack certain regularity properties. All of these issues are addressed in the negotiation model. To address the third issue, which, like the second issue, is a consequence of eliminating legal enforcement in Rubinstein's (1982) model, I define a class of perfect equilibria named *negotiation-compatible equilibria*. The restriction, as the name suggests, is intended to analyze negotiations that involve serious talk in which there is meaningful relationship between the outcome of negotiations and the strategies followed afterwards. The essence of the main assumption in negotiation-compatible equilibrium is that if the parties design a self-enforcing agreement at the bargaining table and both agree to it, then after walking away from the bargaining table they expect the opponent to follow the agreement in the next subgame.

In the main theorem of the paper I show that when the discount factor is sufficiently near one that negotiation-compatible equilibria exist and all negotiation-compatible equilibria in the model are nearly efficient – the players will play a repeated game strategy that is efficient after the first stage. This means that the only set of payoffs which players will not negotiate away from is the Pareto frontier.

Farrell and Maskin (1989) use a postulate from bargaining theory – that the outcome of the bargaining process be Pareto efficient – to refine the set of subgame perfect equilibria. However, there has not been a consensus on what the proper notions of internal and external consistency are for renegotiation proof sets in infinitely repeated games. Yet, when the bargaining process of the players in an infinitely repeated game is explicitly modeled, the issue of internal and external consistency is replaced by different issues – issues related to solving the noncooperative

[3] Rubinstein (1982) is a thorough analysis of the baseline alternating offers model.

game instead of issues about what assumptions to impose on the outcomes of an undefined behavioral process. In this regard, the approach to equilibrium selection taken in the model I analyze differs from the approach taken in the renegotiation literature. The alternating offers protocol of bargaining is used to model *how* the players make decisions. This protocol is as simple as it is rich: it is a model of bilateral communication and a noncooperative mechanism for switching between equilibria, as well as its usual role as a model of an institution. When long-term competition is modeled by a repeated game and the bargaining process is modeled by alternating offers players can negotiate to efficiency. In this context, we further the investigation of renegotiation in repeated games by attempting to more fully integrate bargaining theory into repeated games.

Chapter 2

Learning to be Simple: Adaptation and Complexity in the Repeated Prisoners' Dilemma

2.1 Introduction

Many dynamic models in economics are analyzed by restricting attention to equilibria in which agents use *stationary* (history-independent) strategies. This is true of most analyses of dynamic macro models, search models, and random matching models. This is also true of many models of dynamic games including stochastic games and multi-person bargaining. Indeed such a restriction is the norm rather than the exception. The restriction to stationary equilibria is usually justified on both pragmatic and theoretical grounds. First, stationary equilibria are more tractable for the analyst. In addition, the set of stationary equilibria is typically smaller than the set of all equilibria and so yields a sharper prediction. For instance, in a repeated game the set of equilibria is very large but restricting attention to stationary equilibria yields a small set, sometimes even a singleton. A more substantive justification is based on the hypothesis that the agents themselves prefer history-independent strategies because they are simple computationally and put fewer demands on their limited memories. This paper is an exploration of this hypothesis.

The restriction to stationary equilibria has not been a feature of the analysis of repeated games. Here the focus has been on demonstrating various folk theorems precisely by conditioning in a complex way on past histories. For example, in the infinitely repeated Prisoners' Dilemma, probably the most studied game in this group, cooperation can be sustained by strategies that punish in a history-dependent way.

In this paper I ask whether players can *learn* to play simple, that is, history-independent, strategies. In my model the primary objective of every player is payoff maximization but a player prefers simple decision rules to more complicated ones. It is important to understand that simplicity is only a secondary goal—such players will discard a simple rule in favor of a more complicated rule if it will give them even the slightest gain in payoffs. Will such players learn to play stationary strategies? I explore this idea in the context of the infinitely repeated Prisoners' Dilemma and show that indeed players will play the unique subgame perfect equilibrium in stationary strategies (namely, both players defect in every period). I identify conditions—embodied in behavioral assumptions—under which players "learn to be simple" in this sense. The exact manner in which the learning takes place—via a dynamical process—is as follows.

2.1.1 The Dynamic Learning Process

The adaptive process is recurrent one and at each recurring stage of the process players are assumed to (i) *observe*; (ii) *infer*; and (iii) *choose.*

A choice consists of a strategy in the repeated Prisoners' Dilemma—not an action. Each player begins by choosing some strategy. The pair of strategies so chosen now determines a mode of play in the game. For instance, a particular pair of strategies may lead to the following outcome path

$$(C,D),(D,C),(C,D),(D,C),...$$

Players *observe* only the path of play and not the rule of behavior chosen by their opponents. On the basis of this observation, each player must try to *infer* what strategy the opponent has chosen. Continuing with the example above, suppose player 1 chooses the "Tit-For-Tat" strategy and tries to infer what strategy player 2 might be playing. Player 1 could think that

player 2 is following a strategy which calls on player 2 to do the opposite of what player 1 did in the previous period. Such a guess would be consistent with the observed path. On the other hand, he could infer that player 2 is following a strategy that calls on player 2 to defect in odd numbered periods and this would also be consistent with the observed path. Having inferred the strategy followed by player 2, in a manner yet to be specified, player 1 now *chooses* a new strategy that is optimal—it is a best response—against the inferred strategy of player 2. The new choices lead to a new outcome path which is observed and the whole process is repeated. Following Fudenberg and Maskin (1993), who also have a dynamic game in which the stage game is an infinitely repeated game, a single round of the adaptive process is referred to as an *epoch* instead of a *period* to emphasize that players are primarily choosing rules of behavior and not one-shot actions. This terminology should not mislead one into thinking of the epochs as generations or the model as an evolutionary model[1].

2.1.2 Simplicity

In the process I study in this paper simplicity considerations influence both the choices that players make and the inferences they draw. First, in choosing which strategy to adopt, a player selects one from among those that maximize his payoff given his inference. In other words, the adopted strategy is always one that is a best response, with respect to payoffs, to the inferred strategy of the other player. But if there is more than one best response and one is "simpler" than another, then the simpler one is favored. Notice that this means that there is no trade-off between simplicity and payoffs: a very complicated strategy that yields even the smallest gain in payoffs would be preferred to a simpler strategy.

Second, in inferring which strategy the other player may have adopted, players make use of Occam's Razor, that is, they opt for the simplest explanation that fits the observed facts. In other words, if there are two possible strategies which, if ascribed to the other player, could have led to the observed path and one is more complicated than the other then the complicated strategy is discarded as a possibility.

[1] Our model falls under the category of learning models. It does not involve any population dynamics and, players, instead of being subsumed under a population, make explicit choices. Rationality has been replaced by bounded rationality, but not natural selection.

What is meant by simplicity? Clearly there cannot be a single compelling way to measure the simplicity or complexity of a strategy. Following a suggestion of Aumann (1981), it is assumed that players only choose strategies that can be implemented by finite state *automata*. This allows simplicity/complexity of a strategy to be measured in a straightforward manner. An automaton with fewer states than another is considered to be simpler.

Formally, an automaton is a 4-tuple $M_i = \{A, Q_i, f_i, \tau_i\}$, where A is a finite action set, Q_i is a finite set of states, $f_i : Q_i \rightarrow A$ is an output function, and $\tau_i : Q_i \times A \rightarrow Q_i$ is a transition function that maps states and an action of the other player into a state[2]. One of the states, denoted q_i^1, is a given initial state. Finite state automata rule out strategies in which players condition on their own choices[3]. They only specify an action after histories that are reached with positive probability against any automaton of an opponent. Hence, an automaton representation of a rule of behavior formally corresponds to a plan of action, instead of a strategy (Rubinstein (1998)). This is an exogenous restriction on the abilities of the players to process information and, by the omission of some instructions that must be included in a strategy, implicitly imposes a preference for simplicity on the players.

The restriction to automata rules out arbitrarily complicated strategies such as playing Cooperate in every prime numbered period and Defect otherwise. It also rules out strategies that use a statistical decision rule based on all past information such as playing Defect if and only if the other player has defected more times than I have. Nevertheless, the set of automata is rich enough so that the conclusion of the celebrated "Folk Theorem" obtains even if players are restricted to choose strategies that can be implemented by automata (Osborne and Rubinstein (1994)).

Since players' strategies are restricted to automata it is natural to suppose that when making inferences, that the models players construct of their opponents are also automata. Moreover, out of the infinitely many possible automata that could explain the observed play, a player believes that the opponent's automaton is one with a minimal number of states. That is, players make the simplest possible inferences. Yet there may be more than one simplest automaton

[2] If players did not have the same action sets we would have to distuish between the set of inputs and the set of outputs.

[3] However, see Kalai and Stanford (1988) for an alternative definition of an automaton representation of a rule of behavior.

that explains the facts and how players arrive at a particular inference from this set matters. I introduce and study two alternative—and polar—cases.

Under *optimistic* inferences player 1 infers that player 2's automaton is such that if player 1 were to play a best response against it, player 1's payoff would be at least as high as that from a best response to any other simplest automaton that player 2 could have played. The optimistic rule corresponds to a *maximax* rule of inference from the set of simplest inferences; the inferred automaton is one that would yield player 1 the highest payoff if he were to optimize against it.

Alternatively, under *cautious* inferences player 1 infers that player 2's automaton is such that if player 1 were to play a best response against it, player 1's payoff would be no greater than that from a best response to any other simplest automaton that player 2 could have played. The cautious rule corresponds to a *minimax* rule of inference from the set of simplest inferences; the inferred automaton is one that would yield player 1 the lowest payoff if he were to optimize against it.

The two inference rules lead to two different dynamics and I study properties of both.

2.1.3 Results

I study the *steady states* and the *convergence* properties of the dynamic learning processes outlined above[4]. When players use optimistic inferences I find (Section 3, Proposition A) that the unique steady state of the dynamic learning process consists of both players choosing the one state automaton that always plays Defect (which is also the unique stationary equilibrium of the repeated Prisoners' Dilemma and a refinement of Nash equilibrium). This unique prediction stands in contrast to the multiplicity of equilibria, the entire set of individually rational and feasible payoffs to be precise, obtained by the folk theorem in the infinitely repeated Prisoners' Dilemma. On the other hand, when players make cautious inferences the set of payoffs attained by the steady states corresponds to the set of equilibrium payoffs in the infinitely repeated game with automata when simplicity is a secondary goal (Propositions B and C). Moreover, a steady state need not be a Nash equilibrium, although it must be a self-confirming equilibrium. Thus,

[4] We will use the term *steady states* when referring to the *stationary points* of the dynamical system in order to avoid confusion with the *stationary equilibria* in a game.

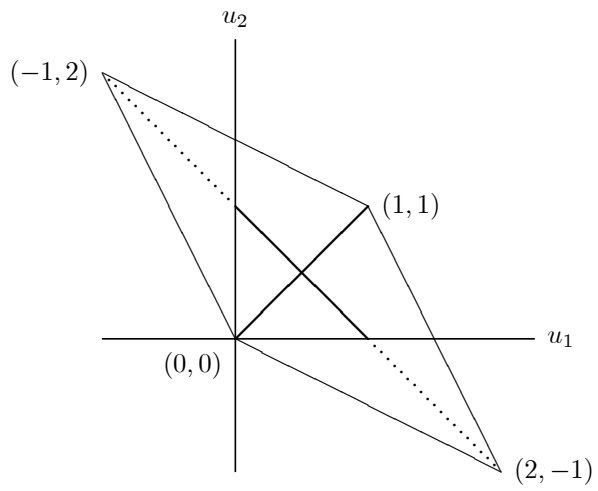

Figure 2.1: Set of Payoffs in a Prisoners' Dilemma

although optimistic inferences obtain a unique prediction and are consistent with stationary behavior the same cannot be said about cautious inferences.

Figure 1 depicts the set of feasible payoffs in a Prisoners' Dilemma—this is the diamond shaped figure. The folk theorem implies that any feasible and individually rational payoff—those feasible payoffs that Pareto dominate $(0,0)$—can arise as an equilibrium of the infinitely repeated game. Proposition A shows that with optimistic inferences the steady state payoff is a singleton, that is, $(0,0)$. Propositions B and C show that with cautious inferences the steady state payoffs are those points on the two diagonals that also Pareto dominate $(0,0)$. Note that this is the same as the set of equilibrium payoffs in the infinitely repeated game with automata when simplicity is a secondary goal (see Abreu and Rubinstein (1988)).

I show that there cannot be a cycle in the class of one and two state automata under either inference rule. That is, the selected strategies converge to the set of steady states from any initial configuration (Propositions D and E). I conjecture that there is global convergence in the set of finite state automata. Traditionally, in dynamical systems, convergence results are more difficult to obtain than steady state results, and this model is not an exception to this rule. Yet convergence is important not only on formal grounds but also for the interpretation

of the model. It enables us to claim that, under optimistic inferences, the players, beginning the repeated game with generic plans of action, *learned* to play the stationary equilibrium of the repeated game.

2.1.4 Equilbrium Selection

Researchers initially hoped that complexity considerations would address the problem of multiple equilibria. In Abreu and Rubinstein (1988) players also choose rules of behavior represented by finite state automata and have a preference for simplicity. Although they obtain that the set of equilibrium payoffs in this game is a strict subset relative to what is obtained by the folk theorem, there still is not an obvious prediction. I obtain a unique prediction by incorporating additional complexity considerations into the behavior of the players in Abreu and Rubinstein. Yet, formally, due to the introduction of an adaptive process, the two models differ considerably.

The problem of selecting stationary strategies has followed an approach in which complexity considerations of the players, or, more generally, elements of bounded rationality, have been incorporated into particular dynamic games. Although a unified approach to this problem would be desirable, it is unclear at this stage whether any single conceptualization of bounded rationality is appropriate for all contexts. In Chatterjee and Sabourian (2000) players in an n-person unanimity bargaining game have a preference for simplicity and, in equilibrium, only choose stationary strategies. In their model, a preference for simplicity is measured by the internal complexity of strategies and, informally, implies that players prefer strategies which require less separation of information, but otherwise yield the same rewards. Using the same formulation of complexity considerations, Sabourian (2000) is able to reduce the set of equilbria in a dynamic matching and bargaining game to only stationary strategies, all of which induce the competitive price.

Maskin and Tirole (1997, 2001) address the hypothesis that play of simple strategies emerges when players have complexity considerations with a model in which this behavior emerges through a learning process. Using a general framework with a large population of players, they show that players would learn to play simple strategies provided that, at the start of the game, a critical proportion of the players is already using simple strategies. For, once players learn this

state of the world, any incentive to choose more complex strategies vanishes[5]. The structure of my model, although it also incorporates complexity considerations into an environment in which learning occurs, differs considerably from their model. I sacrifice some generality in order to obtain more precise characterizations. The population in my model is restricted to two players (which prohibits the result from following from any "mass action" force in the population), I incorporate realistic assumptions into the model about the players' abilities to observe each other's behavior, and I obtain convergence results for strategies.

Binmore and Samuelson (1992) also study equilibrium selection in the repeated Prisoners' Dilemma. They define a version of evolutionary stability for the game with automata when simplicity is a secondary goal and obtain, under limit of the means payoffs, that the symmetric efficient payoff pair is the unique equilibrium payoff. They motivate their model with the idea of the evolution of rules of behavior in the infinitely repeated Prisoners' Dilemma and conceptualize an approach to equilibrium selection that emphasizes the behavioral assumptions of the players over trembles in the environment.

Many equilibrium selection methods modify an environment played by rational players. I do the opposite. The dimensions of the players are modified and the environment is kept the same. This approach establishes a direct correspondence between certain behavioral assumptions and equilibrium predictions. I imagine that players are choosing rules of behavior at the start of a repeated Prisoners' Dilemma situation and modifying them as play progresses. The decision epochs are faithful to this idea, despite their further abstraction of the time dimension beyond that already present in an infinitely repeated game. The gain from this is dynamics in a strategy space, not just an action space, which I regard as a significant advantage of the approach in this paper. Although the players, like their rational kin, are goal-seekers, they must do so with rather ordinary abilities to anticipate contingencies and process information. With limited abilities the players naturally make mistakes in the course of play, not by accident, but by errors in information processing and inference. The degree of their ability is in-between the naive and the sophisticated, arguably more towards the latter, which distinguishes this learning model from many others. The exigencies of having players adapt under scarce information and with ordinary abilities eliminates Nash equilibria present in a more static counterpart. Nash

[5] This learning model does not appear in the published version of their paper.

equilibria are often said to be the only sensible solutions (although the development of the self-confirming equilibrium concept has modified this view), yet equilibrium analysis is silent on how equilibrium play arises and is a static approach in this regard, even in traditional dynamic games. This issue is presented starkly in learning models in which players start with arbitrary plans and only observe realized play.

2.2 The Model

2.2.1 Basic Definitions

The game played in each period is the Prisoners' Dilemma (Figure 2). The action set is denoted $A = \{C, D\}$ and the payoffs are given by $u_i : A^2 \rightarrow \mathbb{R}$, $i = 1, 2$.

	C	D
C	1, 1	$-l, 1+g$
D	$1+g, -l$	0, 0

Figure 2.2: The Prisoners' Dilemma ($g, l > 0$)

An automaton is a 4-tuple $M_i = \{A, Q_i, f_i, \tau_i\}$, where A is a finite action set, Q_i is a finite set of states, $f_i : Q_i \rightarrow A$ is an output function, and $\tau_i : Q_i \times A \rightarrow Q_i$ is a transition function that maps states and an action of the other player into a state. One of the states, denoted q_i^1, is a given initial state. For any automaton M_i the number of states in Q_i is denoted $|M_i|$. Let $q^t = (q_1^t, q_2^t)$ be the pair of states and $f(q^t) = (f_1(q_1^t), f_2(q_2^t))$ be the chosen actions at time t. The superscript on a state is reserved for a period in time, the subscripts refer to the identity of the player (this subscript is often omitted when speaking of a particular automaton) or the identity of a state in an automaton. Every pair of automata (M_1, M_2) generates a sequence of states: $\{q^1, q^2, q^3, ...\}$. The finiteness of both the action set and set of states implies that this sequence must eventually cycle after an introductory phase (possible empty):

$$\{q^1, ..., q^{t_1-1}, \overbrace{q^{t_1}, ..., q^{t_2}}^{Cycle}, \overbrace{q^{t_1}, ..., q^{t_2}}^{Cycle}, ...\}$$

The outcome path that corresponds to the sequence of states, denoted $\pi(M_1, M_2)$, equals $\{f(q^1), f(q^2), f(q^3), ...\}$.

The *state table* representation of a finite state automaton displays the 4-tuple and the initial state in a table. These tables are useful for computations. The initial state is the first state listed in the current state column.

