I wish to thank Kalyan Chatterjee and Vijay Krishna for numerous suggestions that improved the model and exposition. I am also grateful for receiving helpful comments from Drew Fudenberg, Jim Jordan, Dmitri Kvassov, Eric Maskin, Ariel Rubinstein, Hamid Sabourian, and Tomas Sjöström.
Abstract

A learning model is constructed for two-player repeated games. In an inference step players construct minimally complex models of strategies based on observed play, and in an adaptation step players choose minimally complex best responses to an inference. When both players have optimistic inferences, stage game Nash equilibria are played in all steady states in the set of finite state automata, while under cautious inferences the set of steady states is the subset of self-confirming equilibria with Nash outcome paths. Convergence to these steady states is demonstrated in the Prisoner’s Dilemma when the initial conditions have at most two states.

*Journal of Economic Literature* Classification Numbers: C72, D83.

*Key Words:* repeated games, learning, bounded rationality.
1 Introduction

The theory of repeated games is central to models of long-term competition used in economic theory. The main benchmark results in the theory of repeated games are the folk theorems which demonstrate that 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, [10])\(^\dagger\). This multiplicity of equilibria has led some researchers, both theoretical and applied, to restrict attention to certain equilibria. Sometimes these restrictions are imposed directly on the set of equilibrium payoffs and justified by an implicit bargaining argument in which efficient equilibria are focal. Sometimes there are restrictions to certain simple strategies, such as the grim trigger strategy, which are justified by an implicit complexity argument that players have a predilection for simple strategies which in some sense have a minimal cost to implement. Moreover, these types of equilibrium selection arguments are not restricted to repeated games; such restrictions are the norm rather than the exception in dynamic models in economics. Most analyses of dynamic macro models, search models, and random matching models restrict attention to equilibria in which agents use stationary (history-independent) strategies. This is also true of many models of dynamic games including stochastic games and multi-person bargaining. The challenge is to provide equilibrium selection arguments based on modeling principles.

There have been several approaches to equilibrium selection in repeated games that have provided both sharper predictions and explanations for when players might coordinate on efficient or simple equilibria. As the folk theorem obtains partially because players can use potentially complex and history dependent strategies, Abreu and Rubinstein [3] suggested that complexity considerations could address the multiplicity problem. They model players who choose strategies represented by finite state automata and have a preference for simple strategies. Although they obtain that a set of equilibrium payoffs that is a strict subset relative to what is obtained by the folk theorem, there still is not an obvious prediction. Farrell and Maskin [7] adopt a bargaining approach in which they assume a postulate from bargaining theory to refine the set of subgame perfect equilibria in infinitely repeated games. Yet, even though the assumed postulate is that the outcome of a bargaining problem is efficient there is no obvious prediction, nor are the predictions necessarily Pareto efficient. Binmore and Samuelson [4] adopt an evolutionary approach in which they define a version of evolutionary stability for repeated games. They obtain, under limit of the means payoffs, that the symmetric efficient payoff pair in the repeated Prisoners’ Dilemma is the unique equilibrium payoff\(^\ddagger\). Maskin and Tirole [16] hypothesize that players

\(^\dagger\)The full dimensionality condition is a sufficient condition for the folk theorem to hold when there are three or more players.

\(^\ddagger\)See Samuelson and Swinkels [23] for a discussion of the sensitivity of this result to the form of evolutionary stability applied to games with lexicographic preferences.
who have complexity considerations could coordinate on simple strategies 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.

Other research in dynamic games has also impacted equilibrium selection arguments in infinitely repeated games. The development of equilibrium concepts such as self-confirming equilibrium (Fudenberg and Levine, [9]) or equilibrium in justifiable strategies (Spiegler, [24]) has helped to refine the viewpoint that Nash equilibria are the only sensible solutions, yet they have the drawback of increasing the multiplicity of solutions in dynamic games. Further development of methods to measure the complexity of strategies has been used to address the problem of selecting stationary strategies in several classes of dynamic games. In Chatterjee and Sabourian [6] players in an \( n \)-person unanimity bargaining game, with preferences over the degree of complexity of the strategies, only choose stationary strategies in equilibrium. Under the same preferences, Sabourian [22] is able to reduce the set of equilibria in a dynamic matching and bargaining game to only stationary strategies, all of which induce the competitive price.

Jéhie[12] investigates how boundedly rational players might play repeated games and why such players may be led to play simple strategies. Players with restricted hindsight and foresight construct a model or forecast of each other and in an \((n_1, n_2)\) - equilibrium players choose optimally given their forecasts, and the forecasts are correct. A subset of the \((n_1, n_2)\) - solutions, the hyperstable solutions, are shown to have a certain simple structure, although they differ from stationary strategies. Jéhiel [14] further demonstrates equilibrium behavior, in the context of \((n_1, n_2)\) - solutions, which precludes play of the unique stationary equilibrium of the repeated Prisoners’ Dilemma at the same time it does not rule out play of simple nonstationary strategies that attain efficiency.

We adopt a learning approach to equilibrium selection in infinitely repeated games in which the adaptation process serves as the equilibrium selection mechanism. We model boundedly rational players who can begin the game with any feasible strategies and who proceed through recurrent stages where they observe realized play, construct a certain class of models of each other’s strategies based on observed play, and choose new strategies based on the models that they constructed. The adaptation process is a dynamical system that generates infinite sequences of repeated game strategy profiles. The steady states and convergence properties of these sequences are studied.

It is demonstrated that the dynamics of the adaptation process depend crucially on the inference rules the players use to select new strategies. The dynamics when both players make optimistic inferences and when both players make cautious inferences are considered. The steady state results hold for any two-player finite action stage

---

3 This learning model does not appear in the published version of their paper.

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 of a game.
when the set of strategies corresponds to the set of finite state automata. When both players make optimistic inferences we find (Theorem 1) that in any steady state of the adaptation process that Nash equilibria of the stage game must be played in each period. In the main example, the infinitely repeated Prisoners’ Dilemma, this implies that the unique steady state of the adaptation process consists of both players playing Defect in each period. In contrast, when both players make cautious inferences the set of steady states is the subset of self-confirming equilibria with Nash outcome paths (Theorems 2 and 3). It is also demonstrated that the set of steady states that can be reached will, in general, depend on the initial strategies of the players. Finally, we demonstrate convergence to the set of steady states from any initial condition in the set of one and two state automata when the stage game is the Prisoners’ Dilemma (Theorems 4 and 5).

The adaptive process is a 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 infinitely repeated game—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, when the stage game is the Prisoners’ Dilemma, a particular pair of strategies may lead to the following outcome path

\[(C, D), (D, C), (C, D), (D, C), \ldots\]

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. They construct models of each other’s unobserved behavior based on observed behavior. 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 constructed an inference about 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 strategy leads to a new outcome path and the entire process is repeated thereby generating a sequence of strategy profiles.