Current state	*1st Input = C*	*2nd Input = D*	*Output*
q_{i1}	$\tau_i(q_{i1}, C) = q_{i1}$	$\tau_i(q_{i1}, D) = q_{i2}$	$f_i(q_{i1}) = C$
q_{i2}	$\tau_i(q_{i2}, C) = q_{i1}$	$\tau_i(q_{i2}, D) = q_{i2}$	$f_i(q_{i2}) = D$

Figure 2.3: A State Table Representation of Tit-For-Tat.

The two stationary strategies in the infinitely repeated Prisoners' Dilemma are represented by the two one state automata which always play the same action (I will refer to these automata by the names COOPERATE and DEFECT.). Any automaton with at least two reachable states that has at least one cooperative and competitive state is not stationary; whether Cooperate or Defect is chosen depends on the history of play. This permits many of the main ideas and propositions to be illustrated in set of 1 and 2 state automata. Moreover, most strategies traditionally of interest in the repeated Prisoners' Dilemma can be represented by two state automata (e.g., Grim, Tit-For-Tat).

Besides the restriction to choosing a plan in the set of automata, players also have an explicit preference for simplicity in the set of automata. This is motivated by the idea that plans with additional complexity impose additional costs on the players. Following Abreu and Rubinstein (1988), complexity is measured by the number of states in the automaton [6]. Provided two automata attain the same payoff, the players prefer the least complex automaton, where complexity is increasing in the number of states. This preference is represented by lexicographic preferences, denoted $\succ_i, i = 1, 2$, in which discounted payoffs are of primary consideration and complexity costs are secondary.

[6] This measure of complexity is coarser than the measure in Chatterjee and Sabourian. Also see Kalai and Stanford (1988).

Formally, the game played in each epoch is the infinitely repeated Prisoners' Dilemma. The choice set of both players is the set of finite state automata, denoted $\mathcal{M}$. A choice determines a plan of action that implements an action in each period. The payoff in an epoch, denoted $U_i(\pi(M_1, M_2))$, is the discounted average $(0 < \delta < 1)$ of the sequence of payoffs derived from the outcome path in the epoch: $(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_i(f(q^t))$.

The strict preference relation $\succ_i$ on the set of pairs of finite state automata is defined by $(M_1, M_2) \succ_i (M_1', M_2')$ if (i) $U_i(\pi(M_1, M_2)) > U_i(\pi(M_1', M_2'))$, or (ii) $U_i(\pi(M_1, M_2)) = U_i(\pi(M_1', M_2'))$ and $|M_i| < |M_i'|$.

The automaton M_1^* is said to be a *best response* to an automaton M_2 if $(M_1^*, M_2) \succeq_1 (M, M_2)$ for all automata M. A pair of automata is a Nash equilibrium when each automaton is a best response to the other. Given any automaton, M_j, of player j, the set of best responses to M_j, denoted $B_i(M_j)$, is well-defined and consists of automata that have the same number of states and attain the same payoff[7].

2.2.2 The Dynamic Learning Process

I imagine the following sequence of events in the dynamic learning process. In the first epoch, epoch 1, players start with an arbitrary pair of finite state automata. Player 1 has an opportunity to switch automata only in the even numbered epochs (2,4,...) and player 2 can switch automata only in the odd numbered epochs (3,5,...). Starting in epoch 2 and for every epoch thereafter players observe the outcome path of the previous epoch and use it to construct inferences of each other (i.e. the automata which can generate the sequence of actions that appeared on the outcome path), excluding all inferences except those that have a common minimal number of states. These minimal state inferences are their models of the other player. The player who can switch automata at epoch t (generically referred to as player i hereafter) infers exactly one of these models is the true plan of action of player j. Player i chooses an automaton for epoch t that is a best response to the inferred model. Player j continues to use the same automaton as in epoch $t-1$. These choices yield an outcome path for epoch t. Then the process repeats – observe, infer, choose.

[7] An comprehensive introduction to repeated games with automata appears in Osborne and Rubinstein (1994).

In the observation step of the adaptation process, players construct models of each other. Players do not observe each other's plans of action. The only information that the players can use in their decision problem is the information contained in the outcome path of the previous epoch. Information from epochs prior to the previous epoch is not factored into the agents' decision problems. One interpretation of this limited hindsight is that the players consider regimes in the distant past to be irrelevant; they only react to the current regime and do not factor every regime since the beginning of time into today's decision.

The set of automata which could be inferences about a player is infinite (for example, any automaton with redundant states is also an inference). However, if each player believes that the other player is minimizing with respect to complexity costs, i.e. if players take each other's preferences into consideration, then they will infer only *minimal state automata,* and the set of inferences of each player is finite[8].

Informally, an automaton is a minimal state automaton if its behavior can't be duplicated by an automaton that has fewer states. Following Birkhoff and Bartee (1970), an automaton M is said to be a *minimal state automaton* if there does not exist another automaton M' with the same action set such that *(i)* $|M'| < |M|$, and *(ii)* M and M' yield identical output sequences for each input sequence.

For any outcome path $\pi(M_1, M_2)$ generated by a pair of finite state automata, define $\mathcal{I}_1$ to be the set of all minimal state automata with C_1 states such that (i) if $M \in \mathcal{I}_1$, then $\pi(M, M_2) = \pi(M_1, M_2)$, and (ii) there does not exist an automaton M' with fewer than C_1 states such that $\pi(M', M_2) = \pi(M_1, M_2)$. The set $\mathcal{I}_2$ is defined similarly. The sets $\mathcal{I}_1$ and $\mathcal{I}_2$ are called the sets of *minimal state inferences* for players 1 and 2, respectively, or sometimes just *models*. The dependence of these sets on the outcome path does not appear in the notation, but will be clear in context. For any pair of finite state automata (M_1, M_2) the corresponding sets $\mathcal{I}_1$ and $\mathcal{I}_2$ are not empty, unique, finite, and $\pi(M_1', M_2') = \pi(M_1, M_2)$ for all $(M_1', M_2') \in \mathcal{I}_1 \times \mathcal{I}_2$. For player i this relation is symbolized by the multi-valued mapping: $\psi_i : \mathcal{M}^2 \rightrightarrows \mathcal{M}$ where $\psi_i(M_1, M_2) = \mathcal{I}_j$.

Example 1 *Let* $M_1 = M_2 =$ *TIT-FOR-TAT. Then* $\pi(M_1, M_2) = \{(C, C), ...\}$ *and,* $\mathcal{I}_1 =$

[8] Spiegler (2001) uses similar criteria to rank beliefs in a repeated game.

$\mathcal{I}_2 = \{COOPERATE\}$. *Thus, even though TIT-FOR-TAT is a minimal state automaton, it is not a minimal state inference.*

In the inference step of the adaptation process players infer that exactly one of the models they constructed of each other represents the true plan of the other player. After choosing this model, in the optimization step of the adaptation process, the player who may adapt chooses a best response to this model. Formally, these two steps are combined into a single step because players are choosing a model of the other player by comparing the rewards available from each model.

The choice of player i in epoch t is denoted M_i^t for t=1,2,.... If player i infers model M_j about player j in epoch $t-1$ then the set of optimal choices of player i is the best response set $B_i(M_j)$. If $M_i^{t-1} \in B_i(M_j)$, however, then it is required that M_i^{t-1} be chosen in order to minimize *switching costs*[9]. The optimistic player infers a model whose best response is most preferred relative to the best responses to other models and the cautious player infers a model whose best response is least preferred relative to the best responses to other models. Each inference rule defines a different dynamic learning process. I study the two systems that arise when both players use the same inference rule. Formally, if M_i^t is chosen at epoch t under the *optimistic inference rule* it must satisfy condition (O), and under the *cautious inference rule* it must satisfy condition (C). Condition (S), the minimization of switching costs, must be satisfied under both inference rules.

Condition (O) *Optimistic inferences:* $M_i^t \in B_i(M_j)$ *for some* $M_j \in I_j^{t-1}$, *and* $(M_i^t, M_j) \succcurlyeq_i (B_i(M_j'), M_j')$ *for all* $M_j' \in I_j^{t-1}$. *(There is an abuse of notation when* $B_i(M_j')$ *is multi-valued, in which case the condition holds for each element of* $B_i(M_j')$*).*

Condition (C) *Cautious inferences:* $M_i^t \in B_i(M_j)$ *for some* $M_j \in I_j^{t-1}$, *and* $(B_i(M_j'), M_j') \succcurlyeq_i (M_i^t, M_j)$ *for all* $M_j' \in I_j^{t-1}$.

Condition (S) *Players minimize switching costs under both inference rules: If* $M_i^{t-1} \in B_i(M_j)$, *then* $M_i^t = M_i^{t-1}$.

[9]Instead of having switching costs be part of the adaptation procedure we could directly incorporate them into preferences.

The dynamics in the set of finite state automata can be described by the diagram:

$$(M_1^{t-1}, M_2^{t-1}) \overset{\psi_i}{\to} \mathcal{I}_j^{t-1} \overset{B_i}{\to} (M_1^t, M_2^t).$$

Let (M_1^1, M_2^1) be an arbitrary pair of finite state automata. Given this starting point, the dynamic learning process generates sequences of the form $\{(M_1^t, M_2^t)\}_{t=1}^{\infty}$, which are the objects of study in the remainder of the paper. The set of possible automata sequences that the dynamic learning process generates for the initial condition (M_1^1, M_2^1) is denoted $\Gamma(M_1^1, M_2^1)$.

The choice of player i is a single element from a best response set, but may not be uniquely determined. All possible choices must yield the same payoff and have same number of states, yet they may differ in their transition functions. Further reduction of the choices is possible, but the multiplicity of choices cannot be eliminated entirely and indeed is intrinsic to best responses with automata[10]. Nevertheless, I demonstrate that analysis of steady states and convergence is possible.

2.2.3 Examples of the Dynamic Learning Process

The following two examples illustrate the dynamic learning process for particular initial conditions. The first example illustrates basic concepts and the second example illustrates the difficulty of predicting the path that the process will take.

Example 2 *This example illustrates the inference and adaptation procedure for an automaton that has a one period "show-of-strength" and then is willing to cooperate. This automaton is represented by the following table:*

"*SOS*" =

	C	D	
q_1	q_1	q_2	D
q_2	q_2	q_1	C

Consider the process that begins when both players adopt SOS. The outcome path that results

[10] Having the players choose a response that is optimistic or cautious in the same manner as their inference would reduce, but not eliminate, the multiplicity.

is:

$$\pi(SOS, SOS) = \left\{ \binom{D}{D}, \binom{C}{C}, \binom{C}{C}, ... \right\}$$

The set of minimal state inferences each player forms about epoch 1 consists of 4 automata:

$$\mathcal{I} = \left\{ \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_1 & q_2 & D \\ q_2 & q_2 & q_1 & C \end{array} , \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_1 & q_2 & D \\ q_2 & q_2 & q_2 & C \end{array} , \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_2 & D \\ q_2 & q_2 & q_1 & C \end{array} , \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_2 & D \\ q_2 & q_2 & q_2 & C \end{array} \right\}$$

For conciseness and computational purposes it is convenient to represent the set $\mathcal{I}$ with the following *incompletely specified automaton (*Birkhoff and Bartee*, 1970):*

$$\mathcal{I} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_2 & -_b & C \end{array}$$, where $-_a$ and $-_b$ could be any state.

At the epoch 2 decision problem an optimistic player 1 will infer $-_b = q_2$, which permits the highest possible payoff $\delta(1+g)/(1-\delta)$ to be attained with the one state automaton DEFECT. Starting at the initial condition (SOS, SOS) the dynamic learning process, under optimistic inferences, generates a single sequence which quickly converges to the steady state:

$$\left\{ \binom{SOS}{SOS}, \binom{DEFECT}{SOS}, \binom{DEFECT}{DEFECT}, \binom{DEFECT}{DEFECT}, ... \right\}$$

Under cautious inferences a minimax selection is made as follows:

(i) The best response against any inference with $-_b = q_2$ is DEFECT and has the payoff $\delta(1+g)/(1-\delta)$.

(ii) The best response against any inference with $-_b = q_1$ is:

(a) DEFECT with the payoff $\delta(1+g)/(1-\delta)$, if $g \geqq \delta$, or,

(b) The solution given by the mapping $\{q_1 \rightarrow D, q_2 \rightarrow C\}$, which yields the payoff $\delta/(1-\delta)$, if $g < \delta$. The best response set associated to this solution is simply any automaton in the set I.

If $g < \delta$, the cautious player must choose a two state automaton in I, and, to minimize switching costs, must choose SOS. Under this parameter restriction, starting at the initial

condition (SOS, SOS) *, the dynamic learning process with cautious inferences generates a single constant sequence. Indeed,* (SOS, SOS) *is a Nash equilibrium of the game with preferences* $\succ_i$*, and, as will be proven in Section 3 all such equilibria are steady states of the dynamic learning process under cautious inferences.*

Example 3 *Suppose the path observed after the first epoch is*

$$\left\{\binom{C}{D}, \binom{D}{C}\vdots\binom{C}{D}, \binom{D}{C}\vdots\ldots\right\}.$$

Player 1, the player who updates at epoch 2, is the top player. Player 1 constructs 4 minimal state inferences about player 2 before the epoch 2 decision problem:

$$\mathcal{I} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & -_b & q_1 & C \end{array}$$

The optimistic player will infer that $-_a = q_2$ *for all parameters g,l >0, and, if the gain to cheating isn't too large, will also infer* $-_b = q_2$*. The corresponding optimal solution is the mapping* $\{q_1 \to D, q_2 \to C\}$ *which is implemented by the best response set:*

$$B_1(M_2) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_c & q_2 & D \\ q_2 & q_2 & -_d & C \end{array}$$

Since player 1 continues with a two state rule it is not obvious, without further calculation, what the dynamics of the system will be after the first epoch. The epoch 2 outcome path depends not only on the choice of player 1, but also on the actual automaton of player 2. However, players frequently make incorrect inferences, choosing a response with respect to an automaton that is not the true automaton. This creates a divergence between what the updating player expects to occur and what actually occurs, complicating the dynamics considerably. It also tends to slow the speed of convergence, especially when the gains to cheating are small. Despite the relative simplicity of the outcome path in this example it is not obvious what paths the dynamic process could take or how these paths depend on the gain and loss parameters of the

stage game.

Examples 2 and 3 suggest that two common types of behavior exhibited by optimistic agents are expectation of unpunished cheating and expectation of mutual cooperation. Thus, an optimistic player can have an incentive to play both competitively and cooperatively. Example 3, in which a cooperative form of optimism occurs, suggests that the simple intuition that optimistic players will always find it optimal to switch to the dominant action is incorrect.

2.2.4 The Process as a Model of Bounded Rationality

Some general guidelines for incorporating complexity considerations of the players into dynamic games have been proposed in the game theory literature. Rubinstein (1991) suggests that a good model in game theory should be realistic in the sense that it captures the state of the world as it is perceived by the players. Fudenberg and Levine (1998) suggest that an adaptive model in which players are too naive may be inadequate for the study of social processes, even though the model may provide good predictions of animal behavior. The model in this paper attempts to address these considerations by departing from fully rational players in various aspects including: use of *plans of action* instead of *strategies*, a preference for simple plans of action, *rules of thumb* for making inferences about opponents, and *limited foresight*. Despite the resulting limitations on the computational ability of the players, their behavior is richer than the behavior of players in many learning models who use a simple statistical rule to choose a one-shot action.

In my model *limited foresight* and the ability of the players to change their plans during the course of play are important ingredients in the foundation of a dynamic learning process. Rational players have perfect foresight and can anticipate contingencies into the infinite future. Yet, anticipation is a complex and costly undertaking. For example, in the game of chess, rational players choose complete plans of action before the start of the game, and abstract from the practical problem of unraveling the complexity of the vast number of configurations and contingencies. After the game begins, the plans are merely implemented as sets of instructions. Aumann (1986) reminded game theorists that the rationality of players in a dynamic game implies that they necessarily view the game as a one-shot game in which all decisions are actually made before the start of play. On the other hand, limited foresight of the players, in

the sense of a bound on the number and complexity of interactions that they can think through, would prohibit them from completely thinking through a dynamic game before it begins. The extent to which the players can anticipate and predict ensuing developments is restricted.

Once the foresight of the players is restricted, a natural connotation is that the players have incentives to modify their plans during the course of play. Although players may have chosen their plans optimally, these plans are based on information that potentially becomes impertinent. At some date players may find it advantageous to bring their plans into accordance with the current state of the world. A natural way to incorporate limited foresight into a dynamic model is to let the players switch plans at some frequency T, at which time they anticipate only T periods ahead. However, in this model, players make inferences based on the outcome path and it is convenient to assume $T = \infty$. This simplifying assumption implies that the outcome path generated by two automata will not be truncated, in which case it yields the maximum possible information about the players' rules of behavior[11]. This abstraction of the time dimension permits study of learning in a strategy space that is not just an action space.

While limited foresight motivates adaptation, *asynchronous adaptation* places a restriction on the timing of the adaptation. In particular, players may only switch automata in alternate epochs. This can be viewed as part of the state of the world, as it is perceived by the players. They perceive that adaptation is *gradual*: the players act as if their partner will *wait and see* what they do. Formally, asynchronous adaptation temporarily fixes a circumstance that impacts a player's payoffs. The player who adapts, acting as if the circumstance may exist indefinitely, factors it into his or her optimization problem. Having limited powers of anticipation, players believe that when they adopt a new plan of action that the other player will continue with the same plan for some duration of uncertain length (Discounting in an epoch represents uncertainty about the arrival of the next regime change, and it implies that players expect the next regime to arrive after finitely many periods.). In continuous time, the assumption of asynchronous adaptation can be interpreted as an assumption that people or organizations do not always change their plans in lockstep. Adaptation that is always

[11] If there were an upper bound on complexity, and it was assumed that the players knew this upper bound then the length of all cycles would be bounded and the players would know that no more information could be obtained by waiting. However, unless it is assumed that players know this bound then they cannot distinguish between a cycle of length one and a long cycle that has not yet ended.

simultaneous would be problematic because it could create a discontinuity between the past and the future, in the sense that information about the other player's plan of action may immediately become irrelevant. Yet if players believe the models they construct of each other have some relevance to the state of the world, then, for a meaningful model, indeed they do. Simultaneous adaptation, however, would sever this link between the past and future and may lead to naive cycles. In general, as long as adaptation is not simultaneous for a continuous infinite block of time I expect the convergence results to still hold.

2.3 Steady States

Definition 1 *(M_1^*, M_2^*) is said to be a steady state of the dynamic learning process if it is chosen in every epoch after the first epoch in which it appears: If $(M_1^T, M_2^T) = (M_1^*, M_2^*)$ then $\{(M_1^t, M_2^t)\}_{t=T}^{\infty} = \{(M_1^*, M_2^*), (M_1^*, M_2^*), ...ad\ infinitum\}$.*

If (M_1^*, M_2^*) is a steady state then the set $\Gamma(M_1^*, M_2^*)$ contains a single sequence which consists of the infinite repetition of (M_1^*, M_2^*), and the sets of minimal state inferences are the same in each epoch: $\{\mathcal{I}_1^t, \mathcal{I}_2^t\} = \{\mathcal{I}_1^*, \mathcal{I}_2^*\}$ for all t.

Proposition A *Assume optimistic inferences. In set of finite state automata the unique steady state of the dynamic learning process is: $M_1^* = M_2^* = DEFECT$.*

Proposition A states that the unique steady state of the dynamic learning process, under optimistic inferences, is the one state automaton that always plays Defect. Two lemmas will be proved before the proof of Proposition A is presented. Lemma 1 states that if the process is at a steady state then: (i) the automata of both players have the same number of states, and (ii) there is a model of the other player that is the true automaton (Example 1 illustrates why the second implication is not necessarily true if the pair of automata is not a steady state.). Lemma 2 states that every steady state under optimistic inferences is a Nash Equilibrium in the game with preferences $\succ_i$.

The proof of Proposition A is completed by demonstrating that any Nash equilibrium that involves cooperation cannot be a steady state. In standard equilibrium analysis the analyst

can verify whether two plans of action are a Nash equilibrium yet the players themselves would not know this unless it is assumed or else with substantial patience, experimentation, and, perhaps, communication on the part of the players. In this model players are basing their decisions on observed behavior revealed in the realized path of play and constructing models of unobserved behavior that is encoded in their respective plans. The players' responses in the epochs can be viewed as experiments. Experiments can be costly and even irreversible, and, in this model, result in an inevitable chain of events, dependent on their models of the world. The elimination of Nash equilibria involving cooperation by the adaptive process with optimistic inferences suggests that players are underestimating each other's willingness or ability to punish; what in fact is a sufficient deterrent is not observed and not anticipated. For example, some have argued that an automaton which starts with a Show-Of-Strength allows the other player to learn about its punishment capability, thereby providing a deterrent. My model demonstrates that optimistic players will not learn this way, and fail to be convinced that they will actually be punished. They believe they are in the best of possible worlds. Under cautious inferences, on the other hand, even though players construct the same set of models as the optimistic players, play of Nash equilibrium is preserved by the adaptation process; this necessarily implies that players are correctly anticipating some deterrent to rash behavior.

Although the stationary equilibrium of the infinitely repeated Prisoners' Dilemma does not yield an efficient payoff, the main exercise in this paper is to describe a learning approach that can select the stationary equilibria of a dynamic game. One interpretation of the prediction in Proposition A is that even when the players are patient, if they use optimistic models of the world they are led to play stationary strategies in the repeated Prisoners' Dilemma. The substantial anecdotal and experimental evidence that people tend to play simple history-dependent strategies, such as Tit-For-Tat or Grim (trigger) strategies, in a repeated Prisoners' Dilemma situation suggests that it is unlikely that people, on average, are as optimistic as the agents in this model. Hence, this model, insofar as people tend to be unlike its agents, supports the common prediction that people will not play stationary strategies in a repeated Prisoners' Dilemma situation.

A general discussion on convergence will be postponed until section 4, yet I remark at this point that the steady state results under cautious inferences do not reveal anything about

convergence and under optimistic inferences demonstrate that even with Nash equilibrium, a relatively small set of automata, that convergence to the steady state is not necessarily immediate (Also see Example 3).

The following basic results will be used in the proofs.

1. By the *Markov decision problem (MDP) of player i associated to automaton* M_j I mean the following: Choose a sequence of actions $\{a_i^{*t}\}_{t=1}^{\infty}$ that maximizes $(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_i(a_i^t, f_j(q_j^t))$ under the law of motion $q_j^{t+1} = \tau_j(q_j^t, a_i^t)$ that has initial state q_j^1. It is well known that a stationary solution $f_i : Q_j \to A$ exists.

2. An automaton M_i *implements* the optimal solution to the MDP if the outcome path $\pi(M_i, M_j)$ equals the sequence of outputs $\left\{(f_i(q_j^t), f_j(q_j^t))\right\}_{t=1}^{\infty}$ obtained in (1). This outcome path, by definition, attains the optimal value of player *i's* MDP. Of particular interest in the best response set, $B_i(M_j)$, which contains all of the automata that implement M_j which have a common minimal number of states. $B_i(M_j)$ can be constructed by the following 2-step algorithm:

(a) For each solution, $f_i : Q_j \to \{C, D\}$, of the MDP construct the outcome path that is generated when $f_i(q_j^t)$ is the input into M_j. That is, construct the sequence

$$\left\{(f_i(q_j^t), f_j(q_j^t))\right\}_{t=1}^{\infty}.$$

(b) The set of automata that are best responses for player i is obtained by constructing the set of minimal state inferences about player i for each outcome path and then keeping only those automata that have a common minimal number of states.

The algorithm is a convenient method to directly construct the best response set without explicitly referring to the transition function of player j. An immediate implication of these basic results is that a best response in the game with preferences $\succ_i$ does not require any more states than are in the argument. This fact together with the assumption that players build only minimal state models prohibits escalating complexity. Indeed, the maximum number of states in any automaton in a sequence of automata generated by the dynamic learning process is bounded by $\max\{|M_1^1|, |M_2^1|\}$.

Lemma 1 *Under optimistic or pessimistic inferences, if* (M_1^*, M_2^*) *is a steady state of the dynamic learning process then* $|M_1^*| = |M_2^*|$ and $M_j^* \in \mathcal{I}_j^*$, for $j = 1, 2$.

Proof. Since the pair (M_1^*, M_2^*) generates the outcome path $\pi(M_1^*, M_2^*)$ it follows that M_j^*, for $j = 1, 2$, has at least as many states as any minimal state inference about player j: $|M_j^*| \geqq C_j^*$. The claim will follow once it is shown that $|M_j^*| = C_j^*$, for $j = 1, 2$.

Notice that $|M_1^*| \leqq C_2^*$ and $|M_2^*| \leqq C_1^*$: Since player i chooses M_i^* it must be a best response to some $M_j \in \mathcal{I}_j^*$. A stationary solution to the Markov decision problem (MDP) based on the transition function of M_j is well-defined and has $|M_j| = C_j^*$ states. Since M_i^* is chosen by player i, this implies that M_i^* attains the same payoff that is attained by the solution to the MDP and has no more states: $|M_i^*| \leqq C_j^*$.

Since (M_1^*, M_2^*) is a steady state there is an epoch t such that $|M_j^t| = |M_j^{t+1}| = |M_j^{t+2}|$ for $j = 1, 2$. That $|M_j^t| \geqq C_j^t$ and $C_j^t \geqq |M_i^{t+1}|$ implies that $|M_j^{t+2}| \leqq |M_i^{t+1}| \leqq |M_j^t|$, or $|M_j^*| \leqq |M_i^*| \leqq |M_j^*|$. Thus, $|M_i^*| = |M_j^*|$ and $C_i^* \leqq |M_i^*| = |M_j^*| \leqq C_i^*$.■

Lemma 2: *Assume optimistic inferences. If* (M_1^*, M_2^*) *is a steady state in the dynamic learning process then it is a Nash equilibrium in the game with preferences* $\succ_i$.

Proof. Any $(M_1, M_2) \in \mathcal{I}_1^* \times \mathcal{I}_2^*$ yields the same outcome path $\pi(M_1^*, M_2^*)$ and the same payoff pair (v_1^*, v_2^*) for players 1 and 2, respectively. By Lemma 1, $(M_1^*, M_2^*) \in \mathcal{I}_1^* \times \mathcal{I}_2^*$. This result together with $M_i^* \in B_i(M_j)$ for some $M_j \in \mathcal{I}_j^*$ implies that $\pi(M_i^*, M_j) = \pi(M_i^*, M_j^*)$.

Thus, if M_i^* weren't a best response to M_j^* then there would be some best response $B_i(M_j^*)$ to M_j^* such that $(B_i(M_j^*), M_j^*) \succ_i (M_i^*, M_j)$, contradicting the choice of M_i^* under optimistic inferences. Therefore, $M_i^* \in B_i(M_j^*)$.■

Lemma 2, although short, will not necessarily follow if any of the three key ingredients is omitted: (i) there is a model of each player that is the true automaton, (ii) the minimal state inference sets of each player are the same across epochs in a steady state, and (iii) optimistic inferences. It is useful to recall that the players are unaware of the true automaton of the other player when they choose a model of the other player. The consequence of Lemma 2 states that even if their models of the other player are not the true automata, in which case they are choosing best responses to incorrect models, the conclusion is nevertheless that the pair

of automata in a steady state is a mutual best response. This stands in contrast to the case in which optimistic inferences are replaced by cautious inferences and indeed a steady state is not necessarily a Nash equilibrium, but is preserved as a steady state precisely because their models are incorrect.

Proof of Proposition A. Suppose (M_1^*, M_2^*) generates the sequence of states

$$\{q^1, ..., q^{t_1-1}, q^{t_1}, ..., q^{t_2}, q^{t_1}, ..., q^{t_2}, ...\},$$

with t_1 possibly one, and the outcome path $\pi(M_1^*, M_2^*) = \{f(q^1), f(q^2), f(q^3), ...\}$.

Theorem 1 in Abreu and Rubinstein (1988) will permit us to construct an inference which allows an improvement over M_i^*, thereby leading to a contradiction. Their theorem states that if (M_1^*, M_2^*) is a Nash Equilibrium of the game with preferences $\succ_i$, then the states of M_1^* (respectively M_2^*) which appear in the first t_2 periods are distinct. This inference, M_j', say, is constructed using the outcome path and the consequence of the theorem of Abreu and Rubinstein that the first t_2 states are distinct (and hence the number of states in a minimal state inference about player j is t_2).

$M_j' :=$

	C	D	
q_i^1	q_i^2	q_i^2	$f_j(q_j^1)$
q_i^2	q_i^3	q_i^3	$f_j(q_j^2)$
...	...	...	...
$q_i^{t_2}$	$q_i^{t_1}$	$q_i^{t_1}$	$f_j(q_j^{t_1})$

M_j' imitates the states and realized transitions of M_i^* and plays the actions of player j in the order they appear in the outcome path. By construction $|M_j'| = t_2$ and $\pi(M_i^*, M_j') = \pi(M_i^*, M_j^*)$. Thus, M_j' is a minimal state inference about player j: $M_j' \in \mathcal{I}_j^*$.

After Nash equilibrium that involve cooperation are eliminated as candidate steady states one can verify that the pair of automata that are both DEFECT indeed constitutes a steady state. Suppose M_1^* and M_2^* are not DEFECT and constitute a steady state. If $C^* = 1$ and the automata are not DEFECT then they must be COOPERATE, which is not a steady state. Therefore, assume $C^* > 1$ and that $f_i^*(q) = C$ for some $q \in Q_i^*$. (A minimal state automaton

with more than one state has at least one cooperative state.)

By inferring the minimal state inference M_j' player i can increase the payoff in state q by playing D without affecting the payoff in the other states (i.e. without punishment). Thus, the automaton M_i^* will not be chosen, a contradiction to the steady state assumption.■

Propositions B and C characterize the set of steady states under cautious inferences. Although every Nash equilibrium of the game with preferences $\succ_i$ is a steady state (Proposition B), there are steady states that are not Nash equilibria. However, every steady state that is not a Nash equilibrium must generate an outcome path that can be generated by a Nash equilibrium (Proposition C). Thus, the set of steady states under cautious inferences yields the same set of payoffs as the set of Nash equilibrium in the game with preferences $\succ_i$.

Proposition B *Assume cautious inferences. In the set of finite state automata, if* (M_1, M_2) *is a Nash equilibrium of the game with preferences* $\succ_i$ *then it is a steady state of the dynamic learning process.*

Proof. It must be shown that if $(M_1^t, M_2^t) = (M_1, M_2)$ then the choices (M_1^{t+1}, M_2^{t+1}) are (M_1^t, M_2^t).

First I demonstrate that if the selections at epoch t are a Nash equilibrium of the game with preferences $\succ_i$ then they are minimal state inferences: If (M_1^t, M_2^t) is a Nash equilibrium then $(M_1^t, M_2^t) \in \mathcal{I}_1^t \times \mathcal{I}_2^t$. Suppose (M_1^t, M_2^t) is a Nash equilibrium of the game with preferences $\succ_i$ and M_i^t is not a minimal state inference. Then there is an automaton $M_i \neq M_i^t$ that has fewer states, and since $\pi(M_i, M_j^t) = \pi(M_1^t, M_2^t)$ it attains the same repeated game payoff, a contradiction.

This establishes that M_1^t and M_2^t are both best responses to minimal state inferences. If it is shown that they are least preferred responses in the epoch $t+1$ decision problem, then due to the minimization of switching costs they will be selected at epoch $t+1$.

Suppose that player i updates at epoch $t+1$. Notice that all pairs of automata in the set $\mathcal{I}_1^t \times \mathcal{I}_2^t$ yield the same outcome path and payoff. This implies that for all inferences $M_j \in \mathcal{I}_j^t$ it is true that $(M_i^t, M_j) \succcurlyeq_i (M_i^t, M_j^t)$. Since $(B_i(M_j), M_j) \succcurlyeq_i (M_i^t, M_j)$, it follows, by transitivity, that $(B_j(M_j), M_j) \succcurlyeq_i (M_i^t, M_j^t)$ for all $M_j \in \mathcal{I}_j^t$. Thus, Condition (C) is satisfied.■

Proposition C *Assume cautious inferences. In the set of finite state automata, if* (M_1^*, M_2^*) *is a steady state of the dynamic learning process then there is a pair of minimal state inferences* $(M_1, M_2) \in \mathcal{I}_1^* \times \mathcal{I}_2^*$ *that is a Nash equilibrium of the game with preferences* $\succ_i$.

Proof. Since (M_1^*, M_2^*) is a steady state of the dynamic learning process condition (C) implies that $M_1^* \in B_1(M_2)$ for some $M_2 \in \mathcal{I}_2^*$ and $M_2^* \in B_2(M_1)$ for some $M_1 \in \mathcal{I}_1^*$. Every response in $B_i(M_j)$ has the same number of states and attains the same payoff v_i^*, thus, $\mathcal{I}_i^* \subset B_i(M_j)$. Hence, $M_1 \in B_1(M_2)$ and $M_2 \in B_2(M_1)$, which means that (M_1, M_2) is a Nash equilibrium of the game with preferences $\succ_i$.■

The content of Proposition C is that a steady state is a Nash equilibrium in beliefs; the players' inferred models of one another constitute a Nash equilibrium of the game with preferences $\succ_i$.

Proposition C also implies that every steady state is a self-confirming equilibrium in the game with preferences $\succ_i$. This follows solely from the steady state property $\pi(M_i^*, M_j) = \pi(M_1^*, M_2^*)$, where $M_i^* \in B_i(M_j)$ for some $M_j \in \mathcal{I}_j^*$. That is, each player's model is confirmed when (M_1^*, M_2^*) is played next epoch, even though the model M_j may not be the true automaton of player j. In a self-confirming equilibrium players choose a best response to their beliefs, and these beliefs only have to be consistent with the equilibrium path of play (Fudenberg and Levine (1993)). Weaker than a Nash equilibrium, the concept of self-confirming equilibrium is intrinsic to learning models in which only realized play is observed.

2.4 Convergence: Optimistic Inferences

In this section I study pairs of one and two state automata which are not steady states and show that all sequences generated by the dynamic learning process under these pairs must eventually be constant at the unique steady state.

Definition 2 *The dynamic learning process is said to have a cycle if there is a sequence of automata, and no epoch* T *such that the sequence is constant after* T.

First the properties of a cycle are characterized, and then, to prove that a cycle does not exist, these properties are used to obtain contradictions.

Lemma 3 (*Properties of a Cycle*) *Assume optimistic or cautious inferences. If a cycle exists in the set of finite state automata then:*

Property 1 *All automata of both players in the cycle must have the same number of states: there is some time T such that $|M_1^{T+t}| = |M_2^{T+t}|$, for all t.*

Property 2 *An epoch t selection of player i (the player who can update) in the cycle has the same number of states as any minimal state inference about player j in epoch $t-1$ ($|M_i^t| = C_j^{t-1}$) and is in the minimal state inference set at epoch t ($M_i^t \in \mathcal{I}_i^t$).*

Proof. Property 1. As in Lemma 1, the choice of player i in epoch t has at least as many states as any minimal state inference about player i ($M_i^t \geqq C_i^t$). Moreover, optimization implies that the choice of player j in epoch $t+1$ has no more states than the number of states in any minimal state inference about player i at epoch t ($C_i^t \geqq M_j^{t+1}$). Hence,

$$|M_2^1| \geqq |M_1^2| \geqq |M_2^3| \geqq ...$$

This is a bounded monotonic sequence. Since each term in the sequence, after the first, term can assume only one of a finite number of integers then a tail of the sequence, after some period T, is the same integer. Thus, the sequence converges to an integer. Since $|M_1^{t+1}| = |M_1^t|$ if t is even, and $|M_2^{t+1}| = |M_2^t|$ if t is odd, it follows that $|M_1^{T+t}| = |M_2^{T+t}|$ for all t.

Property 2. That $|M_i^t| = C_j^{t-1}$ follows immediately from Property 1 because there cannot be state reduction in a cycle. Likewise, it must be that $M_i^t \in \mathcal{I}_i^t$. For, suppose $|M_i^t| > C_i^t$, where C_i^t is the number of states of each automaton in $\mathcal{I}_i^t$. Then $|M_i^t| > C_1^t \geqq |M_j^t|$, a contradiction to Property 1.■

Although Property 1 is a strong restriction on a cycle, it does not follow that there cannot be a cycle; it is possible for players to choose different automata with the same number of states.

Proposition D *Assume optimistic inferences. The dynamic learning process converges to the unique steady state from any initial condition in the set of 1 and 2 state automata, for any $g, l > 0$.*

The dynamic learning process is a mapping from the set of finite state automata into itself.

The approach of the proof is to partition the set of all outcome paths (that is, all outcome paths which eventually cycle and can be represented by one and two state automata) into sets according to which of the four possible action pairs appear in the outcome path *{(C,C), (C,D), (D,C), (D,D)}*. The partition has 11 elements which are not redundant. Proposition D follows from Lemmata 3 through 10, which are equivalent to showing that all elements of the partition, except one, are transitory. The set which contains the single sequence that repeats (D,D) is the only absorbing set. Although state reduction is necessary to converge to the steady state, it does not follow that optimization always results in state reduction, as Example 3 demonstrated.