In the adaptation process simplicity considerations constrain both the strategies that players choose and the inferences they form about each other’s strategies. First, in choosing which strategy to adopt, a player selects one from among those that maximize his payoff given his inference, but a player prefers simple decision rules to more complicated ones. Clearly there cannot be a single compelling way to measure the simplicity or complexity of a strategy. Following a suggestion of Aumann [1], it is assumed that players only choose strategies that can be implemented by finite state
This allows the simplicity/complexity of a strategy to be measured in a natural and straightforward manner: an automaton with fewer states than another is considered to be simpler. This is motivated by the idea that strategies with additional complexity impose additional costs on the players. A main advantage of representing strategies as automata in this model, besides providing a natural way to measure the complexity of a strategy, is that their mathematical properties facilitate a theoretical study of dynamics in a nontrivial strategy space, without having to resort to numerical simulations.

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. This involves the point, also made by Spiegler [24], that the assumption that players would ever put positive probability on all possible beliefs, or even a large number of them, may be implausible. Mere consistency is a rather coarse theory of belief formation, and, as in Spiegler, we are led to consider players who use Occam’s Razor. A natural justification for using Occam’s Razor is that constructing models is costly in terms of time and labor, rather than any virtue of simple explanations per se. How is information analyzed under such resource constraints? In particular, players may only have the resources to build a small number of models of the opponent. It is plausible that the first models they will construct are the simplest models that explain the facts.

Since players’ strategies are restricted to automata it is natural to suppose that when making inferences, that the models players construct of each other are also automata. Out of the infinitely many possible automata that could explain the observed play the simplest possible inferences are those with a minimal number of states. There may be more than one simplest automaton that explains the facts, and how players arrive at a particular inference from this set matters. We 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 define different dynamical systems and we study properties of both.

## 2 The Model

### 2.1 Basic Definitions

The **Underlying Repeated Game** The game played in each period is a two player, finite action, normal form game. The action set is denoted $A$ and the payoffs are given by $u_i : A^2 \rightarrow \mathbb{R}$, $i = 1, 2$. The payoff matrix for the main example, the Prisoners’ Dilemma, is depicted in Figure 1.

$$
\begin{array}{c|cc}
 & C & D \\
\hline
C & 1, 1 & -l, 1 + g \\
D & 1 + g, -l & 0, 0 \\
\end{array}
$$

Figure 1: The Prisoners’ Dilemma $(g, l > 0)$

It is assumed that players choose strategies that can be implemented by finite state automata. An automaton is a 5-tuple $M_i = \{A_i, A_j, Q_i, f_i, \tau_i\}$, where $A_i$ and $A_j$ are finite action sets, $Q_i$ is a finite set of states, $f_i : Q_i \rightarrow A_i$ is an output function, and $\tau_i : Q_i \times A_j \rightarrow Q_i$ is a transition function that maps states and an action of the other player into a state. One of the states, denoted $q^1_1$, is a given initial state. The set of finite state automata is denoted $\mathcal{M}$. 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 or the identity of a state in an automaton (the subscript for the player is often omitted when speaking of a particular automaton). Every pair of automata $(M_1, M_2)$ generates a sequence of states: $\{q^1, q^2, q^3, \ldots\}$. The finiteness of both the action sets and sets of states implies that this sequence must eventually cycle after an introductory phase (possible empty):

$$\{q^1, \ldots, q^{n-1}, q^n, \ldots, q^2, q^1, \ldots, q^2, \ldots\}$$

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

---

5 When the players have the same action sets the set of inputs into the transition function and the set of outputs are identical.

6 However, see Kalai and Stanford [15] for an alternative definition of an automaton representation of a rule of behavior.
The *state table* representation of a finite state automaton displays the 5-tuple and the initial state in a table (Figure 2). These tables are useful for computations. The initial state is the first state listed in the current state column.

<table>
<thead>
<tr>
<th>Current state</th>
<th>1st Input = C</th>
<th>2nd Input = D</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_1 )</td>
<td>( \tau_i(q_1, C) = q_{i1} )</td>
<td>( \tau_i(q_1, D) = q_{i2} )</td>
<td>( f_i(q_1) = C )</td>
</tr>
<tr>
<td>( q_2 )</td>
<td>( \tau_i(q_2, C) = q_{i1} )</td>
<td>( \tau_i(q_2, D) = q_{i2} )</td>
<td>( f_i(q_2) = D )</td>
</tr>
</tbody>
</table>

Figure 2: A State Table Representation of Tit-For-Tat.

Besides the restriction to choosing a strategy in the set of automata, players also have an explicit preference for simplicity in the set of automata. Following Abreu and Rubinstein [3], complexity is measured by the number of states in the automaton. Provided two automata attain the same payoff, a player prefers 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.

The payoff in the infinitely repeated game, 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:

\[
(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(f(q^t)).
\]

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 \( M'_i \) is said to be a *best response* to an automaton \( M_2 \) if \( (M'_i, M_2) \succ_1 (M, M_2) \) 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.

**The Adaptation Process** We now describe the sequence of events that defines the adaptation process. The players are adaptive agents and allowed to switch their strategies in the course of play. A *decision period* is the frequency \( T \) at which players are permitted to switch their strategies. Although it would be natural to assume that \( T \) is a finite integer, in this model players make inferences based on the outcome path and it is convenient to assume \( T = \infty \).

---

7 This measure of complexity is coarser than the measure in Chatterjee and Sabourian [6]. Also see Kalai and Stanford [15].

8 An introduction to repeated games with automata appears in Osborne and Rubinstein [17].

9 Decision periods in which \( T = \infty \) often appear in evolutionary models (where they are known as *epochs* or *generations*). Yet, the present model is classified as a learning model. The players make explicit choices and there is no population dynamics to drive the results. Rationality has been replaced by bounded rationality, but not by natural selection.
period is the infinitely repeated Prisoners’ Dilemma. 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\(^{10}\). Decision periods are indexed by \(s = 1, 2, \ldots\) to distinguish them from ordinary periods.

At the start of the game play follows the path given by some initial pair of finite state automata. To impose some structure on the adaptation process, we assume that player 1 can switch strategies only after odd numbered decision periods \((s = 1, 3, \ldots)\) and player 2 can switch strategies only after even numbered decision periods \((s = 2, 4, \ldots)\). At the beginning of each decision period \((s > 1)\) players observe the outcome path of the previous decision period and use it to construct inferences of each other. A player’s inferences about the opponent are automata which could have generated the sequence of observed actions, given the player’s own strategy. All inferences except those that have a common minimal number of states are ignored by the players. The player who can switch strategies in the \(s\)-th decision period (generically referred to as player \(i\) hereafter) has a rule, described below, to infer exactly one of these models is the true strategy of player \(j\). At the beginning of the \(s\)-th decision period \((s > 1)\) player \(i\) chooses a strategy, denoted \(M^s_i\), that is a best response to the inferred model. Player \(j\) continues to use the same strategy as he used in the \((s - 1)\)-th decision period. Then the process repeats – observe, infer, choose.

It is assumed that when players observe the outcome path and construct models of each other they do not observe each other’s strategies. Moreover, the only information that the players can use in their decision problem is the information contained in the outcome path of the current decision period. Information from prior decision periods is not factored into the players’ decision problems. After the models are constructed all other data in the history gets discarded.