Lemma 4 *Assume optimistic or cautious inferences. There does not exist a cycle in which automata have one state.*

Proof. By Property 1 of a cycle every automaton in the sequence, after some period T, has the same number of states, Q^*. When $Q^* = 1$, $\mathcal{I}_j^t$ always consists of either COOPERATE and DEFECT, i.e. the players construct a unique model of each other. If $\mathcal{I}_j^t = \{\text{COOPERATE}\}$ then player i chooses DEFECT, under optimistic or cautious inferences. Likewise, if $\mathcal{I}_j^t = \{\text{DEFECT}\}$ then player i also selects DEFECT. This means that eventually both players always select DEFECT, which is the steady state.■

Lemma 5 *Assume optimistic inferences. There does not exist a cycle in which automata have two states and an element (M_1^t, M_2^t) of the cycle has a path $\pi(M_1^t, M_2^t)$ that consists only of the terms $\binom{C}{C}$ and $\binom{D}{D}$.*

Proof. Suppose there is such a cycle. All possible choice problems based on the sets of minimal state inferences associated to this class of paths will be considered. There are two possible sets of minimal state inferences to consider. Notice that if $\binom{C}{C}$ and $\binom{D}{D}$ are the only terms that appear in the outcome path then Defect is never observed when Cooperate is played and vice versa.

Case 1: The set of minimal state inferences at epoch $t-1$ consists of the 4 automata represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & C \\ q_2 & -_b & q_k & D \end{array}$$

Player i can attain the maximum possible payoff $(1+g)/(1-\delta)$ by choosing DEFECT. Hence, there is state reduction in a cycle, a contradiction to Property 1. The state q_k represents q_1 or q_2.

Case 2: The set of minimal state inferences at epoch $t-1$ are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_k & -_b & C \end{array}$$

Given that player j is expected to defect in the first period, player i can attain the highest possible payoff, $\delta(1+g)/(1-\delta)$, with DEFECT, a contradiction to Property 1.■

Lemma 6 *Assume optimistic inferences. There does not exist a cycle in which automata have two states and an element (M_1^t, M_2^t) of the cycle has a path $\pi(M_1^t, M_2^t)$ that consists only of the terms $\binom{C}{D}$ and $\binom{D}{C}$.*

Proof. Suppose there is such a cycle. There are two possible sets of minimal state inferences to consider. Notice that if $\binom{C}{D}$ and $\binom{D}{C}$ are the only terms that on the outcome path then Defect is never observed when Defect is played and Cooperate is never observed when Cooperate is played.

Case 1: The set of minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_k & -_b & D \end{array}$$

There is an inference for which COOPERATE attains $1/(1-\delta)$. Alternatively, if player i

plays competitively, the highest attainable payoff from an inference is $(1+g)/(1-\delta^2)$, which can be attained by DEFECT when $-_b = q_2$. Hence, for any parameters g, $l > 0$ a one state automaton is chosen at epoch t, a contradiction to Property 1.

Case 2: The minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & -_b & q_k & C \end{array}$$

It is always optimal for player i to leave player j's competitive state for free by inferring $-_a = q_2$ and choosing $\{q_1 \to D\}$. Since one state solutions lead to an immediate contradiction, assume that $q_k = q_1$ and $-_b = q_2$ to obtain the candidate two state solution, $\{q_1 \to D, q_2 \to C\}$. Player i's solution is implemented by the best response set:

$$B_i(M_j) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_c & q_2 & D \\ q_2 & q_2 & -_d & C \end{array}$$

Many possible outcome paths could result at epoch t, depending on which automaton in $\mathcal{I}_j^{t-1}$ is player j's actual automaton (recall that Property 2 of a cycle requires that one of the minimal state inferences is the true automaton) and which automaton in $B_i(M_j)$ player i chooses. To obtain a contradiction consider player j's decision problem at epoch $t+1$.

Without specifying which automaton in $\mathcal{I}_j^{t-1}$ is player j's actual automaton or which automaton in $B_i(M_j)$ player i chooses the outcome path at epoch t must start $\left\{\binom{D}{D}, \binom{C}{?}, ?\right\} = \pi(B_i(M_j), \mathcal{I}_j^{t-1})$, i.e. iterate the two incompletely specified automata until an undefined term is reached. Hence, the transition $\tau_i(q_1, D) = q_2$ for every inference about player i in $\mathcal{I}_i^t$. This means that at player j's epoch $t+1$ decision problem, the competitive state of player i can be exited for free, $\{q_1 \to D\}$, and the only solution of player j that is not one state is $\{q_1 \to D, q_2 \to C\}$. This solution is implemented by

$$B_j(M_i) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_e & q_2 & D \\ q_2 & q_2 & -_f & C \end{array}.$$

By Property 2 of a cycle, $M_i^t \in \mathcal{I}_i^t$. However, notice that for any $(M_i^t, M_j^{t+1}) \in B_i(M_j) \times B_j(M_i)$ that $\pi(M_i^t, M_j^{t+1}) = \left\{\binom{D}{D}, \binom{C}{C}, \binom{C}{C}, ...\right\}$. Under asynchronous updating, $M_i^{t+1} = M_i^t$. Thus, $\pi(M_i^{t+1}, M_j^{t+1}) = \pi(M_i^t, M_j^{t+1})$, a contradiction to Lemma 5.■

When there is a unique model of the other player, the optimal choice leads to an outcome path that cannot be part of a cycle. To establish this result I need to characterize some properties of the best response sets.

Lemma 7 *Let M_j be a finite state automaton and assume that every automaton in the best response set$B_i(M_j)$ has $|M_j|$* states. *If $M_i \in B_i(M_j)$ then the outcome path $\pi(M_i, M_j)$ is characterized by exactly one of the following:*

(a) $\pi(M_i, M_j)$ consists only of terms $\binom{D}{D}$ and $\binom{C}{C}$.

(b) $\pi(M_i, M_j)$ consists only of terms $\binom{C}{D}$ and $\binom{D}{C}$.

Proof. Similar to Theorem 1 in Piccione (1992).■

Lemma 8 *Assume optimistic inferences. There does not exist a cycle in which automata have two* states *and player i, who updates at epoch t, has a unique inference about player j at epoch $t-1$. Hence, there does not exist a cycle in which automata have two states and all four action pairs occur in the outcome path.*

Proof. The unique inference about player j must be the actual automaton of player j. By Lemma 7 $\pi(M_i^t, M_j^t)$ consists only of the terms $\binom{D}{D}$ and $\binom{C}{C}$, or only of the terms $\binom{C}{D}$ and $\binom{D}{C}$. By Lemmata 5 and 6, respectively, these outcome paths cannot be part of a cycle.

Observe that if all four action pairs occur on an outcome path and the path can be generated by a pair of two state automata then the minimal state inference sets are singletons. For example, since $\binom{C}{D}$ and $\binom{C}{C}$ both appear in the outcome path then the two transitions in the cooperative state of the top player can be deduced from the outcome path.■

Lemma 9 *Assume optimistic inferences. There does not exist a cycle in which automata have two states and an element (M_1^t, M_2^t) of the cycle has a path $\pi(M_1^t, M_2^t)$ that consists only of the terms $\binom{C}{C}, \binom{D}{C}$, and $\binom{C}{D}$.*

Proof. Notice that when the action D is played only the action C is observed. Thus, each set of minimal state inferences consists of exactly two automata that can be represented as one incompletely specified automaton. This gives rise to the following cases.

Case 1: The two minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & q_j & q_k & C \end{array}.$$

An optimistic player will infer $-_a = q_2$ and optimally choose the action D in q_1. Hence, the only two state solution to consider is $\{q_1 \to D, q_2 \to C\}$. If $q_k = q_2$, or if $q_k = q_1$ and $q_j = q_1$ this solution would not be optimal. It remains to consider $q_k = q_1$ and $q_j = q_2$:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & q_2 & q_1 & C \end{array}$$

Player *i's* solution yields the path $\left\{ \binom{D}{D}, \binom{C}{C}, \binom{C}{C}, ... \right\}$ which is implemented by:

$$B_i(M_j) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_b & q_2 & D \\ q_2 & q_2 & -_c & C \end{array}$$

The rest of the argument is identical to Case 2 of Lemma 6.

Case 2: The two minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_j & C \\ q_2 & q_k & -_a & D \end{array}$$

If $q_j = q_1$ a one state solution is chosen. Otherwise, suppose $q_j = q_2$. Notice that if player i infers $-_a = q_1$ then the choice D is optimal in q_2 for any q_k. Thus, depending on the parameters *l, g,* the solution is always a one state automaton, a contradiction.■

Lemma 10 *Assume optimistic inferences. There does not exist a cycle in which automata have two states and an element* (M_1^t, M_2^t) *of the cycle has a path* $\pi(M_1^t, M_2^t)$ *that consists only of the terms* $\binom{D}{D}, \binom{D}{C}$, *and* $\binom{C}{D}$.

Proof. Notice that when the action C is chosen only the action D is observed.

Case 1: The two possible minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_i & C \\ q_2 & q_j & q_k & D \end{array}$$

Since the initial state of a minimal state inference can't be isolated it is not possible to have $q_i = q_1$. Thus, $q_i = q_2$. If player 1 infers $-_a = q_1$, then $1/(1-\delta)$ is expected from COOPERATE. If this isn't optimal then the only candidate two state solution is $\{q_1 \to D, q_2 \to C\}$. This could only be optimal if the two minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_1 & q_2 & D \end{array}$$

Thus, the candidate two state solution only involves transitions that are common to both inferences, and hence player *j's* actual automaton. The outcome path that results at epoch t is $\left\{\binom{D}{C}, \binom{C}{D} \vdots \binom{D}{C}, \binom{C}{D}, ...\right\}$. By Lemma 6 this outcome path cannot be part of a cycle.

Case 2: Suppose the two minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_j & D \\ q_2 & - & q_k & C \end{array}.$$

Part A: $q_k = q_2$. For a two state solution the inference set must be:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & - & q_2 & C \end{array} .$$

The two state solution $\{q_1 \to D, q_2 \to C\}$ only involves transitions that are common to both inferences, and hence player *j's* actual automaton. Thus, the outcome path that player j expects at epoch t, and actually occurs, is $\left\{ \binom{C}{D}, \binom{D}{C}, \binom{D}{C}, \binom{D}{C}, ... \right\}$. By Lemma 6 this outcome path cannot be part of a cycle.

Part B: $q_k = q_1, q_j = q_2$:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_2 & D \\ q_2 & -_a & q_1 & C \end{array}$$

If player *i's* solution is two state then it must be $\{q_1 \to D, q_2 \to C\}$. If player *j's* actual automaton, M_j^{t-1}, has $-_a = q_2$ then the outcome path at epoch t is the same as the outcome path that player i expects, which leads to a contradiction. Otherwise, suppose $-_a = q_1$ in M_j^{t-1}. The optimal path of player i $\left\{ \binom{D}{D}, \binom{C}{C}, \binom{C}{C}, ... \right\}$ is implemented by:

$$B_i(M_j) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_c & q_2 & D \\ q_2 & q_2 & -_d & C \end{array}$$

The rest of the argument is identical to Case 2 in Lemma 6.

Part C: $q_k = q_1, q_j = q_1$ (This implies that $q_i = q_2$ since the initial state cannot be isolated.):

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & -_a & q_1 & C \end{array}$$

The only candidate two state solution to player i's selection problem is $\{q_1 \to C, q_2 \to D\}$. This solution does not rely on $-_a$ and hence the outcome path that player i expects at epoch t actually occurs, which leads to a contradiction.■

Lemma 11 *Assume optimistic inferences. There does not exist a cycle in which automata have two states and an element (M_1^t, M_2^t) of the cycle has a path $\pi(M_1^t, M_2^t)$ that consists only of the terms $\binom{D}{D}, \binom{C}{C}$, and $\binom{C}{D}$.*

Proof. Due to the lack of symmetry in the terms that appear on the outcome path the possible sets $\mathcal{I}_1^{t-1}$ and $\mathcal{I}_2^{t-1}$differ. Since either player could have the opportunity to update the decision problems for both players will be considered.

Case 1: Assume that player i (the player who updates*)* is the top player. Notice that player j only observes the action C when C is chosen.

Part A: The minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & -_a & C \\ q_2 & q_j & q_k & D \end{array}.$$

Player i optimally chooses COOPERATE, a contradiction.

Part B: The minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_j & D \\ q_2 & q_k & -_a & C \end{array}$$

If $q_j = q_2$ then DEFECT attains the highest feasible value for player j. Otherwise $q_j = q_1$:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & q_k & -_a & C \end{array}$$

The candidate two state solution is $\{q_1 \to C, q_2 \to D\}$, for any q_k. Player i implements this

solution with:

$$B_i(M_j) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_b & q_2 & C \\ q_2 & q_2 & -_c & D \end{array}$$

Without specifying q_k or $-_a$ in $\mathcal{I}_j^{t-1}$,or $-_b$ and $-_c$ in $B_i(M_j)$ the first three terms in the epoch t outcome path can be determined by iterating the following incompletely specified automata: $\pi(B_i(M_j), \mathcal{I}_j^{t-1}) = \left\{ \binom{C}{D}, \binom{D}{C}, \binom{D}{?}, ? \right\}$. Thus, player j's minimal state inference set about player i in epoch t must specify the following transitions and actions:

$$\mathcal{I}_i^t = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & ? & q_2 & C \\ q_2 & q_2 & ? & D \end{array}$$

It is obvious that player *j's* epoch $t+1$ selection must be COOPERATE or DEFECT.

Case 2: Now assume that the player who updates at epoch t, player i, is the bottom player. Notice that player j only observes the action D when D is chosen.

Part A: The models of player j are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_j & q_k & C \end{array}.$$

$\{q_1 \rightarrow D\}$ is optimal. For a two state solution to be optimal the transitions must be as follows:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_2 & q_1 & C \end{array}$$

However, the two state solution $\{q_1 \rightarrow D, q_2 \rightarrow C\}$ does not depend on $-_a$, which leads to a contradiction.

Part B: The models of player j are:

.

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_j & C \\ q_2 & -_a & q_k & D \end{array}$$

If $q_j = q_1$ then a one state automaton attains the highest payoff in the repeated game. Otherwise, suppose $q_j = q_2$ and $q_i = q_1$. Then COOPERATE can attain the highest feasible payoff of any solution that involves mutual cooperation. Thus, the only candidate two state solution is $\{q_1 \to D, q_2 \to C\}$. For this to be optimal the inference set must be:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_1 & q_2 & C \\ q_2 & -_a & q_2 & D \end{array}$$

Player i implements $\{q_1 \to D, q_2 \to C\}$ with:

$$B_i(M_j) = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_b & D \\ q_2 & -_c & q_1 & C \end{array}$$

If the actual automaton of player j has $-_a = q_1$, then a contradiction has been reached. Otherwise the actual automaton of player j has $-_a = q_2$. Then the outcome path $\pi(M_i^t, M_j^t)$ must begin $\left\{ \binom{D}{C}, \binom{C}{D}, \binom{D}{D}, \binom{?}{D}, ? \right\}$. No matter the value of the transition $-_b$ a contradiction has been reached with Lemma 10.

Part C: Thus, it remains to consider:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_2 & C \\ q_2 & -_a & q_k & D \end{array}$$

$\{q_1 \to D\}$ is always optimal. Thus, the candidate two state solution is $\{q_1 \to D, q_2 \to C\}$, which could be optimal only if $q_k = q_2$:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_2 & C \\ q_2 & -_a & q_2 & D \end{array}$$

These automata cannot be minimal state inferences for the class of outcome paths considered in this lemma; as the first state is transitory and the second state is absorbing, there cannot be two observations of player j playing C.■

2.5 Convergence: Cautious Inferences

Proposition E *Assume cautious inferences. The dynamic learning process converges to the unique steady state from any initial condition in the set of 1 and 2 state automata, for any* $g, l > 0$.

Lemma 12 *Assume cautious inferences. There does not exist a cycle in which automata have two states and an element* (M_1^t, M_2^t) *of the cycle has a path* $\pi(M_1^t, M_2^t)$ *that consists only of the terms* $\binom{C}{C}$ *and* $\binom{D}{D}$.

Proof. *Case 1*: Suppose the set of minimal state inferences at epoch $t-1$ for both players is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & C \\ q_2 & -_b & q_1 & D \end{array} \qquad or \qquad \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_1 & -_b & C \end{array}$$

The solution is DEFECT for all four inferences in each set for all parameters l, g. A contradiction to Property 1 of a cycle.

Case 2: Suppose the set of minimal state inferences at epoch $t-1$ for both players is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & C \\ q_2 & -_b & q_2 & D \end{array}$$

The cautious player selects a best response to a model which has a payoff no greater than a best response to any other model.

BR1: If $-_a = q_1$ the solution is DEFECT with payoff $(1+g)/(1-\delta)$.

BR2: If $-_a = q_2$ and $-_b = q_2$, the solution is DEFECT with payoff $(1+g)$.

BR3: If $-_a = q_2$ and $-_b = q_1$, the solutions depend on the parameters:

a. DEFECT with payoff $(1+g)$.

b. $\{q_1 \to D, q_2 \to C\}$ with payoff $(1+g-\delta l)/(1-\delta^2)$.

Notice that when the two state response in BR3 is a solution then the one state solution in BR2 is chosen. Thus, a one state automaton is always chosen, a contradiction to Property 1 of a cycle.

Case 3: Suppose the set of minimal state inferences at epoch $t-1$ for both players is:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & D \\ q_2 & q_2 & -_b & C \end{array}$$

BR1: If $-_b = q_2$ the solution is DEFECT, with payoff $\delta(1+g)/(1-\delta) = [\delta(1+\delta)(1+g)]/(1-\delta^2)$.

BR2: If $-_b = q_1$,the solutions depend on the parameters:

a. DEFECT with payoff $\delta(1+g)/(1-\delta^2)$.

b. $\{q_1 \to D, q_2 \to C\}$, with payoff $\delta/(1-\delta) = [\delta(1+\delta)]/(1-\delta^2)$.

If $g < \delta$ the two state automaton in BR2 is a solution and is chosen for the next epoch by a cautious player. If $g \geqq \delta$ a one state automaton is chosen for the next epoch.

When the two state solution is chosen the set of automata that implement it is exactly $\mathcal{I}_j^{t-1}$. Since this set includes M_i^{t-1}, Condition (S) in the adaptation procedure requires that

$M_i^t = M_i^{t-1}$. A steady state has been reached.

Thus, if the dynamic learning process reaches an outcome path that consists only of the terms $\binom{C}{C}$ and $\binom{D}{D}$ then it converges to a steady state.■

Lemma 13 *Assume cautious inferences. There does not exist a cycle in which automata have two states and an element (M_1^t, M_2^t) of the cycle has a path $\pi(M_1^t, M_2^t)$ that consists only of the terms $\binom{C}{D}$ and $\binom{D}{C}$.*

Proof. *Case 1*: The set of minimal state inferences at epoch $t-1$ for players i and j are derived from the path $\left\{\binom{C}{D}, \binom{D}{C}, \binom{D}{C}, ...\right\}$ and respectively represented by:

$$\mathcal{I}_i^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_2 & -_b & D \end{array} \qquad \textit{and} \qquad \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_c & D \\ q_2 & -_d & q_2 & C \end{array}$$

For all four models of player i the solution is either COOPERATE or DEFECT. A contradiction to Property 1 of a cycle.