The set of automata which could be the inferences about a player is infinite (for example, adding redundant states to an inference yields a new 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. 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 [5], 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.

\(^{10}\)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.
For any outcome path $\pi(M_1, M_2)$ generated by a pair of finite state automata, define $I_1$ to be the set of all minimal state automata with $C_1$ states such that \((i)\) if $M \in I_1$, 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 $I_2$ is defined similarly. The sets $I_1$ and $I_2$ 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 $I_1$ and $I_2$ are not empty, unique, finite, and $\pi(M_1', M_2') = \pi(M_1, M_2)$ for all $(M_1', M_2') \in I_1 \times I_2$. For player $i$ this relation is symbolized by the multi-valued mapping: $\psi_i : \mathcal{M}^2 \Rightarrow \mathcal{M}$ where $\psi_i(M_1, M_2) = I_j$.

**Example 1.** Let $M_1 = M_2 = TIT – FOR – TAT$. Then $\pi(M_1, M_2) = \{(C, C), \ldots\}$ and, $I_1 = I_2 = \{\text{COOPERATE}\}$, where COOPERATE denotes the one-state automaton that always plays the action $C$. Thus, even though $TIT – FOR – TAT$ is a minimal state automaton, it is not a minimal state inference.

Now we describe how the players evaluate their minimal state inferences in order to choose new strategies. To each model $M_j$ that player $i$ infers about player $j$ in decision period $s$ we associate a best response set $B_i(M_j)$. The optimistic player infers a model $M_j$ which has a best response that is most preferred relative to the best responses to the other models. The cautious player infers a model which has a best response that is least preferred relative to the best responses to the other models. The player who adapts switches to a strategy that is a best response to this model. We study the two dynamical systems that arise when both players use the same inference rule. Formally, if $M_i^s$ is chosen at decision period $s$ under the optimistic inference rule it must satisfy Condition (O), and if it is chosen under the cautious inference rule it must satisfy Condition (C).

**Condition O** (Optimistic inferences). $M_i^s \in B_i(M_j)$ for some $M_j \in I_j^{s-1}$, and $(M_i^s, M_j) \succ_i (B_i(M_j'), M_j')$ for all $M_j' \in I_j^{s-1}$.

**Condition C** (Cautious inferences). $M_i^s \in B_i(M_j)$ for some $M_j \in I_j^{s-1}$, and $(B_i(M_j'), M_j') \succ_i (M_i^s, M_j)$ for all $M_j' \in I_j^{s-1}$.

When $B_i(M_j')$ is multi-valued, we require that the conditions hold for each element of $B_i(M_j')$.

It may be the case that the current strategy used by a player is one of several strategies that satisfy the adaptation rule. In this case, an additional condition, minimization of switching costs, must be satisfied under both inference rules: If $M_i^{s-1} \in B_i(M_j)$, then $M_i^s = M_i^{s-1}$. That is, we require that $M_i^{s-1}$ be chosen in the $s$-th decision period in order to minimize switching costs$^{11}$.

$^{11}$Instead of incorporating switching costs into the adaptation procedure we could directly incorporate them into the preferences.
The adaptation process in the set of finite state automata can be described by the diagram:

\[(M_1^{s-1}, M_2^{s-1}) \xrightarrow{\psi_i} I_{s-1} \xrightarrow{B_i} (M_1^s, M_2^s).\]

Let \((M_1^1, M_1^2)\) be an arbitrary pair of finite state automata. Given this initial condition, the adaptation process generates sequences of automata, \(\{(M_1^s, M_2^s)\}_{s=1}^{\infty}\), which are the objects of study in the remainder of the paper. The set of possible sequences that the adaptation process generates for the initial condition \((M_1^1, M_1^2)\) is denoted \(\Gamma(M_1^1, M_1^2)\). This set may not be a singleton because even though the selection of player \(i\) is a single element from a best response set, the best response set may not be a singleton. All best responses must yield the same payoff and have same number of states, yet they may differ in their transition functions.

### 2.2 Examples of the Adaptation Process

The following two examples illustrate the adaptation process for particular initial conditions. The first example illustrates the observation, inference, and choice steps when both players start with an automaton that has a one period “show-of-strength” and then is willing to cooperate:

\[SOS = \begin{array}{cc}
  C & D \\
  q_1 & q_1 q_2 \\
  q_2 & q_2 q_1
\end{array}\] 

The outcome path that results in the first decision period is:

\[\pi(SOS, SOS) = \left\{ (D), (C), (C), \ldots \right\} \]

The set of minimal state inferences each player forms about play in the first decision period consists of 4 strategies:

\[\mathcal{I} = \left\{ \begin{array}{cc}
  C & D \\
  q_1 & q_1 q_2 \\
  q_2 & q_2 q_1
\end{array}, \begin{array}{cc}
  C & D \\
  q_1 & q_1 q_2 \\
  q_2 & q_2 q_1
\end{array}, \begin{array}{cc}
  C & D \\
  q_1 & q_1 q_2 \\
  q_2 & q_2 q_1
\end{array}, \begin{array}{cc}
  C & D \\
  q_1 & q_1 q_2 \\
  q_2 & q_2 q_1
\end{array} \right\} \]

For conciseness and computational purposes it is convenient to represent the set \(\mathcal{I}\) with the following *incompletely specified automaton* (Birkhoff and Bartee, [5]):

\[\mathcal{I} = \begin{array}{cc}
  C & D \\
  q_1 & -a q_2 \\
  q_2 & -b q_2
\end{array}\]

where \(-a\) and \(-b\) could be any state. 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 that always plays the action $D$. Starting at the initial condition $(SOS, SOS)$ the adaptation process, under optimistic inferences, generates a single sequence which converges to a steady state:

$$\left\{ \begin{pmatrix} SOS \end{pmatrix}, \begin{pmatrix} DEFECT \\ SOS \end{pmatrix}, \begin{pmatrix} DEFECT \\ DEFECT \end{pmatrix}, \begin{pmatrix} DEFECT \\ DEFECT \end{pmatrix}, \ldots \right\}$$

Thus, even though $(SOS, SOS)$ can be a Nash equilibrium of the game with preferences $\succ_i$, the players have an incentive to depart from it – it is not a steady state. What optimistic types infer about the latent behavior of the opponent creates an incentive to switch strategies.

Under cautious inferences a minimax selection is made by inspecting the following set of best responses:

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

2. The best response against any inference with $-b = q_1$ is:
   
   (a) DEFECT with the payoff $\delta(1 + g)/(1 - \delta)$, if $g \geq \delta$, or,
   
   (b) The solutions corresponding to the mapping $\{q_1 \rightarrow D, q_2 \rightarrow C\}$, yielding the payoff $\delta/(1 - \delta)$, if $g < \delta$. The best response set associated to this mapping is 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 adaptation 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 adaptation process under cautious inferences. It is less obvious that there are strategies which are not Nash equilibrium of the game with preferences $\succ_i$, but have the same outcome path as $(SOS, SOS)$, that are also steady states.