For player j's decision problem:

BR1: If $-_a = q_2$, then DEFECT is optimal and yields payoff $\delta(1+g)/(1-\delta)$.

BR2: If $-_a = q_1$ then $\{q_1 \to C, q_2 \to D\}$ yields payoff $-l + \delta(1+g)/(1-\delta)$.

The response in BR2 does not depend on $-_a$ or $-_b$, which leads to a contradiction via Lemma 12.

Case 2: The set of minimal state inferences at epoch $t-1$ for players i and j are derived from the path $\left\{\binom{C}{D}, \binom{D}{C}\vdots\binom{C}{D}, \binom{D}{C}\vdots, ...\right\}$ and represented by:

$$\mathcal{I}_i^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_1 & -_b & D \end{array} \qquad \textit{and} \qquad \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_c & D \\ q_2 & -_d & q_1 & C \end{array}$$

Consider player *j's* choice problem against player *i*.

BR1: If $-_a = q_1$ and $-_b = q_1$, the solutions depend on the parameters:

a. DEFECT with payoff $(1+g)/(1-\delta^2)$.

b. COOPERATE with payoff $(1+\delta)/(1-\delta^2)$.

BR2: If $-_a = q_1$ and $-_b = q_2$, the solutions depend on the parameters:

a. DEFECT with payoff $(1+g)$.

b. COOPERATE with payoff $(1+\delta)/(1-\delta^2)$.

c. $\{q_1 \to D, q_2 \to C\}$ with payoff $A_2 = (1+g-\delta l)/(1-\delta^2)$.

BR3: If $-_a = q_2$ and $-_b = q_1$, DEFECT with payoff $(1+g)/(1-\delta^2)$.

BR4: If $-_a = q_2$ and $-_b = q_2$, the solutions depend on the parameters:

a. DEFECT with payoff $1+g$.

b. $\{q_1 \to D, q_2 \to C\}$ with payoff $A_1 = (1+g-\delta l)/(1-\delta^2)$.

Suppose in BR4 that the two response is optimal. Then, since $A_1 = A_2$, the responses in BR1 and BR2 indicate that a two state response is chosen for the epoch if cooperating forever is preferred to alternating $\binom{D}{C}$ and $\binom{C}{D}$ forever. Otherwise, a one state solution is selected.

Consider player *i's* choice problem against player *j*.

BR1: If $-_c = q_1$ and $-_d = q_1$, the solutions depend on the parameters:

a. DEFECT yields payoff 0.

b. $\{q_1 \to C, q_2 \to D\}$ yields payoff $A_3 = -l + \delta(1+g-\delta l)/(1-\delta^2)$.

BR2: If $-_c = q_1$ and $-_d = q_2$, the solutions depend on the parameters:

a. DEFECT yields payoff 0.

b. COOPERATE yields $A_4 = -l + \delta(1+\delta)/(1-\delta^2)$.

BR3: If $-_c = q_2$ and $-_d = q_1$, DEFECT yields payoff $\delta(1+g)/(1-\delta^2)$.

BR4: If $-_c = q_2$ and $-_d = q_2$, the solutions depend on the parameters:

a. DEFECT yields payoff $\delta(1+g)/(1-\delta^2)$.

b. $\{q_1 \to D, q_2 \to C\}$ yields payoff $\delta/(1-\delta) = \delta(1+\delta)/(1-\delta^2)$.

If a two state response is optimal in BR1 it is selected if $A_3 \leqq A_4$. This is equivalent to saying that cooperating forever is preferred to alternating $\binom{D}{C}$ and $\binom{C}{D}$ forever. The two state response in BR4 is never chosen for the epoch by a cautious player. Otherwise, a one state solution is selected at the epoch, a contradiction to Property 1 of a cycle.

Assume that player i selects the two state solution. Then this solution can be implemented exactly with the set of automata $\mathcal{I}_i^{t-1}$. Likewise, if player j selects a two state solution it can be implemented with the set $\mathcal{I}_j^{t-1}$. And for any pair of automata in $\mathcal{I}_i^t \times \mathcal{I}_j^t$ the outcome path is $\left\{ \binom{C}{D}, \binom{D}{C} \vdots \binom{C}{D}, \binom{D}{C} \vdots, ... \right\}$. Hence, any pair of automata in $\mathcal{I}_i^t \times \mathcal{I}_j^t$ is a steady state.■

Remark Lemmata 12 and 13 permit Lemma 8 to follow under cautious inferences: There does not exist a cycle in which automata have two states and all four action pairs appear on the outcome path.

Lemma 14 *Assume cautious inferences. There does not exist a cycle in which automata have two states and an element* (M_1^t, M_2^t) *of the cycle has a path* $\pi(M_1^t, M_2^t)$ *that consists only of the terms* $\binom{C}{C}, \binom{D}{C}$, *and* $\binom{C}{D}$.

Proof. *Case 1*: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & q_j & q_2 & C \end{array} \quad \text{or } \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & q_1 & q_1 & C \end{array}$$

If $-_a = q_2$ a one state automaton is the solution which contradicts Property 1 of a cycle.

If $-_a = q_1$ then a solution is a two state automaton, but the transition $-_a$ will not be used. Thus, the only information used is what is common with M_j^{t-1}. This is in the nature of a unique inference and leads to a contradiction via Lemma 13.

Case 2: The two minimal state inferences are:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & D \\ q_2 & q_2 & q_1 & C \end{array}$$

BR1: If $-_a = q_2$, the solution depends on the parameters:

a. DEFECT yields the payoff $A_1 = \delta(1+g)/(1-\delta^2)$.

b. $\{q_1 \to D, q_2 \to C\}$ yields the payoff $A_2 = \delta(1+\delta)/(1-\delta^2)$.

BR2: If $-_a = q_1$, the solution depends on the parameters:

a. COOPERATE yields the payoff $A_3 = -l + \delta(1+\delta)/(1-\delta^2)$.

b. $\{q_1 \to C, q_2 \to D\}$ yields the payoff $A_4 = -l + \delta(1+g-\delta l)/(1-\delta^2)$.

The only choice at the epoch that would not immediately lead to a contradiction is BR1b. However, for this response to be optimal it must be that $A_2 \geqq A_1$. This implies that $A_3 \geqq A_4$. Then, since $A_2 > A_3$, BR1b will not be selected.

Case 3: A one state solution is always optimal if the set of minimal state inferences is:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_1 & C \\ q_2 & q_k & -_a & D \end{array} \quad or \quad \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_2 & C \\ q_2 & q_2 & -_a & D \end{array}$$

Case 4: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_2 & C \\ q_2 & q_1 & -_a & D \end{array}$$

BR1: If $-_a = q_1$, DEFECT yields the payoff $(1+g)/(1-\delta)$.

BR2: If $-_a = q_2$, the solutions depend on the parameters:

a. DEFECT yields the payoff $(1+g)$.

b. $\{q_1 \to D, q_2 \to C\}$ does not involve the transition $-_a$.

Thus, all possible solutions will lead to a steady state.

The last case, when the two minimal state inferences are represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_1 & q_2 & C \\ q_2 & q_1 & -_a & D \end{array}$$

is very similar to Case 4.■

Lemma 15 *Assume cautious inferences. There does not exist a cycle in which automata have two states and an element* (M_1^t, M_2^t) *of the cycle has a path* $\pi(M_1^t, M_2^t)$ *that consists only of the terms* $\binom{D}{D}, \binom{D}{C}$, *and* $\binom{C}{D}$.

Proof. *Case 1*: The solution is always a one state automaton when the set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_2 & q_k & D \end{array} \quad or \mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_1 & q_1 & D \end{array}$$

Case 2: The solutions are either one state automata or do not rely on the transition $-_a$ for all minimal state inferences represented by $\mathcal{I}_j^{t-1} =$

$$\begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_a & q_2 & C \\ q_2 & q_1 & q_2 & D \end{array} \quad or \quad \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_j & D \\ q_2 & -_a & q_2 & C \end{array} \quad or \quad \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & -_a & q_1 & C \end{array}$$

Case 3: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_i & q_2 & D \\ q_2 & -_a & q_1 & C \end{array}$$

BR1: If $-_a = q_1$ then DEFECT yields the payoff $A_1 = \delta(1+g)/(1-\delta^2)$.

BR2: If $-_a = q_2$ the solution depends on the parameters:

a. DEFECT yields the payoff $A_1 = \delta(1+g)/(1-\delta^2)$.

b. $\{q_1 \to D, q_2 \to C\}$ yields the payoff $A_2 = \delta(1+\delta)/(1-\delta^2)$.

Notice that if $A_2 > A_1$ then BR2b is not selected.■

Lemma 16 *Assume cautious inferences. There does not exist a cycle in which automata have two states and an element* (M_1^t, M_2^t) *of the cycle has a path* $\pi(M_1^t, M_2^t)$ *that consists only of the terms* $\binom{D}{D}, \binom{C}{C}$, *and* $\binom{C}{D}$.

Proof. *Case 1:* Assume that player i, the player who updates at epoch t, is the top player.

Part A: The solution is DEFECT for all inferences represented by $\mathcal{I}_j^{t-1} =$

	C	D	
q_1	q_2	$-_a$	C
q_2	q_j	q_1	D

or

	C	D	
q_1	q_2	$-_a$	C
q_2	q_2	q_2	D

or

	C	D	
q_1	q_i	q_2	D
q_2	q_1	$-_a$	C

Part B: The set of minimal state inferences is represented by:

$\mathcal{I}_j^{t-1} =$

	C	D	
q_1	q_2	q_2	D
q_2	q_2	$-_a$	C

BR1: If $-_a = q_2$ then DEFECT yields $A_1 = \delta(1+g)(1+\delta)/(1-\delta^2)$.

BR2: If $-_a = q_1$ the solution depends on the parameters:

a. DEFECT yields the payoff $A_1 = \delta(1+g)/(1-\delta^2)$.

b. The solution $\{q_1 \to D, q_2 \to C\}$ does not depend on the transition $-_a$.

Part C: The set of minimal state inferences is represented by:

$\mathcal{I}_j^{t-1} =$

	C	D	
q_1	q_2	q_1	D
q_2	q_1	$-_a$	C

BR1: If $-_a = q_1$ the solution depends on the parameters:

a. DEFECT yields the payoff zero.

b. $\{q_1 \to C, q_2 \to D\}$ yields the payoff $-l + \delta(1+g-\delta l)/(1-\delta^2)$.

BR2: If $-_a = q_2$ the solution depends on the parameters:

a. DEFECT yields the payoff zero.

b. $\{q_1 \to C, q_2 \to D\}$ yields the payoff $-l + \delta(1+g)/(1-\delta)$.

Thus, if DEFECT is not chosen at the epoch then BR1b is selected. This solution can be implemented with:

$$B_i = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_b & q_2 & C \\ q_2 & q_1 & -_c & D \end{array}$$

If $-_a = q_1$ in M_j^{t-1}, so that player i's inference is correct then the process will converge to a steady state by Lemma 13. Otherwise $-_a = q_2$ in M_j^{t-1} and $\pi(B_i, M_j^t) = \left\{\binom{C}{D}, \binom{D}{C}, \binom{C}{C}, \binom{?}{D}, ?\right\}$. This means that no matter what automaton in B_i is selected at epoch t by player i that the 4th play will lead us to a contradiction by the remark after Lemma 13, or by Lemma 14.

Part D: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & q_2 & -_a & C \end{array}$$

BR1: If $-_a = q_1$ the solution depends on the parameters:

a. COOPERATE yields the payoff $-l + \delta/(1-\delta)$.

b. $\{q_1 \to C, q_2 \to D\}$ yields the payoff $-l + \delta(1 + g - \delta l)/(1 - \delta^2)$.

BR2: If $-_a = q_2$ the solution depends on the parameters:

a. DEFECT yields the payoff zero.

b. $\{q_1 \to C, q_2 \to D\}$ yields the payoff $-l + \delta(1+g)/(1-\delta)$.

If BR1b is selected it can be implemented with:

$$B_i = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & -_b & q_2 & C \\ q_2 & q_1 & -_c & D \end{array}$$

If $-_a = q_1$ in M_j^{t-1}, so that player *i's* model of player j is correct then the outcome path

next epoch will result in a contradiction via Lemma 13. Otherwise $-_a = q_2$ in M_j^{t-1} and $\pi(B_i, M_j^{t-1}) = \left\{ \binom{C}{D}, \binom{D}{C}, \binom{C}{C}, \binom{?}{C}, \right\}$. This means that no matter what automaton in B_i is selected that the 4th term in the outcome path will to a contradiction via Lemma 14.

Part E: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_a & C \\ q_2 & q_1 & q_2 & D \end{array}$$

BR1: If $-_a = q_1$ the solution DEFECT yields the payoff $(1+g)/(1-\delta)$.

BR2: If $-_a = q_2$ the solution $\{q_1 \to D, q_2 \to C\}$ yields the payoff $(1+g-\delta l)/(1-\delta^2)$.

When the two state solution is selected it can be implemented with:

$$B_i = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & -_b & D \\ q_2 & -_c & q_1 & C \end{array}$$

If $-_a = q_2$ in M_j^{t-1}, so that player i's model of player j is correct then the outcome path next epoch will result in a contradiction via Lemma 13.

Otherwise $-_a = q_1$ in M_j^{t-1} and $\pi(B_i, M_j^{t-1}) = \left\{ \binom{D}{C}, \binom{C}{C}, \binom{?}{D}, \right\}$.

If $-_c = q_2$ then the path is $\pi(B_i, M_j^{t-1}) = \left\{ \binom{D}{C}, \binom{C}{C}, \binom{C}{D} \vdots \binom{D}{C}, \right\}$ which, by Lemma 14, results in a contradiction.

If $-_c = q_1$ and $-_b = q_2$ then the path is $\pi(B_i, M_j^{t-1}) = \left\{ \binom{D}{C}, \binom{C}{C}, \binom{D}{D}, \binom{C}{D} \vdots, \right\}$ which will result in singleton inference sets and thus a contradiction.

If $-_c = q_1$ and $-_b = q_1$ then the path is $\pi(B_i, M_j^{t-1}) = \left\{ \binom{D}{C}, \binom{C}{C}, \binom{D}{D}, \binom{D}{D}, \right\}$. Player j infers that player i used one of the automata:

$$\mathcal{I}_i^t = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_2 & q_1 & D \\ q_2 & q_1 & - & C \end{array}$$

By *Part C* of *Case 1* of this Lemma a contradiction has been reached.

Case 2: Assume that player i, the player who updates at epoch t, is the top player. Notice that player j only observes the action D when D is chosen.

Part A: All solutions are either one state solutions or two state solutions that do not depend on the transition $-_a$ when the set of minimal state inferences is:

$\mathcal{I}_j^{t-1} =$

	C	D	
q_1	$-_a$	q_2	D
q_2	q_j	q_k	C

Part B: The solution to every inference is a one state automaton when the sets of minimal state inferences are $\mathcal{I}_j^{t-1} =$

	C	D	
q_1	q_2	q_2	C
q_2	$-_a$	q_1	D

or

	C	D	
q_1	q_2	q_1	C
q_2	$-_a$	q_k	D

or

	C	D	
q_1	q_1	q_2	C
q_2	$-_a$	q_1	D

Part C: The set of minimal state inferences is represented by:

$\mathcal{I}_j^{t-1} =$

	C	D	
q_1	q_2	q_2	C
q_2	$-_a$	q_2	D

BR1: If $-_a = q_1$ the solution depends on the parameters:

a. DEFECT yields the payoff $(1 + g)$.

b. $\{q_1 \to D, q_2 \to C\}$ yields the payoff $(1 + g - \delta l)/(1 - \delta^2)$.

BR2: If $-_a = q_2$ then DEFECT yields the payoff $(1 + g)$.

Notice that if the two state response is ever a solution then it is not chosen at the epoch.

Part D: The set of minimal state inferences is represented by:

$$\mathcal{I}_j^{t-1} = \begin{array}{c|cc|c} & C & D & \\ \hline q_1 & q_1 & q_2 & C \\ q_2 & -_a & q_2 & D \end{array}$$

BR1: If $-_a = q_1$ the solution depends on the parameters:

a. COOPERATE yields $(1+\delta)/(1-\delta^2)$.

b. $\{q_1 \to D, q_2 \to C\}$ yields the payoff $(1+g-\delta l)/(1-\delta^2)$.

BR2: If $-_a = q_2$ the solution depends on the parameters:

a. COOPERATE yields the payoff $(1+\delta)/(1-\delta^2)$.

b. DEFECT yields the payoff $(1+g)$.

Notice that if the two state response is ever a solution then it is not chosen at the epoch.■

2.6 Conclusion

The paper has discussed the issue of stationarity in a game – the infinitely repeated Prisoners' Dilemma – in which a multiplicity of equilibria exists. A set of behavioral assumptions has been identified which leads players to choose stationary strategies. In particular, the players' behavior is restrained by their computational abilities: they use strategies represented by finite state automata and have a preference for simple automata. By themselves, these assumptions do not imply that players will choose stationary strategies. It is the limited foresight of the players, together with the concomitant incentives to adapt their strategies and use of a rule of thumb – optimistic inferences – for modeling each other, by which they learn to play stationary strategies.

2.7 References

Aumann, Robert (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern.* Mannheim, Bibliographisches Institut.

Aumann, Robert (1997): "Rationality and Bounded Rationality: The 1986 Nancy L. Schwartz Memorial Lecture," *Games and Economic Behavior*, 21, 2-14.

Abreu, Dilip, and Ariel Rubinstein (1988): "The Structure of Nash Equilibria in Repeated Games with Finite Automata," *Econometrica*, 56, 1259-1282.

Binmore, Kenneth, and Larry Samuelson (1992): "Evolutionary Stability in Repeated Games Played by Finite Automata," Journal of Economic Theory, 57, 278-305.

Birkhoff, Garrett, and Thomas Bartee (1970): Modern Applied Algebra. New York: McGraw-Hill.

Chatterjee, Kalyan, and Hamid Sabourian (2000): "Multiperson Bargaining and Strategic Complexity," *Econometrica*, 68, 1491-1509.

Fudenberg, Drew, and David Levine (1993): "Self-Confirming Equilibrium," *Econometrica*, 61, 523-545.

Fudenberg, Drew, and David Levine (1998): *The Theory of Learning in Games.* Cambridge, Mass.: The MIT Press.

Fudenberg, Drew, and Eric Maskin (1993): "Evolution and Repeated Games," Harvard Working Paper.

Kalai, Ehud, and William Stanford (1988): "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, 397-410.

Maskin, Eric, and Jean Tirole (1997): "Markov Perfect Equilibrium, I: Observable Actions," manuscript.

Maskin, Eric, and Jean Tirole (2001): "Markov Perfect Equilibrium: I. Observable Actions," *Journal of Economic Theory*, 100, 191-219.

Osborne, Martin and Ariel Rubinstein (1994): *A Course in Game Theory.* Cambridge, Mass.: The MIT Press.

Rubinstein, Ariel (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 39, 83-96.

Rubinstein, Ariel (1991): "Comments on the Interpretation of Game Theory," *Econometrica*, 59, 909-924.

Rubinstein, Ariel (1998): *Modeling Bounded Rationality.* Cambridge, Mass.: The MIT Press.

Sabourian, Hamid (2000): "Bargaining and Markets: Complexity and the Walsrasian Outcome," *Cowles Foundation Discussion Paper.*

Spiegler, Ran (2001): "Equilibrium in Justifiable Strategies: A Model of Reason-Based Choice in Extensive-Form Games," manuscript.

Chapter 3

Negotiation in Repeated Games

3.1 Introduction

This paper is concerned with the question of how bargaining can be used to select equilibria in repeated games. One of the most salient results in the theory of repeated games is the folk theorem: any individually rational and feasible payoff vector of a two player normal form game is a perfect equilibrium outcome of the repeated game when the discount factor is sufficiently near one (Fudenberg and Maskin, 1986).[1] An implication of the folk theorem for economic theory is that long-term competition situations are characterized by a multiplicity of enforceable outcomes. In bargaining problems, however, the existence of a multiplicity of enforceable outcomes is an ingredient. A bargaining process selects a particular allocation from the set of outcomes that is subsequently enforced by an outside power or the players themselves.

The model in this paper can be viewed as a model of renegotiation in repeated games in which the bargaining process is explicitly modeled. The players engage in negotiations while they play a repeated game. The players follow a current agreement – a repeated game strategy – while they exchange messages to try to reach a new agreement. The negotiations, which have no direct payoff consequences themselves, are an alternating offers bargaining protocol. The messages comprise the language that players use to negotiate with each other. We investigate a class of perfect equilibria, negotiation-compatible equilibria, in which the strategies followed

[1] The full dimensionality condition is a sufficient condition for the folk theorem to hold when there are three or more players.

in the repeated game are consistent with the outcome of the players' negotiations. We find in Theorem 1 that when the discount factor is sufficiently near 1 that all negotiation-compatible equilibria in the model are nearly efficient and that the players eventually play a repeated game strategy that is efficient. Hence, the only set of enforceable payoffs which players will not negotiate away from is the Pareto frontier.

The behavior of members of a cartel is illustrative of the situation modeled in this paper. In a cartel players vie to obtain their preferred outcomes by simultaneously choosing actions that are directly payoff relevant while negotiating with each other to attempt to reach a compromise outcome. An outcome for a cartel is not merely a division of a fixed quantity of a resource – it is a division of an output whose magnitude is determined by the joint actions of the players. An agreement by members of a cartel is rarely enforced by legal means, rather it is enforced by the players' threats which necessarily consist of payoff relevant actions. These self-enforcing strategies guarantee each player his or her share of the allocation and deter encroachment by the opponent.

The game can be described as follows. At each stage players (1) choose actions in a simultaneous move stage game that are prescribed by a prevailing agreement, and (2) negotiate with each other, relative to some threat point, to attempt to reach a compromise on a new outcome and the terms to enforce it.

The bargaining protocol is based on the protocol of the Rubinstein alternating offers model (1982). In Rubinstein's model the players alternate between the roles of proposer and responder: A proposer tenders an offer which is followed with a response by the responder. An acceptance results in each player receiving his or her part of the offer. A rejection leads to a repetition of the process with the players switching roles. The outcome of the bargaining process is a payoff vector or a disagreement outcome.

The bargaining protocol we use also permits a player to make an offer after he or she deviates from the prevailing agreement. In this situation – a challenge to the current agreement – the offer of the challenger is intended to present the opponent with an ultimatum: accept my demands or make good on your threat. The bargaining protocol requires that the opponent respond to a challenge. If the defender accepts this offer then the players switch to a new agreement whose payoff is this offer. If the opponent rejects this offer then the players choose

actions according to what the prevailing agreement prescribes at the subgame that follows the deviation. As the name *challenge* suggests it is a demand which, together with the deviation, is an attempt to improve the bargaining position of the challenger. Put another way, a challenge also signals an intent to renegotiate the current agreement. By deviating, a challenger potentially changes the bargaining positions of the players that are embodied in the state of the current agreement; players have the power to change the threat point.

There are several important differences between Rubinstein's alternating offers model and its application in a repeated game framework, with regard to enforcement and commitment in particular. First, with regard to enforcement, in the baseline model an accepted offer is enforced by legal means – means that are exogenous to the model. Enforcement is with respect to the ownership rights of a particular division of a divisible object, such as a piece of land. An accepted offer in a repeated game context, however, is a continuation payoff that must be enforced by a repeated game strategy – means that are endogenous to the model and effectively substitute for legal enforcement. Here enforcement is with respect to an outcome that is determined by the players' joint actions, such as a total quantity of oil ready to be sold in a market. It is as if the players offer both a payoff vector and the terms of the agreement in their negotiations, and an acceptance is an acceptance of the entire agreement, not just the payoff. Indeed, this viewpoint is implicit in any solution to the model; it would be redundant for the players to actually offer agreements as part of their messages.

Second, with regard to commitment, in the baseline alternating offers model offers and responses are merely messages exchanged in negotiations whose meaning is entirely derived from the fact that the messages are commitments enforced by legal means. An offer tendered by a proposer is a commitment on the part of the proposer that is enforced by legal means were the responder to accept the offer. Likewise, an acceptance by a responder is a commitment on the part of the responder that is enforced by legal means. In a repeated game there is no legal enforcement to impart any meaning to the messages exchanged by the players during their negotiations. Since the model in this paper is a standard repeated game with additional moves that are messages there exist equilibria in which the players ignore all messages. For example, without the power of legal enforcement there are strategies in which players always offer the payoff vector x, always accept x, and always choose in order to obtain the payoff vector $y \neq x$

in each stage.

We will focus our investigation on equilibria in which the messages equip the players with a language to conduct negotiations and the messages are meaningful in this context. In these *negotiation-compatible equilibria* players have common expectations about the meaning of the messages they use in negotiations; at the simplest level these expectations impose that "yes" means yes, "no" means no, and payoff "x" means payoff x. In a negotiation-compatible equilibria, if the responder says "yes" to an offer x then the new agreement actually yields x. In addition, if the agreement that enforces this offer x is self-enforcing then both players expect the other player to subsequently follow the new agreement. Likewise, if the responder says "no" to an offer x and the current agreement is self-enforcing then both players expect the other player to subsequently follow the current agreement. Without these common expectations a self-enforcing agreement is not necessarily a meaningful way to consummate a deal and is not equivalent to legal enforcement.

Disagreement outcomes are not independent outside options as they are in the baseline alternating offers model. The consequence of disagreement – which occurs when an offer is rejected – is that players continue to follow the current agreement. Once the players adopt a self-enforcing agreement it is feasible to continue with this agreement indefinitely, even if it is Pareto dominated by another agreement, by having one player veto the negotiations. A player can always effectively veto the negotiations by rejecting offers that are not identical to the current payoff vector d and by always tendering the offer d. Under this veto the opponent necessarily does not have a profitable deviation and both players anticipate receiving the payoff d of the current agreement. Thus, the alternating offers mechanism endows the bargaining process with the principle of market theory that a potential trade in a free market can be vetoed by one of the participants – you walk away, opting for the status quo. An advantage of modeling the bargaining process explicitly is that negotiation is a choice, which stands in contrast to models which only make assumptions about the outcome of the bargaining process.

Farrell and Maskin (1989) use a postulate from bargaining theory to refine the set of subgame perfect equilibria. However, the postulate that they use, a requirement that the outcome of the bargaining process be Pareto efficient, is a group rationality concept. In their model they seek to reconcile the goal of obtaining Pareto efficient equilibria in repeated games with

the multiplicity problem by characterizing the solution set that results under this postulate. We further their investigation of renegotiation in repeated games by attempting to more fully integrate bargaining theory into repeated games in order to investigate the interdependence of the bargaining problem with strategies and events in the repeated game.

Ray (1994), modifies the requirements of weak renegotiation proof equilibria with the stronger notion of internally renegotiation-proof sets (IRP). He shows that as the discount factor approaches one that the limit IRP sets are either singletons or subsets of the Pareto frontier, provided an IRP set exists. Datta (1994) demonstrates an example in which an IRP set fails to exist for all sufficiently high discount factors. Ray's limiting result also suggests that the issue of external consistency becomes negligible when the discount factor is high.

There has not been a consensus on what the proper notions of internal and external consistency are for renegotiation proof sets in infinitely repeated games. Yet, when the bargaining process of the players in an infinitely repeated game is explicitly modeled the issue of internal and external consistency is replaced by different issues – issues related to solving the noncooperative game instead of issues about what assumptions to impose on the outcomes of an undefined behavioral process. In this regard, the approach to equilibrium selection taken in this paper differs from the approach taken in the renegotiation literature. The alternating offer protocol of bargaining is used to model *how* the players make decisions. This protocol is as simple as it is rich: it is a model of bilateral communication and a noncooperative mechanism for switching between equilibria, as well as its usual role as a model of an institution. When long-term competition is modeled by a repeated game and the bargaining process is modeled by alternating offers players can negotiate to efficiency.

Busch and Wen (1995) study a bargaining model in which players follow the alternating offers protocol, bargain over a pie, and, whenever an offer is rejected, play a one-shot game to determine the payoff in that stage. Although their model is not concerned with negotiation over outcomes of a repeated game it contributes to this investigation by demonstrating the relevance of Rubinstein's alternating offers model to a variety of dynamic settings. In particular, they derive the solution to the alternating offers model when the sequence of disagreement payoffs is nonstationary.

3.2 Model

3.2.1 Preliminaries

The game, denoted Γ, is a two person extensive game. The stage game consists of a finite action set A_i with elements a_i and a utility function $u_i : A \longrightarrow \mathbb{R}$ for each player, where A is the product set of the players' action sets and has elements a. The mapping $u : A \longrightarrow \mathbb{R}^2$ assigns a point in utility space to each outcome in A. The minmax payoff of player 1 is:

$$v_1 := \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$$

The minmax payoff of player 2 is defined analogously. We assume throughout that the stage game is scaled so that $v_i = 0$. The set of possible payoffs is $con(u(A))$. Define the set of feasible and strictly individually rational payoff vectors of the stage game to be the set

$$V^* = \{x \in con(u(A)) : x_i > 0\}.$$

In every proposition we assume that the Pareto frontier of V^* can be represented by a bijective function $g : [0, \overline{u}_1] \longrightarrow R$ with $g(0) > 0$ and $g(\overline{u}_1) = 0$. This assumption means that each player has a unique most preferred agreement in the closure of V^* which is located on the player's axis. For $d \in V^*$ let $S(d) = \{x \in V^* : x_i \geqq d_i\}$.

An infinitely repeated game is obtained by playing the stage game across time $(0, 1, 2, ...)$. A strategy σ_i in the repeated game specifies an action for each finite history and a strategy vector $\sigma = (\sigma_1, \sigma_2)$ determines a unique *outcome path* $\mathbf{a}(\sigma)$ which is a history of action pairs, $\{a(t)\}_{t=1}^{\infty}$, induced by σ. The payoff to each player in the repeated game is the discounted average $(0 < \delta < 1)$ of the sequence of payoffs determined by $\mathbf{a}(\sigma)$

$$U_i(\mathbf{a}(\sigma)) := (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(a(t)).$$

A strategy $\sigma(d)$ has payoff vector $d := (U_1(\mathbf{a}(\sigma(d)), U_2(\mathbf{a}(\sigma(d)))$.

When the players use a correlated strategy in the stage game to realize a payoff vector the definitions have natural analogues.

One method by which individuals resolve conflicts is to reach a compromise through bargaining. The data of the bargaining problem consists of a set of *outcomes* $S \subset V^*$, an initial *threat point* $d^0 \in S$, the set of strategies $\Sigma(S)$ in the repeated game which yield a payoff in the set S, and a *bargaining protocol* which specifies the rules by which the players bargain with each other.

Players have two kinds of messages which comprise their negotiation language. Players can propose *offers* which are outcomes in S. The offers of players 1 and 2 are denoted x and y, respectively. Players can *accept* or *reject* offers, denoted Y and N, respectively. Neither offers nor responses are directly payoff relevant, as indicated by the definition of $U_i(\mathbf{a}(\sigma))$.

3.2.2 Bargaining in Repeated Games

The bargaining protocol is used as a noncooperative mechanism for switching between self-enforcing strategies in the repeated game in addition to being a mechanism to select from a multiplicity of enforceable outcomes. The outcome of the bargaining problem is an outcome $d \in S$ which is enforced by a repeated game strategy, $\sigma(d)$. A repeated game strategy is referred to as an *agreement in* Γ to differentiate it from a strategy in Γ which is denoted ϕ. A strategy ϕ includes stage game actions and negotiations, whereas agreements σ only include stage game actions. Players negotiate to reach new agreements. The messages in ϕ determine *which* agreement is followed, but an agreement only depends on the past actions in the stage game which were taken since the agreement was adopted.

To impose structure on how players exchange messages – who can say what when – we use a bargaining protocol. As in the baseline alternating offers model, negotiations are required to follow the bargaining protocol as if there were an arbitrator present to enforce the protocol. In a *stage* of Γ the stage game is played and players potentially exchange messages. The order of these three substages is always (1) stage game, (2) proposal, (3) response.

Definition 1 *The* **bargaining protocol** *of* Γ :

BP1. *Player 1 makes the first offer at the start of* Γ *and after any acceptance.*

BP2. *When neither player deviates from the current agreement players continue to alternate between the roles of proposer and responder.*

BP3. *When a player deviates from the current agreement an offer must follow it: a challenge. The opponent responds. If the opponent rejects the challenge the opponent has the privilege of making the next offer, according to* **BP**2 *or* **BP**3.

BP4. *If both players simultaneously deviate from the current agreement, then there are no proposals or responses in this stage and in the next stage of the game they continue under* **BP**2.

It is convenient to define a **challenge** to be the offer a player makes after deviating from the current agreement.

The bargaining protocol is part of the definition of the extensive game Γ. In Γ, an acceptance is logically equivalent to the current agreement σ appearing as a message in the history: at the time of an acceptance both players have given their consent to the current agreement as if they had also accepted the terms of the agreement. It is irrelevant to Γ which agreement σ is the current agreement, all that is relevant is whether or not the agreement is followed. If it is followed Γ proceeds as in BP2, if it is not followed Γ proceeds as in BP3.

Another viewpoint is that Γ is purely a bargaining game whose inputs are a set of offers and an agreement to enforce each offer. It is the analogue of Rubinstein's alternating offers model when players must enforce the outcomes of the bargaining process themselves. In order to solve Γ we must also solve a repeated game.

To complete the definition of Γ we will define the possible types of histories. At the initial history, denoted $\emptyset$, the players choose actions in the stage game according to some initial agreement $\sigma^0(d^0)$. Player 1 makes an offer to player 2 after the stage game is played.

A length-3 history in which neither player has challenged the initial agreement has the form

$$\emptyset, a^0(1), x^1.$$

If player 2 rejects the offer, stage 1 is represented by the length-4 history

$$\emptyset, a^0(1), x^1, N.$$

In stage 2 the players will reverse roles

$$\emptyset, \vdots \underbrace{a^0(1), x^1, N,}_{\text{STAGE 1}} \vdots a^0(2), y^2.$$

The pair of actions t stages into the initial agreement is $a^0(t)$, the superscript is an index of the agreements and indicates that players are following the initial agreement $\sigma^0(d^0)$. The superscript on the offers is an index over stages. Recall that the proposals and responses occur in between plays of the stage game and are classified in the same stage as the previous play of the stage game. The time the players need for proposals and responses does not alter the discounting of the sequence $\{u_i(a(t))\}_{t=0}^{\infty}$.

The general form of a history that ends in a rejection by player 1 and does not involve a deviation from the current agreement has the form

$$\emptyset, \vdots a^0(1), x^1, N, \vdots a^0(2), y^2, N, \vdots ..., \vdots a^0(t), y^t, N$$

and, when player 1 accepts,

$$\emptyset, \vdots a^0(1), x^1, N, \vdots a^0(2), y^2, N, \vdots ..., \vdots a^0(t), y^t, Y.$$

After an offer d^k is accepted the process repeats: a new agreement $\sigma^k(d^k)$ is followed and they can negotiate towards another agreement

$$\emptyset, \vdots a^0(1), x^1, Y, \vdots a^1(1), y^2, N, \vdots ..., \vdots a^1(t-1), x^t, N, \vdots a^1(t), y^{t+1}, N.$$

The time index on action profiles starts at 1 when a new agreement is adopted to correspond to the outcome path of the new agreement,

$$\mathbf{a}(\sigma^1(x^1)) = (a^1(1), a^1(2), ...).$$

Other types of histories in which there are no deviations from the current agreement can be defined analogously. These include histories in which no tendered offer is ever accepted,

histories that end in a response by player 2, and histories that end in an offer instead of a response.

There are also types of histories in which a player deviates from the current agreement while the bargaining protocol is being followed. The symbol$\widehat{\ }$ on an action signifies that a deviation from play prescribed by the current agreement has occurred. An offer d^k tendered after a deviation $\widehat{a}_i$ from the current agreement will be referred to as a *challenge by player i.*

Histories which involve deviations from the current agreement are depicted below. These histories illustrate the cases when player 2 has challenged the current agreement. Histories in which player 1 has challenged the current agreement are defined analogously.

A history in which player 2 challenges the current agreement during negotiations, player 1 accepts the offer, and they continue to bargain is represented as

$$\emptyset, \begin{pmatrix} a_1^0(1) \\ \widehat{a}_2 \end{pmatrix}, y^1, Y, \vdots a^1(1), x^2, N.$$

A history in which player 2 challenges the current agreement during negotiations, player 1 rejects the offer, and then they negotiate again is represented as

$$\emptyset, \begin{pmatrix} a_1^0(1) \\ \widehat{a}_2 \end{pmatrix}, y^1, N, \vdots \widehat{a}^0(2), x^2, N,$$

where $\widehat{a}^0(2)$ indicates play by $\sigma^0(d^0)$ after the previous deviation.

A strategy of a player, denoted ϕ_i, maps histories to stage game actions and messages, the latter in accordance with the bargaining protocol. After any offer is accepted or rejected ϕ prescribes the agreement $\sigma(d)$ that is subsequently followed and the messages exchanged in negotiations towards another agreement.

3.3 Characterizations of the Solution

3.3.1 Negotiation-Compatible Equilibria

The solution concept is subgame perfect Nash equilibria. We will restrict attention to the perfect equilibria which are compatible with the players' negotiations.