The second example illustrates the diversity of behavior that the adaptation process permits. Suppose the path observed after the first decision period is

$$\left\{ \begin{pmatrix} C \\ D \end{pmatrix}, \begin{pmatrix} D \\ C \end{pmatrix}, \begin{pmatrix} C \\ D \end{pmatrix}, \begin{pmatrix} D \\ C \end{pmatrix}, \ldots \right\}.$$ 

Player 1, the player who updates at the start of the second decision period, constructs 4 minimal state inferences about player 2:

$$I = \begin{array}{c|cc}
q_1 & C & D \\
q_2 & -a & q_1 \\
\end{array}$$

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 solutions corresponding
to the mapping \( \{q_1 \rightarrow D, q_2 \rightarrow C\} \) are given by the best response set:

\[
\begin{array}{c|cc}
  & C & D \\
\hline
q_1 & -c & q_2 \\
q_2 & q_2 & -d
\end{array}
\]

Since player 1’s new strategy has two states it is not obvious, without further calculation, what steady state the system will converge to. The outcome path in the second decision period is determined by the new strategy of player 1 and the true strategy of player 2. However, if player 1 makes an incorrect inference there will be a divergence between what player 1 expects to occur and what actually occurs, complicating the dynamics considerably.

These two examples suggest that two common types of behavior exhibited by optimistic agents are expectation of unpunished cheating and expectation of mutual cooperation. Thus, optimistic players can have incentives to play both competitively and cooperatively. The second example, 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.

### 2.3 Discussion

Some general guidelines for incorporating aspects of bounded rationality into dynamic games have been proposed in the game theory literature. Rubinstein [20] 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 [9] suggest that an adaptive model in which players are too naive may be inadequate for the study of social processes, even though it may provide good predictions of animal behavior. The model in this paper attempts to address these considerations by departing from fully rational players in several aspects including the use of *rules of thumb* for constructing a finite number of inferences and *limited foresight*. A by-product of incorporating these aspects of bounded rationality into the model is that we have defined an alternative way for a long-term competition situation to be played.

Limited foresight and the ability of the players to switch strategies during the course of play are important ingredients in the foundation of a model with adaptive agents. To have genuine adaptive agents it is necessary that there be limited foresight. 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 [2] 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. In contrast, limited foresight of the players prohibits them from completely thinking through a dynamic game before it begins. A natural way to incorporate limited foresight into a dynamic model is to let the players switch strategies at some frequency $T$, at which point they anticipate only $T$ periods ahead. Jéhiel [12], in the context of an equilibrium model, came to a similar conclusion in the development of his concept of a limited horizon forecast.

The asynchronous adaptation, in which players can only switch strategies in alternate decision periods, is a minor assumption. In continuous time, the assumption of asynchronous adaptation would be interpreted as an assumption that there is a zero chance of agents switching their strategies at the same time – they do not change their plans in lockstep. Adaptation that is always 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 strategy 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. Players build models of each other because they believe that the opponent will continue with the same strategy for some unknown duration\textsuperscript{12}. I expect the convergence results to still hold with some simultaneous adaptation, as long as adaptation is not simultaneous for an infinite number of consecutive decision periods.

3 Steady States

The solution concept is a steady state of the adaptation process.

**Definition 1.** A pair of finite state automata, $(M^*_1, M^*_2)$, is said to be a steady state of the adaptation process if whenever $(M^S_1, M^S_2) = (M^*_1, M^*_2)$ then $\{(M^s_1, M^s_2)\}_{s=S}^{\infty} = \{(M^*_1, M^*_2), (M^*_1, M^*_2), ...\}$.

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 decision period: $\{I^s_1, I^s_2\} = \{I^*_1, I^*_2\}$ for all $s$.

In Theorem 1 it is shown that in any steady state of the adaptation process under optimistic inferences the players will play a Nash equilibrium of the stage game in each period. Corollary 1 states that when the stage game is the Prisoners’ Dilemma the unique steady state of the adaptation process, under optimistic inferences, is the one-state strategy that always plays Defect. Figure 3 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.

\textsuperscript{12}Discounting in a decision period can be interpreted as representing uncertainty about when the opponent switches strategies.
Corollary 1 shows that with optimistic inferences the steady state payoff is a singleton, that is, (0,0). Two lemmas will be proved before the proof of Theorem 1 is presented. In Lemma 1 it is shown that if the process is at a steady state then: (1) the strategies of both players have the same number of states, and (2) one of the models of the other player is the true strategy of the other player (Example 1 illustrated that the second implication does not necessarily follow if the pair of strategies is not a steady state.). In Lemma 2 it is shown that every steady state under optimistic inferences is a Nash Equilibrium in the game with preferences $\succ_i$.

**Theorem 1** (steady states under optimistic inferences). *For any finite action stage game, in any steady state of the adaptation process under Condition (O) a Nash equilibrium of the stage game is played in each period.*

**Corollary 1.** *When the stage game is the Prisoners’ Dilemma the unique steady state of the adaptation process under Condition (O) is $M^*_1 = M^*_2 = \text{DEFECT}$. *

An implication of Theorem 1 is that there are Nash equilibria that are not steady states of the adaptation process under Condition (O). Predictions in dynamic games using standard equilibrium analysis are obtained by verifying whether or not strategies are an equilibrium, even though in most situations that economists model the assumption that the players themselves would know whether or not they are at an equilibrium is a strong assumption. When strategies are private information players cannot know the complete strategy of the opponent unless it is assumed or else they
elicit the information through experimentation or communication. Yet, once this assumption is removed, the players have incentives to switch strategies. For example, even if the players are playing equilibrium strategies in a repeated game they can construct many alternative models of each other's strategies that are consistent with observed play. Yet, if one of these models were believed to be the true strategy of the opponent, optimal behavior could require switching strategies.

Adopting a new strategy can be viewed as an experiment. Such experiments can be costly and even irreversible, and, in the adaptation process, result in an inevitable chain of events, dependent on the models they construct of each other. For example, in the Prisoners’ Dilemma, the elimination of Nash equilibria involving cooperation by the adaptive process under 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. Some researchers have argued that an automaton which starts with a "show-of-strength" allows the opponent to learn about its punishment capability, thereby providing a deterrent. The adaptation process under optimistic inferences 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.

One interpretation of the prediction in Corollary 1 is that even when the players are patient, if they construct optimistic models of the world they are led to play stationary strategies in the infinitely 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 the adaptation process under Condition (O).

The following basic results will be used in the proofs.