Definition 2 *A perfect equilibrium ϕ of Γ is said to be* **negotiation-compatible** *if it satisfies:*

NC1. *When a responder rejects an offer the current agreement is played in the subgame that follows.*

NC2. *When a responder accepts an offer d an agreement $\sigma(d)$ is played in the subgame that follows.*

Conditions **NC**1 and **NC**2 imply that it is common knowledge between the players which strategy the players will follow in the subgame after a response. Negotiation-compatible equilibria rule out equilibria in which player i responds with Y to the message $d' \in S$ then the players ignore their own messages and choose according to some strategy $\sigma(x)$ with $x \neq d'$. This means they also rule out equilibria in which both players expect that the other player will follow the current agreement $\sigma(d)$ when the responder has just accepted a new agreement.

NC1 and **NC**2 eliminate the need for an assumption that an exogenous power enforces the offer, or the need for an assumption that an exogenous power must enforce the players' self-enforcing agreements, which would be rather paradoxical. In a negotiation-compatible equilibrium the players can walk away from the bargaining table after designing a new self-enforcing agreement and expect that the agreement will be followed in the immediate subgame – they don't need an arbitrator to say "Go!".

$NC(\delta)$ denotes the set of negotiation-compatible payoffs when the discount factor is δ. It is a subset of the set of perfect equilibrium payoffs of Γ.

Negotiation-compatible equilibria have the property that after an offer d is accepted the agreement that follows has the payoff d. In particular, challenges never occur on the outcome path determined by a negotiation-compatible equilibrium ϕ, or on the outcome path of ϕ at any subgame after an acceptance. Moreover, to not contradict condition **NC**2, in a negotiation-compatible equilibrium players will only accept offers that can be enforced by agreements for which it is not optimal for a player to deviate from the agreement and tender an offer. These agreements must be perfect equilibria in a repeated game, for otherwise a player could gain by deviating and proposing an absurd offer that is surely rejected. In addition, these agreements must not leave open an opportunity for a player to deviate and follow up with an offer that

is accepted. Other offers are not credible and cannot be part of a negotiation-compatible equilibrium. In other words, in a negotiation-compatible equilibrium it is common knowledge between the players which offers are "serious offers."

To illustrate why challenges do not occur in a negotiation compatible equilibrium, consider the case when the players want to design an agreement to enforce a payoff on the Pareto frontier. It must not only be robust to deviations, it must be robust to deviations that are followed by an offer, that is, a challenge. If it is not robust to a challenge from player 1, say, then player 1 can profit from a stratagem in which he or she deviates from the agreement and tenders a stingy offer designed to provide a sufficient incentive to prevent the opponent from retaliating with a costly punishment – to buy off the punishment. The end result would be a net gain for player 1, a net loss for player 2, and a movement towards player 1's most preferred agreement in V^*. Thus, a minimal standard for an agreement to enforce a payoff that is Pareto efficient is that no challenges occur.

Next, consider the case when players want to design an agreement to enforce a payoff that is not Pareto efficient. For this agreement to be enforced then necessarily a player does not have an incentive to deviate from it and tender an offer that is expected to be rejected. Moreover, and analogous to the Pareto efficient case, there must not be an incentive to deviate from the agreement and tender an offer that would be accepted. This is a type of no-trade condition which depends on the outcome of negotiations when neither player deviates relative to the effective threat point of the current agreement that results after a deviation. In a negotiation-compatible equilibrium, the balance of power between the players is such that they prefer to bargain at their current position rather than fail in an attempt to obtain a better position by force. A main problem in solving the model is to show the existence of such challenge-free agreements that can be used to enforce accepted offers in negotiation-compatible equilibria.

Observe that players may negotiate to a new agreement on or off the outcome path of a negotiation-compatible equilibrium ϕ. Players follow the current agreement while they negotiate, and when an offer is accepted they switch to a new agreement. This is partly a consequence of the alternating offers model of bargaining, which provides a process to switch between self-enforcing agreements. Off the equilibrium outcome path, a strategy ϕ could potentially include an infinite number of agreements which the players plan to follow. In Theorem 1 we will see

that when the discount factor is sufficiently large every negotiation-compatible equilibria ϕ of Γ has at most two distinct agreements on its outcome path; if the first agreement is not Pareto efficient the players will negotiate to a Pareto efficient agreement.

3.3.2 Efficient Renegotiation

Definition 3 *An equilibrium is said to be* **nearly efficient** in V^* *if either it is efficient or else it is efficient after the first stage.*

Theorem 1 *If $d^0 \in V^*$ then there exists a $\underline{\delta} \in (0,1)$ such that for each discount factor $\delta \in [\underline{\delta}, 1)$:*

(1) *There exists a nearly efficient negotiation-compatible perfect equilibrium of Γ in which the initial agreement has payoff d^0.*

(2) *All negotiation-compatible equilibria in which the initial agreement has payoff d^0 are nearly efficient.*

Theorem 1 follows from Propositions 1 and 2. Proposition 1 proves existence of negotiation compatible equilibria and Proposition 2 demonstrates near efficiency.

The outcome path of a negotiation-compatible equilibrium under Theorem 1 has at most two distinct agreements, and the second agreement has a Pareto efficient payoff. Moreover, it will be shown that at any accepted offer in $S(d^0)$ the players will negotiate to an efficient payoff. Hence, the only subset of $S(d^0)$ that is not "renegotiable" is its frontier. One of the roles of the high discount factor is to permit each offer to be enforced by an agreement which no player will challenge. Yet, the set of negotiation-compatible equilibria does not grow with the discount factor because the bargaining process selects a particular payoff from the set of all enforceable payoffs.

The near efficiency result is an implication of the alternating offers model of bargaining. The ability to apply the alternating offers protocol in this environment depends on four factors. First, the set of possible outcomes of the bargaining process are now continuation payoffs, rather than divisions of a pie, and must be enforced without legal enforcement. Second, we need to define negotiation-compatible equilibria in order for the players negotiations to not just be cheap talk. Third, to obtain regularity properties on the set of enforceable offers, we

make claims that certain convex sets in V^* are enforceable for a fixed discount factor, which is a uniformity result instead of a folk theorem result. Fourth, the negotiation process is continual, in particular, players are able negotiate after each acceptance. Each of these factors differentiate this bargaining model from Rubinstein's model and also from the model of Busch and Wen.

There are additional consequences from having the players negotiate with alternating offers in a repeated game. The set of negotiation-compatible equilibria is distinct from the set of perfect equilibria of the repeated game. It is these consequences of using the alternating offers protocol which are of the most interest from the standpoint of investigating issues in renegotiation. For the question of how a repeated game should be played should not be divorced from the question of how individuals should bargain in long-term competition. The strategy in the repeated game – in particular, the threats – should be coordinated with the bargaining strategy in order to protect the players' bargaining positions and achieve their bargaining goals. A strategy in a situation of long-term competition should coordinate negotiations with actions rather than delegating production and pricing to individual A and negotiation to individual B without analyzing their interdependence. It is difficult to investigate this interdependence by studying repeated games in the absence of an explicit model of bargaining. The use of an alternating offers protocol in a repeated game is our attempt to explicitly model long-term competition and bargaining together. The example could be cartels seeking to restrict trade, farmers seeking area in the commons for their cows, or politicians from different parties seeking to achieve their policy goals by trading votes with each other on various bills. What is common to situations of long-term competition is the central role that bargaining plays in determining the outcomes – the strategy of the players in the repeated game is subsidiary to the bargaining problem.

Bold $\mathbf{a} = (a^1, a^2, ...)$ denotes an infinite sequence of action pairs of the stage game. A tail of the sequence is denoted by $\mathbf{a}_k = (a^k, a^{k+1}, ...)$. Let $b^1(\delta, \mathbf{0})$ denote the unique perfect equilibrium outcome of Rubinstein's alternating offers model when the set of outcomes is V^*

Proposition 1 of Busch and Wen (1995) establishes the solution to the alternating offers model when the sequence of threat point payoffs is nonstationary. This proposition is presented below in the context of Γ:

Alternating Offers Solution: *Modify the definition of* Γ *so players (1) are committed to playing* $\mathbf{a}$ *in the stage game before an acceptance, (2) can make offers in* V^**, and (3) are committed to play to obtain the offer, in expectation, in every stage after an offer is accepted. Then for all* δ *the modified* Γ *has a perfect equilibrium in which player 1's offer* $b^1(\delta, \mathbf{a}_2)$ *is accepted in the first stage, where*

$$b_1^1(\delta, \mathbf{a}_2) = b_1^1(\delta, \mathbf{0}) + (1-\delta)\sum_{j=0}^{\infty} \delta^{2j}[\delta u_1(a(2j+3)) - u_2(a(2j+2))],$$

and

$$b_2^1(\delta, \mathbf{a}_2) = b_2^1(\delta, \mathbf{0}) + (1-\delta)\sum_{j=0}^{\infty} \delta^{2j}[u_2(a(2j+2) - \delta u_1(a(2j+3))].$$

Any other perfect equilibrium is payoff equivalent. Subsequent offers $b^i(\delta, \mathbf{a}_k)$ *are defined analogously.*

Although the outcome path of every negotiation-compatible equilibrium has the property that no challenges occur we must show that there exist equilibria for which challenges are indeed not optimal. An *assumption* of existence of a negotiation-compatible equilibrium ϕ can be quite strong; it asserts that some set of offers can be enforced with agreements that no player will challenge. A *proof* of the existence of negotiation-compatible equilibria is important to establish that there are strategies that can enforce a set of offers for some discount factor.

Proposition 1 *If* $d^0 \in V^*$ *then there exists a* $\underline{\delta} \in (0,1)$ *such that for each discount factor* $\delta \in [\underline{\delta}, 1)$ *there exists a nearly efficient negotiation-compatible perfect equilibrium of* Γ *in which the initial agreement has payoff* d^0*.*

Proof. The negotiation-compatible equilibrium ϕ^* and the agreement $\sigma(d)$ for $d \in S(d^0)$ are defined as follows.

S1. (*payoff of the agreement)* The agreement $\sigma(d)$ yields the expected payoff d in every stage of the outcome path $\mathbf{a}(\sigma(d))$.

S2. (*proposals*) When neither player has deviated from the agreement $\sigma(d)$ the players alternate offers; player 1 makes the offer $b^1(\delta, \mathbf{d})$ and player 2 makes the offer $b^2(\delta, \mathbf{d})$, where

$b^1(\delta, \mathbf{d})$ and $b^2(\delta, \mathbf{d})$ are the payoff vectors that players 1 and 2 offer, respectively, in the solution to the alternating offers model. If d is Pareto efficient then the only offer that is relevant, when there are no deviations from $\sigma(d)$, is the payoff of the current agreement.

S3. (*decision rule of responders*) At each history which ends in an offer b by player j, and without a deviation from $\sigma(d)$, player i accepts the offer if $b_i \geqq b_i^j(\delta, \mathbf{d})$ and rejects the offer if $b_i < b_i^j(\delta, \mathbf{d})$.

S4. (*punishments*) Define $\overline{d}_1^0 := \max\{u_1 \geqq 0 : (u_1, d_2^0) \in V^*\}$. Since $(\overline{d}_1^0, d_2^0)$ is strictly individually rational define $(\overline{d}_1^0, d_2^0) + (\gamma_1, -\alpha_2)$ to be Pareto efficient in the strictly individually rational set $B((\overline{d}_1^0, d_2^0), r) \cap V^*$ for some $r, \alpha_i, \gamma_i > 0$, where $B(x, r)$ denotes an open ball around a point x with radius r. Define $(d_1^0, \overline{d}_2^0) + (-\alpha_1, \gamma_2)$ analogously.

The following two-phase punishment of player i, represented by the outcome path that results, $\mathbf{a}^j$,is used to show that there exists equilibrium agreements that can be used to punish the players. In the $L-$ length *punishment phase* of $\mathbf{a}^j$ player i chooses m_i^j to minmax player j, where $m^j \in A$ is an action profile that holds player j to his minmax payoff. In every period after the punishment phase, the players are in a *reward phase* and choose to obtain the expected stage game payoff $(d_1^0, \overline{d}_2^0) + (-\alpha_1, \gamma_2)$ if it is the reward phase of $\mathbf{a}^1$ and $(\overline{d}_1^0, d_2^0) + (\gamma_1, -\alpha_2)$ if it is the reward phase of $\mathbf{a}^2$.

The payoff to player i from $\mathbf{a}^j$ is

$$w_i^i := (1 - \delta^L) u_i(m^j) + \delta^L(\overline{d}_i^0 + \gamma_i),$$

and the payoff to player j from following $\mathbf{a}^j$ is

$$x_j^i := \delta^{L+1}(d_j^0 - \alpha_j).$$

Let (x_j^i, w_i^i) generically denote either (x_1^2, w_2^2) or (w_1^1, x_2^1).

S5. (*challenges*) Consider a history that ends in a challenge (u_i, u_j) by player j to an agreement $\sigma(d)$. Player i rejects the offer if $u_i \leqq w_i^i$.

This completes the definition of ϕ^* except for the parameter L, which is chosen together with the discount factor in the following incentive conditions. We will choose δ and L so that any offer accepted in $S(d^0)$ is an equilibrium offer in a negotiation-compatible equilibrium.

IC1 (*initial deviations).* First, we consider histories in which no player has deviated from the current agreement $\sigma(d)$. When there have not been any deviations the expected payoff to player i from following ϕ^* when player i is the responder is $(1-\delta)d_i+\delta b_i^j(\delta,\mathbf{d})$ and when player i is the proposer is $(1-\delta)d_i+\delta b_i^i(\delta,\mathbf{d})$. Since the latter expression equals $b_i^j(\delta,\mathbf{d})$ it suffices to consider the case when player i is the responder.

It must *not* be optimal for player i to deviate from $\sigma(d)$ and make an offer that player j rejects. The expected payoff of player i from challenging $\sigma(d)$ with an offer that player j rejects is bounded above by

$$R_i := (1-\delta)M + \delta^{L+1}(d_i^0 - \alpha_i)$$

where $M > 0$ is the largest payoff to either player in the stage game. A sufficient incentive constraint is

$$R_i \leqq (1-\delta)d_i + \delta b_i^j(\delta,\mathbf{d}) \tag{C1}$$

For a fixed L, when δ is larger than a threshold δ_1 we have that $R_i \in [d_i^0 - \alpha_i, d_i^0)$. Since $d_i^0 \leqq d_i \leqq b_i^j(\delta,\mathbf{d})$ condition (C1) will be satisfied.

Also, it must *not* be optimal for player i to deviate from $\sigma(d)$ and make an offer that player j accepts. Consider a history h that ends in a challenge by player i to $\sigma(d)$. What needs to be shown is that challenges by player i with a profitable own-demand are rejected by player j – there are no gains to trade. For it to not be optimal to challenge $\sigma(d)$ it must not be optimal for player i to deviate from $\sigma(d)$ and make an offer (u_i, u_j) such that

$$(1-\delta)u_i(\widehat{a}) + \delta u_i > U_i(\mathbf{a}(\phi_h))$$

and player j accepts (u_i, u_j), where $\widehat{a}$ is the action profile when player i deviates and player j does not, and ϕ_h is the strategy ϕ conditioned on the history h. By construction of the punishments, after a deviation by player i, player j can guarantee w_j^j by rejecting challenges by player i. For a challenge (u_i, u_j) by player i:

$$(1-\delta)u_i(\widehat{a}) + \delta u_i > (1-\delta)d_i + b_i^j(\delta,\mathbf{d}) \Longrightarrow u_j \leqq w_j^j.$$

We need to state a sufficient incentive constraint. Given a δ, define $\underline{u}_i^i$ to be the number that

satisfies the equality

$$(1-\delta)M + \delta \underline{u}_i^i = (1-\delta)d_i + \delta b_i^j(\delta, \mathbf{d}).$$

For player 2, $\underline{u}_2^2$ is a lower bound on profitable own-demands which puts player 2 on the margin between negotiating under the current agreement and challenging the current agreement. It follows that the number

$$\underline{u}_1^2 := \max\{u_1 \geqq 0 : (u_1, \underline{u}_2^2) \in V^*\}$$

is the *upper* bound on what player 2 could offer player 1 in a challenge with a profitable own-demand. The sufficient incentive constraint is

$$\underline{u}_j^i \leqq w_j^j. \tag{C2}$$

Since $\underline{u}^i$ is Pareto efficient and $\lim_{\delta \nearrow 1} \underline{u}_i^i(\delta) = b_i^j(1, \mathbf{d})$, we also have that $\lim_{\delta \nearrow 1} \underline{u}_j^i(\delta) = b_j^j(1, \mathbf{d})$. The difference

$$w_j^j - b_j^j(\delta, \mathbf{d})$$

is positive and increases in the discount factor when punishment is costly relative to $\overline{d}_j^0 + \gamma_j$. Moreover, $b_j^j(d, \delta) \leqq \overline{d}_j^0$. Thus, condition (C2) holds for all discount factors above a threshold δ_2. When punishment is not costly relative to the reward, then $\overline{d}_j^0 + \gamma_j$ is a lower bound for w_j^j and permits an analogous argument.

IC2. (*punishments*) First, we establish a condition on the discount factor for the set of paths $\Pi = \{\ \mathbf{a}^1, \mathbf{a}^2\}$ to be a perfect equilibrium in the underlying repeated game, not Γ. This construction is used to prove that there exists an equilibrium offer in the negotiation-compatible equilibrium which has a payoff in $S(x_j^i, w_i^i)$. These particular offers are continuation equilbria which are used as punishments.

When the punishment $\mathbf{a}^j$ is used and player j deviates from it the continuation payoff of player i is always w_i^i. Moreover, when punishing is costly to player i relative to $\overline{d}_i^0 + \gamma_i$, the continuation payoff of player i at any point of the sequence $\mathbf{a}^j$ is at least as large as w_i^i.

If player j deviates from $\mathbf{a}^j$ the punishment is restarted. This deviation is not optimal

during the punishment phase if

$$\delta^{L+1}(d_j^0 - \alpha_j) \leqq \delta^s(d_j^0 - \alpha_j),$$

where $s \leqq L_1$. During the reward phase this deviation is not optimal if

$$(1-\delta)M + \delta^{L+1}(d_j^0 - \alpha_j) \leqq (d_j^0 - \alpha_j),$$

or,

$$M < \frac{1-\delta^{L+1}}{1-\delta}(d_j^0 - \alpha_j). \tag{C3}$$

Since

$$\lim_{\delta \nearrow 1} \frac{1-\delta^{L+1}}{1-\delta} = L$$

there exists an L and δ_3 such that (C3) is satisfied for $\delta \in [\delta_3, 1)$.

Now we argue, given $\underline{\delta} = \max\{\delta_1, \delta_2, \delta_3\}$ and L, that if the payoff (x_j^i, w_i^i) is not an equilibrium offer in a negotiation-compatible equilibrium then the perfect equilibrium punishment we constructed which yields (x_j^i, w_i^i) necessarily permits a challenge

$$d' \in S^*(x_j^i, w_i^i) := S(x_j^i, w_i^i) - \{(x_j^i, w_i^i)\}$$

that is an equilibrium offer in a negotiation-compatible equilibrium.