1. By the Markov decision problem (MDP) of player i associated to automaton $M_j$ is meant the following: Choose a sequence of actions $\{a^{(t)}_i\}^\infty_{t=1}$ that maximizes $(1-\delta)\sum_{t=1}^\infty \delta^{t-1}u_i(a^{(t)}_i, f_j(q^{(t)}_j))$ under the law of motion $q^{(t+1)}_j = \tau_j(q^{(t)}_j, a^{(t)}_i)$ that has initial state $q^{(1)}_j$. It is well known that a stationary solution $f_i : Q_j \rightarrow A_i$ 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 $\{(f_i(q^{(t)}_j), f_j(q^{(t)}_j))\}^\infty_{t=1}$ obtained in (1). This outcome path, by definition, attains the optimal value of player i’s MDP. Of particular interest is the best response set, $B_i(M_j)$, which contains

---

This notion of experimentation stands in contrast to the notion of experimentation in repeated games in which players can costlessly "test" an opponent who plays a fixed strategy.
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, $f_i : Q_j \rightarrow A_i$, 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 $\{(f_i(q_j^t), f_i(q_j^t'))\}_{t=1}^\infty$.

(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 repeated 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 adaptation process is bounded by $\max\{|M_1^*|, |M_2^*|\}$.

**Lemma 1.** For any finite action stage game, if $(M_1^*, M_2^*)$ is a steady state of the adaptation process under Condition (O) or (C) 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 $j$: $|M_j^*| \geq C_j^*$. The claim will follow once it is shown that $|M_j^*| = C_j^*$ for $j = 1, 2$.

Notice that $|M_1^*| \leq C_2^*$ and $|M_2^*| \leq C_1^*$: 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: $|M_i^*| \leq C_j^*$.

Since $(M_1^*, M_2^*)$ is a steady state there is a decision period $s$ such that $|M_j^s| = |M_j^s+1| = |M_j^s+2|$ for $j = 1, 2$. That $|M_j^s| \geq C_j^s$ and $|M_j^s| \geq |M_j^s+1|$ implies that $|M_j^{s+2}| \leq |M_j^{s+1}| \leq |M_j^s|$, or $|M_j^s| \leq |M_j^s| \leq |M_j^s|$. Thus, $|M_1^*| = |M_2^*|$ and $C_i^* \leq |M_i^*| = |M_j^*| \leq C_i^*$. □

**Lemma 2.** For any finite action stage game, if $(M_1^*, M_2^*)$ is a steady state of the adaptation process under Condition (O) then it is a Nash equilibrium of the repeated 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 \(\pi(M_1^*, M_2^*) = \pi(M_1^*, M_j^*)\).

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, \(M_i^* \in B_i(M_j^*)\). \(\blacksquare\)

Lemma 2, although short, will not necessarily follow if any of the three key ingredients are 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 decision periods in a steady state, and (iii) optimistic inferences. The consequence of Lemma 2 states that even if the model of the opponent is not the opponent’s true strategy, in which case the new strategy is a best responses to an incorrect model, nevertheless the pair of strategies in a steady state is a mutual best response. This stands in contrast to the adaptation process under cautious inferences when a steady state is not necessarily a Nash equilibrium, but is preserved as a steady state precisely because incorrect models are believed.

**Proof** (Theorem 1). Suppose the steady state \((M_1^*, M_2^*)\) generates the sequence of states

\[
\{q^1, \ldots, q^{t_1-1}, q^{t_1}, \ldots, q^{t_2}, q^{t_3}, \ldots, q^{t_2}, \ldots\},
\]

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

Assuming the steady state were to involve play that is not a Nash equilibrium of the stage game we construct an inference which permits an improvement over \(M_i^*\), thereby leading to a contradiction. By Theorem 1 in Abreu and Rubinstein [3], if \((M_1^*, M_2^*)\) is a Nash Equilibrium of the repeated game with preferences \(\succ_i\) the states of \(M_i^*\) (respectively \(M_2^*\)) which appear in the first \(t_2\) periods are distinct. Therefore, there exists a minimal state inference, denoted \(M'_j\), which has exactly \(t_2\) states and has the following structure.

\[
M'_j := \begin{pmatrix}
q^1 & a_{i1} & a_{i2} & \ldots & a_{ik} & f_j(q^1) \\
q^2 & q^2 & q^3 & \ldots & q^3 & f_j(q^2) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
q^{t_2} & q^{t_1} & q^{t_1} & \ldots & q^{t_1} & f_j(q^{t_2})
\end{pmatrix}
\]

The \(k\) actions of player \(i\) are the inputs and the set of states for each player is represented by the common set \(Q = \{q^1, q^2, \ldots, q^{t_2}\}\). \(M'_j\) is an inference about player \(j\) that imitates the states and realized transitions of \(M_i^*\) and plays the actions of player \(j\) in the order that 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 \(j\): \(M'_j \in \mathcal{I}_j^*\).

Suppose \(M_1^*\) and \(M_2^*\) constitute a steady state and there is a period in which play is not a Nash equilibrium of the stage game. If \(C^* = 1\) and \(f^*_i(q)\) is not a
best response to \( f^*_i(q) \) there cannot be a steady state. Therefore, assume \( C^* > 1 \) and that \( f^*_i(q) \) is not a best response to \( f^*_i(q) \) for some \( q \in Q_i^* \). By inferring the minimal state inference \( M_i^* \) player \( i \) can increase the payoff in state \( q \) by playing a best response to \( f^*_i(q) \) without affecting the payoff in the other states (i.e. without punishment). Thus, the strategy \( M_i^* \) will not be chosen, a contradiction to the steady state assumption. ■

When the stage game is the Prisoners’ Dilemma all Nash equilibrium that involve cooperation are eliminated as candidate steady states. In this case one can verify that the pair of automata (DEFECT, DEFECT) is a steady state.

Theorems 2 and 3 characterize the set of steady states under cautious inferences. Although every Nash equilibrium of the game with preferences \( \succ_i \) is a steady state (Theorem 2), 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 (Theorem 3). 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 \( \succ_i \). In the repeated Prisoners’ Dilemma example, Theorems 2 and 3 demonstrate that the set of steady state payoffs of the adaptation process under cautious inferences are those points on the two diagonals in Figure 3 that also Pareto dominate \((0,0)\). Note that this is the same as the set of equilibrium payoffs in the repeated game with preferences \( \succ_i \) (see Abreu and Rubinstein [3]).

**Theorem 2** (steady states under cautious inferences). For any finite action stage game, if \((M_1, M_2)\) is a Nash equilibrium of the repeated game with preferences \( \succ_i \) then \((M_1, M_2)\) is a steady state of the adaptation process under Condition (C).

**Proof.** It must be shown that if \((M_i^*, M_j^*) = (M_{i1}, M_{j2})\) then \((M_i^{*+1}, M_j^{*+1}) = (M_i, M_j)\).

First we demonstrate that if the decision period \( s \) strategies are a Nash equilibrium of the repeated game with preferences \( \succ_i \) then they are minimal state inferences: If \((M_i^*, M_j^*)\) is a Nash equilibrium then \((M_i^*, M_j^*) \in T_i^s \times T_j^s\). Suppose \((M_i^*, M_j^*)\) is a Nash equilibrium of the game with preferences \( \succ_i \) and \( M_i^* \) is not a minimal state inference. Then there is an automaton \( M_i \neq M_i^* \) that has fewer states, and since \( \pi(M_i, M_i^*) = \pi(M_i^*, M_j^*) \) it attains the same repeated game payoff, a contradiction.

This establishes that \( M_i^* \) and \( M_j^* \) are both best responses to minimal state inferences. If it is shown that they are least preferred responses in the decision period \( s + 1 \) problem, then due to the minimization of switching costs they will be selected at decision period \( s + 1 \).

Suppose that player \( i \) updates at decision period \( s + 1 \). Notice that all pairs of automata in the set \( T_i^s \times T_j^s \) yield the same outcome path and payoff. This implies that for all inferences \( M_j \in T_j^s \) it is true that \( (M_i^*, M_j) \succ_i (M_i^*, M_j^*) \). Since \((B_i(M_j), M_j) \succ_i (M_i^*, M_j)\), it follows, by transitivity, that \((B_j(M_j), M_j) \succ_i (M_i^*, M_j^*)\) for all \( M_j \in T_j^s \). Thus, Condition (C) is satisfied. ■

19
Theorem 3 (steady states under cautious inferences). For any finite action stage game, if \((M_1^*, M_2^*)\) is a steady state of the adaptation process under Condition (C) then there is a pair of minimal state inferences \((M_1, M_2)\) that is a Nash equilibrium of the repeated game with preferences \(\succ_i\).

Proof. Since \((M_1^*, M_2^*)\) is a steady state of the adaptation process, condition (C) implies that \(M_1^* \in B_1(M_2)\) for some \(M_2 \in I_2\) and \(M_2^* \in B_2(M_1)\) for some \(M_1 \in I_1\). Every response in \(B_i(M_j)\) has the same number of states and attains the same payoff \(v_i^*\), thus, \(I_i^* \subset B_i(M_j)\). 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 repeated game with preferences \(\succ_i\).

The content of Theorem 3 is that a steady state is a Nash equilibrium in beliefs; the players’ inferred models of each other constitute a Nash equilibrium of the game with preferences \(\succ_i\). Theorem 3 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 \(\pi(M_1^*, M_j) = \pi(M_1^*, M_2^*)\), where \(M_i^* \in B_i(M_j)\) for some \(M_j \in I_j\). That is, each player’s model is confirmed when \((M_1^*, M_2^*)\) is played next decision period, even though the model \(M_j\) may not be the true strategy 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 [8]). Weaker than a Nash equilibrium, the concept of self-confirming equilibrium is intrinsic to learning models in which only realized play is observed. Although the existence of steady states in addition to Nash equilibria might not appear to be in the spirit of equilibrium selection, observed that the entire set of steady states is not reachable from every initial condition. To the extent that economists expect outcomes of real coordination problems to depend on initial conditions, this type of path dependence acts to sharpen the predictions obtained in Theorems 2 and 3.

4 Convergence

Convergence concerns what happens to the adaptation process when it is started at initial strategies which are not steady states. The steady state results reveal little about convergence. In this section we study the adaptation process when the stage game is the Prisoners’ Dilemma. It is shown that all sequences of strategies generated by the adaptation process from any pair of strategies in the set of one and two state automata must eventually reach a steady state. First, we characterize the properties that must be satisfied by a cycle, and then, to prove that a cycle does not exist, these properties are used to obtain contradictions.

Definition 2. The adaptation process is said to have a cycle if there is a sequence of automata, and no decision period \(S\) such that the sequence is constant after \(S\).

Lemma 3 (Properties of a Cycle). If a cycle exists then:
**Property 1:** All strategies of both players in the cycle must have the same number of states: there is some decision period $S$ such that $|M_1^{S+s}| = |M_2^{S+s}|$, for all $s$.

**Property 2:** A decision period $s$ selection of player $i$ (the player who can update) in the cycle: (a) has the same number of states as any minimal state inference about player $j$ in decision period $s-1$ : $|M_i^s| = C_{j}^{s-1}$, and, (b) is in the minimal state inference set at epoch $s$ : $(M_i^s \in \mathcal{T}_i^s)$.

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

$$|M_1^s| \geq |M_2^s| \geq |M_2^s| \geq ...$$

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 decision period $S$, is the same integer. Thus, the sequence converges to an integer. Since $|M_i^{s+1}| = |M_i^s|$ if $s$ is even, and $|M_j^{s+1}| = |M_j^s|$ if $s$ is odd, it follows that $|M_1^{S+s}| = |M_2^{S+s}|$ for all $s$.

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

Although Property 1 is a strong restriction on a cycle, it does not immediately follow that there cannot be a cycle – Property 1 does not rule out the possibility that players choose different strategies which have the same number of states. Moreover, although state reduction is necessary to converge to the steady state under optimistic inferences, it does not follow that optimization always results in state reduction, as was demonstrated in the examples. However, the restrictions imposed by these properties do lend credence to the more general conjecture that when the stage game is the Prisoners’ Dilemma there is global convergence in the set of finite state automata.

The adaptation process is a mapping from the set of finite state automata into itself. The approach of the convergence proofs is to partition the set of all outcome paths that can be generated by one and two state automata according to which of the four possible action pairs appear in the outcome path $\{(C,C), (C,D), (D,C), (D,D)\}$. The proof of Theorem 4, convergence under optimistic inferences, appears in the appendix and demonstrates that all sections of the partition, except one, are transitory. The section of the partition which contains the single sequence that repeats $(D,D)$ is the only absorbing section. The proof of Theorem 5, convergence under cautious inferences, also appears in the appendix and demonstrates that for every section of the partition the adaptation process either converges to a steady state in that section or else enters another section which is absorbing.
Theorem 4 (convergence under optimistic inferences). When the stage game is the Prisoners’ Dilemma, the adaptation process under Condition (O) converges to the unique steady state from any initial condition in the set of 1 and 2 state automata, for any $g, l > 0$, and all $0 < \delta < 1$.

Theorem 5 (convergence under cautious inferences). When the stage game is the Prisoners’ Dilemma, the adaptation process under Condition (C) converges to the set of steady states from any initial condition in the set of 1 and 2 state automata, for any $g, l > 0$, and all $0 < \delta < 1$.

5 Conclusion

Many equilibrium selection methods modify the environment in which rational players compete. The approach to equilibrium selection taken in this paper is to modify the decision processes of the players and retain the basic environment. This approach establishes a correspondence between the outcomes that are selected by the adaptation process and the behavioral assumptions that define the adaptation rule. Two behavioral types, optimistic and cautious, are used to illustrate this method of equilibrium selection by behavioral type. 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 an accidental tremble, but by errors in 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 with scarce information and ordinary abilities of inference eliminates Nash equilibria present in the standard repeated game in which it is assumed that players do not adapt their strategies\textsuperscript{14}. Equilibrium analysis is silent on how equilibrium play arises and is a static approach in this regard, even in standard dynamic games. This deficiency of standard equilibrium analysis appears as a significant problem in learning models in which players are allowed to start the game with arbitrary strategies and it is assumed players can only observe realized play.