A lower bound on the challenger's payoff to deviating from Π is given by the payoff obtained from deviating and having the challenge be rejected:

$$(1-\delta)u_j(\widehat{a}) + \delta^{L+1}(d_j^0 - \alpha_j).$$

Since player i can guarantee w_i^i by rejecting the offer, player j can guarantee x_j^i by not deviating in the first place, and the perfect equilibrium Π can be challenged, this means the accepted challenge is an element of $S^*(x_j^i, w_i^i)$.

Thus, if the payoff (x_j^i, w_i^i) is an equilibrium offer in a negotiation-compatible equilibrium then the punishment for player j is an equilibrium agreement that yields (x_j^i, w_i^i). Otherwise, if (x_j^i, w_i^i) is not an equilibrium offer in a negotiation-compatible equilibrium then the punishment

for player j is an equilibrium agreement that yields $d' \in S^*(x_j^i, w_i^i)$.

To summarize, choose $\underline{\delta} = \max\{\delta_1, \delta_2, \delta_3\}$, and L according to IC1 and IC2. This establishes that any continuation payoff d in the set $S(d^0)$ can be enforced by an agreement $\sigma(d)$ which no player will challenge. The alternating offers bargaining problem from any of these points is well-defined and has a unique solution given by the formulas for $b^1(\delta, \mathbf{d})$ and $b^2(\delta, \mathbf{d})$. ■

A corollary of Proposition 1 is that there exist strategies in a standard repeated game such that the set of perfect equilibrium payoffs, $PEP(\delta)$, contains a nontrivial convex set when the discount factor is sufficiently high. This claim is not implied by the folk theorem, which states that every strictly individually rational payoff vector is a member of $PEP(\delta)$ when the discount factor is sufficiently high. It suffices to use *simple strategy profiles* (Abreu, 1988) which, for two players, are completely defined by an initial path $\mathbf{a}^0$ and punishment paths $\mathbf{a}^1$ and $\mathbf{a}^2$.

Corollary 1 *If the payoff $d^0 \in V^*$ is strictly individually rational then there exists a $\underline{\delta} \in (0,1)$ such that for each discount factor $\delta \in [\underline{\delta}, 1)$ there exists a subgame perfect equilibrium in simple strategies of the repeated game and $S(d^0) \subset PEP(\delta)$.*

Proof. The simple strategy is derivative of the strategy in Proposition 1. For $x \in S(d^0)$ the players choose to obtain the expected payoff x in each period on the initial path $\mathbf{a}^0$, using correlated strategies in the stage game if necessary. The punishment path for player 1, $\mathbf{a}^1$, has a punishment phase in which an action profile m^1 that holds player 1 to zero is chosen for L periods and a reward phase in which players choose to obtain the Pareto efficient expected payoff $(d_1^0, \overline{d}_2^0) + (-\alpha_1, \gamma_2)$ thereafter. The punishment path for player 2, $\mathbf{a}^2$, is defined analogously. There are four incentive constraints to verify for each player. Only one of these depends on the initial payoff vector x and is relevant for establishing a uniform incentive constraint for all $x \in S(d^0)$. For any $x \in S(d^0)$ this constraint is

$$(1-\delta)M + \delta^{L+1}(d_i^0 - \alpha_i) \leqq x_i$$

The number d_i^0 is the minimal value of the ith coordinate over all vectors in $S(d^0)$. To establish

the uniformity condition observe that the condition

$$(1-\delta)M + \underline{\delta}^{L+1}(d_i^0 - \alpha_i) \leqq d_i^0$$

implies the former condition for $\underline{\delta}$ chosen to satisfy the latter condition. ■

Example. *Cournot duopoly.* The stage game is a static Cournot model with inverse demand function $p = 2 - x$, quantities $x_i \in [0, 2]$, and marginal cost equal to zero.

The players will have an initial agreement $\sigma^0(u^0)$. For illustration, assume each payoff on the outcome path of $\sigma^0(u^0)$ is $u^0 = (4/9, 4/9)$, the payoff of the Cournot-Nash equilibrium action pair $a = (2/3, 2/3)$.

The players negotiate with each other while they are choosing levels of output to attempt to compromise on a continuation payoff. An offer $u \in S(u^0)$ that differs from u^0 is a proposal for how to split the gains to restricting output below the competitive level.

Assume that the discount factor satisfies the condition in Proposition 1 so that all payoffs $S(u^0) = \{x \in V^* : x_i \geqq 4/9\}$ can be enforced in a negotiation-compatible equilibrium.

To illustrate the distinction between a perfect equilibrium of the repeated game and an agreement followed by the players in a negotiation-compatible equilibrium suppose that Cournot-Nash reversion is the initial agreement used to obtain $(1/2, 1/2)$, the point of $S(u^0)$ that splits the monopoly output. Now suppose that player 1 challenges the agreement by choosing a best-response in the stage game and the offer $1/2$. If player 2 were to accept this offer the expected payoff of player 1 would be

$$(1-\delta)\frac{9}{16} + \delta\frac{1}{2},$$

which exceeds the payoff to not challenging, $1/2$. The decision player 2 faces is whether to accept the offer. Accepting yields the payoff $1/2$. If player 2 rejects the offer, and both players follow the current agreement in the next subgame, then player 2 would obtain the Cournot-Nash payoff for one period followed by the bargaining outcome:

$$(1-\delta)\frac{4}{9} + \delta\frac{5+4\delta}{9(1+\delta)}.$$

Since the latter expression is less than 1/2 player 2 would accept the challenge of player 1. We conclude that Cournot-Nash reversion cannot be a challenge-free agreement.

We also note that an agreement that always consists of play of the static Cournot-Nash equilibrium would not be part of a negotiation-compatible equilibrium when $S(u^0)$ is enforced in a negotiation-compatible equilibrium. Since a deviation would have no payoff consequence for either player the players effective bargaining positions do not change. However, the bargaining protocol permits the challenger to make an offer after a deviation. Without any punishment from the opponent a player would profit from using a deviation to become the proposer.

In the remainder of the example we characterize the set of negotiation-compatible equilibrium payoffs when the set of offers $S(u^0)$ is enforceable in a negotiation-compatible equilibrium. Observe that the payoff u^0 is effectively a threat point – when an offer is refused the players will expect play to follow $\sigma^0(u^0)$. Thus, in any negotiation-compatible equilibrium the offers will be identical to the solution of the alternating offers model when the threat point is u^0. From the alternating offers solution, Player 1 always offers

$$b^1(\delta, u^0) = \left(\frac{5+4\delta}{9(1+\delta)}, \frac{5\delta+4}{9(1+\delta)}\right)$$

and accepts any offer b with

$$b_1 \geqq \frac{5\delta+4}{9(1+\delta)}.$$

Player 2 always offers

$$b^2(\delta, u^0) = \left(\frac{5\delta+4}{9(1+\delta)}, \frac{5+4\delta}{9(1+\delta)}\right)$$

and accepts any offer b with

$$b_2 \geqq \frac{5\delta+4}{9(1+\delta)}.$$

In any negotiation-compatible equilibrium with an initial agreement $\sigma^0(u^0)$ player 1 offers the Pareto efficient vector $b^1(\delta, u^0)$, player 2 accepts it, and it is subsequently enforced by an agreement $\sigma(b^1(u^0, \delta))$.

The payoff to player 1 in the negotiation-compatible equilibrium is

$$(1-\delta)\frac{4}{9} + \delta\frac{5+4\delta}{9(1+\delta)},$$

and the payoff to player 2 is

$$(1-\delta)\frac{4}{9}+\delta\frac{5\delta+4}{9(1+\delta)}.$$

Since only the first stage payoff is not Pareto efficient, the payoff is nearly efficient when the discount factor is high. Also, observe that in any subgame that ends in the acceptance of some offer $u \in S(u^0)$ the bargaining problem is identical to the bargaining problem at the initial agreement except that now the threat point is u and the effective set of offers is $S(u) \subset S(u^0)$. Thus, in any negotiation-compatible equilibrium whenever a payoff in $S(u^0)$ that is not Pareto efficient is accepted the players will subsequently negotiate to a payoff vector that is Pareto efficient and remain there. This means that in any negotiation-compatible equilibrium the only set of payoffs in $S(d^0)$ that players will not negotiate away from is the Pareto frontier.

Proposition 2 *If $d^0 \in V^*$ and the discount factor is sufficiently large then all negotiation-compatible equilibria in which the initial agreement has payoff d^0 are nearly efficient.*

Proof. Several key facts are used. Proposition 1 showed that all payoffs in $S(d^0)$ can be enforced in a negotiation-compatible equilibrium when the discount factor is sufficiently large. In any negotiation-compatible equilibrium it is not optimal to challenge the agreements, and once there are no challenges there is a well-defined bargaining problem that has a unique prediction under the alternating offers bargaining protocol.

Consider the bargaining problem $\langle S(d), \mathbf{a}\rangle$ that the players have at an initial agreement, where $\mathbf{a}$ is a sequence of action pairs that yields the payoff d^0 or some other payoff in $S(d^0)$. The assumptions on preferences imply that each player's preferences on $S(d^0)$ are represented by $\delta^t u_i$ for $u \in S(d^0)$. The alternating offers solution provides an efficient payoff vector $b^i(\delta, \mathbf{a})$ that solves the bargaining problem. In every perfect equilibrium of the bargaining problem the first proposer, player i, offers $b^i(\delta, \mathbf{a})$ in the first stage of the negotiations and it is accepted.

We want to argue that the alternating offers solution to the bargaining problem $\langle S(d), d\rangle$ is equivalent to the alternating offers solution to the bargaining problem $\left\langle S(d^0), d\right\rangle$. Since $S(d^0) \cap S(d) = S(d)$ and each player can guarantee himself d_i there does not exist a perfect equilibrium alternating offers solution to the bargaining problem $\left\langle S(d^0), d\right\rangle$ that lies in the set $S(d^0) - S(d)$. At the same time there is a unique solution in the set $S(d)$. Thus, any perfect equilibrium alternating offers solution to $\langle S(d), \mathbf{a}\rangle$ yields the payoff $b^i(\delta, \mathbf{a})$. ■

Proposition 3 demonstrates that one of the consequences of incorporating bargaining into a repeated game framework is that the offer of a challenger serves as the threat that effectively enforces the challenger's own punishment. Restrictions on the payoff of this threat to the challenger are derived.

Consider a history of Γ that ends in a challenge by player j to an agreement which has a payoff vector $d \in S(d^0)$. In particular, consider the challenge (u_i^i, u_j^i) by player j in which u_i^i equals the continuation payoff under ϕ player i receives when she rejects this challenge and u_j^i is defined to be the maximum own-demand of player j given u_i^i. Let $\mathbf{a}$ be a sequence of action profiles that yields the payoff d.

Proposition 3 *If $d^0 \in V^*$ and the discount factor is sufficiently large then in all negotiation-compatible equilibria in which the initial agreement has payoff d^0 we have*

a. (No-Trade Condition) $u_i^i \geqq b_i^i(\delta, \mathbf{a})$, $u_j^i \leqq b_j^i(\delta, \mathbf{a})$.

b. (Concession Principle) The largest own-demand of player j in a challenge which is accepted by the defender is bounded above by $b_j^i(\delta, \mathbf{a}) - \epsilon(\delta)$ where $\epsilon(\delta) \searrow 0$ as $\delta \nearrow 1$.

Proof. Consider the case when player 1 is the challenger. The case when player 2 is the challenger is analogous. The strategy of player 2 must have a response to all histories of Γ in which player 1 has challenged with some offer (u_1, u_2). For it to be optimal for player 2 to reject the offer it must be that u_2 is less than or equal to player 2's continuation payoff from choosing reject: $u_2 \leqq u_2^2$. Since the strategy is a negotiation-compatible equilibrium and an offer $u_2 = u_2^2$ would be the threshold above which offers are accepted by player 2 the corresponding maximum demand u_1^2 defined above is worse for player 1 than choosing to bargain: $u_1^2 \leqq b_1^2(\delta, \mathbf{a})$. This is true if $u_2^2 \geqq b_2^2(\delta, \mathbf{a})$.

It is possible to be more precise about the relation of the own-demand, u_1, in a challenge (u_1, u_2) which player 2 accepts, and the own-demand, $b_1^1(\delta, \mathbf{a})$, in the offer player 1 makes under the alternating offers protocol. Denote by $\underline{u}_1$ the lower bound on any own-demand of the challenger in which choosing to challenge is at least as profitable as choosing to bargain

$$(1-\delta)u_1(\widehat{a}) + \delta\underline{u}_1 \geqq (1-\delta)d_1 + \delta b_1^2(\delta, \mathbf{a}), \text{ or}$$

$$\underline{u}_1 \geqq b_1^2(\delta, \mathbf{a}) - (1-\delta)(u_1(\widehat{a}) - d_1)/\delta,$$

where $\widehat{a}$ is the action profile when the challenger deviates. Since it is not optimal to choose to challenge in a negotiation-compatible equilibrium it must be that all offers (u_1, u_2) which player 2 accepts have $u_1 \leqq \underline{u}_1$. Moreover, since

$$\lim_{\delta \nearrow 1} (b_1^2(\delta, \mathbf{a}) - (1-\delta)(u_1(\widehat{a}_1, a_2) - d_1)/\delta) = b_1^2(\delta, \mathbf{a})$$

it can be said that for δ near 1 that the largest own-demand of player 1 in a challenge that player 2 accepts is approximately bounded above by the expected bargaining settlement, with the discrepancy accounted for by the short run gains to cheating in the initial agreement. Conversely, the smallest profitable own-demand of player 1 in a challenge that player 2 rejects is approximately bounded below by the expected bargaining settlement. ■

The no trade condition in Proposition 3 means that both players anticipate that any offer that follows a deviation from the current agreement will be rejected, thereby deterring any challenges. By choosing to reject the challenge, the defender has effectively chosen to delay any negotiations until after he has punished the challenger. This choice defends the original bargaining position of the defender, while accepting the challenge would result in an inferior bargaining position. As a consequence, the only gains to trade that the players anticipate occur at the bargaining table – no one breaks the current agreement.

How should an individual who violates an agreement respond when the opponent fails to deter the violation? The prediction of the model is that the challenger will demand more in the future. For, if the opponent fails to punish the challenger he has effectively failed to defend his bargaining position and weakened his opposition to the challenger. The *concession principle*, as stated in Proposition 3, implies that were a player to challenge the current agreement with a profitable own-demand then the challenger's payoff is at least as high as the payoff were he never to have challenged the current agreement. We might even state this prediction as a principle for how to act in the event of a concession: *If an individual has made a concession then the other individual should respond with an attempt to increase his demand.* Since the offer of the challenger is also the subsequent threat of the challenger that effectively enforces the challenger's own punishment this means that the threat of the challenger, when chosen optimally, is constrained by the continuation payoff of the original strategy. Hence, the players

have two distinct types of threats they can make depending on whether they are in the situation of a challenger or a defender. One threat is a punishment path, as in the folk theorem of Fudenberg and Maskin: "If you challenge the agreement then I will hold you down." The other threat is a threat to demand more for themselves at their opponent's expense: "If you don't hold me down then I will increase my demands." In particular, the challenger does not threaten to punish the defender for conceding – instead the challenger threatens to demand more.

One of the implications of Proposition 3 is that there exists a distinction between credible threats of an individual in the position of a *challenger* and credible threats of an individual in the position of a *defender*. It is credible for players to use costly punishment when in the position of a defender but not when in the position of a challenger. Firms who are cheated on are willing to retaliate, because the consequence of not retaliating is that the challenger will continue to demand more at their expense. Both the challenger and the defender fully anticipate the new demands which are completely in accord with the incentives of the challenger. Individuals acting under their natural incentives in the face of a concession will demand more from an opponent who has relaxed his resistance – the unintended consequence of this is that they enforce their own punishments. This explanation of the existence of costly punishment stands in sharp contrast to theory in repeated games which predicts that firms willfully enforce their own punishments by holding down their opponents and themselves – a distortion of the meaning of incentives justified by subgame perfection. This explanation of costly punishment also stands in contrast to the argument provided by Farrell and Maskin to support their claim that economic actors will not use costly punishment. Farrell and Maskin state that economists should be skeptical of subgame perfect equilibria in which punishments hurt both players. Instead, they argue, the players would strike a deal that is mutually preferred. However, this argument that costly punishment is an incredible threat and, by implication, not a good prediction of economic behavior is not confirmed by empirical evidence. Costly retaliation is a defining feature of most observed situations of long-term competition.

3.4 Conclusion

A model of long-term competition is constructed in which a repeated game is played while negotiations are conducted over continuation payoffs of the repeated game. The solution concept is negotiation-compatible equilibria, a subset of the set of perfect equilibria of the repeated game in which the players' negotiations are meaningful and not just cheap talk. Theorem 1, the main proposition, states that when the discount factor is sufficiently high negotiation-compatible equilibria exist, all negotiation-compatible equilibria are nearly efficient, and after one stage play in the repeated game is efficient.

3.5 References

Abreu, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56, 383-396.

Abreu, D., and D. Pearce (2000): "Bargaining, Reputation, and Equilibrium Selection in Repeated Games," mimeo.

Abreu, D., Pearce, D., and E. Stacchetti. (1993): "Renegotiation and Symmetry in Repeated Games," *Journal of Economic Theory*, 60, 217-240.

Benoit, J., and Krishna, V. (1993): "Renegotiation in Finitely Repeated Games," *Econometrica*, 61, 303-323.

Bernheim, D., and D. Ray (1989): "Collective Dynamic Consistency in Repeated Games," *Games and Economic Behavior*, 1, 295-326.

Busch, L., and Q. Wen (1995): "Perfect Equilibria in a Negotiation Model," *Econometrica*, 63, 545-565.

Datta, Saikat (1994): "Existence of Internally Renegotiation Proof Sets in Infinitely Repeated Games," Ph.D. Thesis, Indian Statistical Institute.

Farrell, J., and E. Maskin (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1, 327-360.

Fudenberg, D., and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54, 533-554.

Osborne, M., and A. Rubinstein (1994): *A Course in Game Theory.* Cambridge, Mass.:

The MIT Press.

Ray, D. (1994): "Internally Renegotiation-Proof Equilibrium Sets: Limit Behavior with Low Discounting," *Games and Economic Behavior*, 6, 162-177.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 128-140.

Vita
Eliot Maenner

Date & Place of Birth: February 5, 1970; Wisconsin

Citizenship: USA

Education: Ph.D. in Economics, Pennsylvania State University, 2002

B.A. (with distinction), Economics and Political Science, University of Wisconsin Madison, 1992

Publications: "Convex Potentials with an Application to Mechanism Design," (with Vijay Krishna), *Econometrica*, 69(4), 2001, pp. 1113 -1119

Experience: Instructor, Intermediate Microeconomics, 2000-2001

Teaching Assistant, Microeconomic Theory (graduate), Fall 1999, Fall 2001

Referee for Econometrica

Research Assistant, Board of Governors of the Federal Reserve System, 1992-1994