In the context of a class of games which have a multiplicity of equilibria – infinitely repeated games – this paper investigated how the nature of the strategies that players coordinate on crucially depends on their adaptive behavior. The problem was presented in a general model of adaptation in repeated games in which players can begin the game with any feasible strategies and are permitted to adapt their strategies in the course of play. The dynamics occur in a strategy space, not just an action space, which is a significant advantage of representing strategies by finite state automata.

\textsuperscript{14}The use of subgame perfect equilibrium as the solution concept in a dynamic game implicitly assumes that the equilibrium strategies are common knowledge to the players— that is, that players observe their opponent’s strategy.

22
The rules that define the mode of adaptation are a set of behavioral assumptions. In particular, the players’ behavior is restrained by complexity considerations, both in the strategies that they choose and in the models that they construct of each other’s strategies. These assumptions, together with the assumption that players can only observe realized play, provide the players incentives to adapt their strategies. Under one type of behavior, optimistic inferences, players eventually only coordinate on outcomes that can be obtained by Nash equilibria of the stage game. Under another type of behavior, cautious inferences, players can coordinate on strategies that sustain outcomes that cannot be supported by Nash equilibria of the stage game, in particular the subset of self-confirming equilibria with Nash outcome paths.

6 Appendix

Proof of Theorem 4. Theorem 4 follows from Lemmas 4 through 11.

Lemma 4: The adaptation process under Condition (O) or (C) does not have a cycle in one-state strategies.

Proof: By Lemma 3 every automaton in the sequence, after some decision period $s$, has the same number of states, $Q^*$. When $Q^* = 1$, the players have a unique minimal state inference of each other. Eventually both players will select DEFECT, which is the steady state.

Lemma 5: The adaptation process under Condition (O) does not have a cycle in two-state strategies with a path $\pi(M_1^s, M_2^s)$ that consists only of the terms $(D)$ and $(O)$.

Proof: Suppose there is such a cycle. There are two possible sets of minimal state inferences to consider.

Case 1: The set of minimal state inferences at decision period $s - 1$ is:

$$T_j^{s-1} = \begin{bmatrix} C & D \\ q_1 & q_2 \\ q_2 & q_{k} \end{bmatrix}$$

where the state $q_k$ represents $q_1$ or $q_2$. Player $i$ can attain the maximum possible payoff $(1 + g)/(1 - \delta)$ by choosing DEFECT, a contradiction to Lemma 3.

Case 2: The set of minimal state inferences at decision period $s - 1$ is:

$$T_j^{s-1} = \begin{bmatrix} C & D \\ q_1 & q_2 \\ q_2 & q_{k} \end{bmatrix}$$
Given that player \(j\) chooses \(D\) in the first period, player \(i\) can attain the highest possible payoff, \(\delta(1 + g)/(1 - \delta)\), with DEFECT, a contradiction to Lemma 3.

**Lemma 6:** The adaptation process under Condition (O) does not have a cycle in two-state strategies with a path \(\pi(M^*_1, M^*_2)\) that consists only of the terms \((D)\) and \((C)\).

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

\[
I_{j}^{s-1} = \begin{array}{cc}
C & D \\
\frac{1}{-a} q_2 & \frac{1}{-b} q_k & C \\
\frac{1}{-a} q_2 & \frac{1}{-b} q_k & D
\end{array}
\]

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 decision period \(s\), a contradiction to Lemma 3.

Case 2: The set of minimal state inferences is:

\[
I_{j}^{s-1} = \begin{array}{cc}
C & D \\
\frac{1}{-a} q_2 & \frac{1}{-b} q_k & C \\
\frac{1}{-a} q_2 & \frac{1}{-b} q_k & D
\end{array}
\]

It is always optimal for player \(i\) to infer \(-a = q_2\). Since one-state solutions lead to a contradiction, assume that \(q_k = q_1\) and \(-b = q_2\) to obtain the candidate two-state solution, \(\{q_1 \rightarrow D, q_2 \rightarrow C\}\). The best responses that implement this solution which do not immediately lead to a one-state strategy being selected in the next decision period are:

\[
B_i(M_j) = \begin{array}{cc}
C & D \\
\frac{1}{-c} q_2 & \frac{1}{-d} q_2 & D \\
\frac{1}{-c} q_2 & \frac{1}{-d} q_2 & C
\end{array}
\]

Recall that Property 2 of a cycle requires that one of the minimal state inferences in \(I_{j}^{s-1}\) be the true strategy. To obtain a contradiction consider player \(j\)’s decision problem at decision period \(s + 1\). Iterate the two incompletely specified automata until an undefined term is reached: \(\pi(B_i(M_j), I_{j}^{s-1}) = \{(D), (C), \}\). Hence, the transition \(\tau_i(q_1, D) = q_2\) for every inference about player \(i\) in \(I_i^s\). This means that at player \(j\)’s decision period \(s + 1\) decision problem it is optimal to play \(D\) in the first state. Then the only solution of player \(j\) that is not one-state is \(\{q_1 \rightarrow D, q_2 \rightarrow C\}\). This solution is implemented by

\[
B_j(M_i) = \begin{array}{cc}
C & D \\
\frac{1}{-e} q_2 & \frac{1}{-f} q_2 & D \\
\frac{1}{-e} q_2 & \frac{1}{-f} q_2 & C
\end{array}
\]
By Property 2 of a cycle, $M^s_i \in \mathcal{I}^s_i$. However, notice that for any $(M^s_i, M^{s+1}_j) \in B_i(M_j \times B_j(M_i)$ that $\pi(M^s_i, M^{s+1}_j) = \{(D), (C), (C), \ldots\}$. Under asynchronous adaptation, $\pi(M^{s+1}_i, M^{s+1}_j) = \pi(M^s_i, M^{s+1}_j)$, 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 we need to characterize some properties of the best response sets. The proof of Lemma 7 is similar to Theorem 1 in Piccione [18].

**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 $(\ddag)$ and $(\ddag)$.
(b) $\pi(M_i, M_j)$ consists only of terms $(\dagger)$ and $(\ddag)$.

**Lemma 8:** The adaptation process under Condition (O) does not have a cycle in two-state strategies when player $i$, who updates at decision period $s$, has a unique inference about player $j$ at decision period $s - 1$. Moreover, there does not exist a cycle in two-state strategies in which 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, $\pi(M^s_i, M^s_j)$ consists only of the terms $(\ddag)$ and $(\ddag)$, or only of the terms $(\dagger)$ and $(\ddag)$. By Lemmas 5 and 6, respectively, these outcome paths cannot be part of a cycle. Moreover, when all four action pairs occur on an outcome path and the path can be generated by a pair of two-state automata the minimal state inference sets are singletons.

**Lemma 9:** The adaptation process under Condition (O) does not have a cycle in two-state strategies with a path $\pi(M^s_1, M^s_2)$ that consists only of the terms $(\dagger)$, $(\ddag)$, and $(\ddag)$.

**Proof:** Each set of minimal state inferences consists of exactly two automata. 

*Case 1:* The two minimal state inferences are:

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

An optimistic player will infer $-a = q_2$ and optimally choose the action $D$ in $q_1$. The only two-state solution to consider is $\{q_1 \rightarrow D, q_2 \rightarrow C\}$. If $q_2 = q_2$, or if $q_k = q_1$ and $q_j = q_1$ this solution would not be optimal. It remains to consider when $q_k = q_1$ and $q_j = q_2$. Player $i$’s best responses which do not immediately lead to a one-state solution in the next decision period are:
\[ B_i(M_j) = \begin{array}{cc|cc}
C & D \\
q_1 & -b & q_2 & D \\
q_2 & q_2 & -c & C 
\end{array} \]

The resulting path is \( \{(D), (D), (C), \ldots\} \), a contradiction to Lemma 5.

Case 2: The two minimal state inferences are:

\[ T_j^{s-1} = \begin{array}{cc|cc}
C & D \\
q_1 & q_i & q_j & C \\
q_2 & q_k & -a & D 
\end{array} \]

If \( q_j = q_1 \) a one-state solution is chosen. Otherwise, suppose \( q_j = q_2 \). If player \( i \) infers \(-a = q_1\) then the action \( D \) is optimal in \( q_2 \) for any \( q_k \). A one-state solution is always optimal, a contradiction.

Lemma 10: The adaptation process under Condition (O) does not have a cycle in two-state strategies with a path \( \pi(M_1^s, M_2^s) \) that consists only of the terms \( (D), (D), (C) \), and \( (D) \).

Proof. Case 1: The two minimal state inferences are:

\[ T_j^{s-1} = \begin{array}{cc|cc}
C & D \\
q_1 & -a & q_i & C \\
q_2 & q_j & q_k & D 
\end{array} \]

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 \( \{q_1 \rightarrow D, q_2 \rightarrow C\} \). This could only be optimal if the two minimal state inferences are:

\[ T_j^{s-1} = \begin{array}{cc|cc}
C & D \\
q_1 & -a & q_i & C \\
q_2 & q_1 & q_2 & D 
\end{array} \]

Since the candidate two-state solution only involves transitions that are common to both inferences, the outcome path that results at decision period \( s \) is \( \{(C), (C), (D), (C), \ldots\} \). By Lemma 6 this outcome path cannot be part of a cycle.

Case 2: When the two minimal state inferences are given by

\[ T_j^{s-1} = \begin{array}{cc|cc}
C & D \\
q_1 & q_i & q_j & D \\
q_2 & q_k & -a & C 
\end{array} \]

arguments similar to those given in Case 1 obtain contradictions via Lemmas 5 or 6.
Lemma 11. The adaptation process under Condition (O) does not have a cycle in two-state strategies with a path $\pi(M^s_1, M^s_2)$ that consists only of the terms $(D_D), (C_C), \text{ and } (D_D)$.

Proof. Similar to Lemma 10.

Proof of Theorem 5. Theorem 5 follows from Lemma 4, and Lemmas 12 through 14.

Lemma 12. The adaptation process under Condition (C) does not have a cycle in two-state strategies with a path $\pi(M^s_1, M^s_2)$ that consists only of the terms $(C_C)$ and $(D_D)$.

Proof. Case 1: Suppose the set of minimal state inferences at decision period $s - 1$ for both players is represented by:

$$I_{j-1}^s = \begin{bmatrix} C & D \\ q_1 & q_2 \\ -a & q_1 \\ q_2 & -b \\ \\ D \\ q_2 & q_1 \\ b & q_2 \end{bmatrix}$$

The solution is DEFECT for all four inferences in each set for all parameters $l, g$, which contradicts Lemma 3.

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

$$I_{j-1}^s = \begin{bmatrix} C & D \\ q_1 & q_2 \\ -a & q_1 \\ q_2 & -b \\ \\ D \\ q_2 & q_1 \\ b & q_2 \end{bmatrix}$$

The cautious player selects a best response to a model which has a payoff no greater than a best response to any other model. We compare the following best responses:

- 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 Lemma 3.

Case 3: Suppose the set of minimal state inferences at decision period $s - 1$ for both players is:

$$I_{j-1}^s = \begin{bmatrix} C & D \\ q_1 & q_2 \\ -a & q_1 \\ q_2 & -b \\ \\ D \\ q_2 & q_1 \\ b & q_2 \end{bmatrix}$$
We compare the following best responses:

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 \(\delta(1 + g)/(1 - \delta^2)\).

If \(g < \delta\) the two-state automaton in BR2 is a solution and is chosen for the next decision period by a cautious player. If \(g \geq \delta\) a one-state automaton is chosen for the next decision period.

When the two-state solution is chosen the set of automata that implement it is exactly \(\mathcal{I}_{s-1}^i\). Since this set includes \(M_{s-1}^i\), the condition for minimization of switching costs requires that \(M_{s-1}^i = M_{s-1}^j\). A steady state has been reached.

Thus, if the adaptation process reaches an outcome path that consists only of the terms \((C)\) and \((D)\) then it converges to a steady state.

**Lemma 13.** The adaptation process under Condition (C) does not have a cycle in two-state strategies with a path \(\pi(M_1^i, M_2^j)\) that consists only of the terms \((C)\) and \((D)\).

**Proof.** Case 1: The sets of minimal state inferences at decision period \(s - 1\) for players \(i\) and \(j\) are derived from the path \(\{(...,(C),(D),(D),...\}\) and represented by:

\[
\mathcal{I}_{s-1}^i = \begin{pmatrix}
q_1 & C & D \\
q_2 & -a & q_2 \\
d & q_2 & -b \\
d & D & C
\end{pmatrix}
\]

and

\[
\mathcal{I}_{s-1}^j = \begin{pmatrix}
q_1 & C & D \\
q_2 & -c & q_2 \\
d & q_1 & -d \\
d & D & C
\end{pmatrix}
\]

For all four models of player \(i\) the solution is either COOPERATE or DEFECT, a contradiction to Lemma 3.

For player \(i\)’s decision problem with respect to models \(\mathcal{I}_{s-1}^j\) we compare the following best responses:

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

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

The response in BR2 does not depend on \(-c\) or \(-d\), which leads to a contradiction via Lemma 12.

Case 2: The sets of minimal state inferences at decision period \(s - 1\) for players \(i\) and \(j\) are derived from the path \(\{(...,(C),(D),(D),...\}\) and represented by:

\[
\mathcal{I}_{s-1}^i = \begin{pmatrix}
q_1 & C & D \\
q_2 & -a & q_2 \\
d & q_1 & -b \\
d & D & C
\end{pmatrix}
\]

and

\[
\mathcal{I}_{s-1}^j = \begin{pmatrix}
q_1 & C & D \\
q_2 & -c & q_2 \\
d & q_1 & -d \\
d & D & C
\end{pmatrix}
\]

Consider player \(j\)’s decision problem using the models \(\mathcal{I}_{s-1}^i\).

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

Suppose that the two-state response in BR4 is optimal. Then, since \(A_1 = A_2\), the responses in BR1 and BR2 indicate that a two-state response is chosen for the decision period if cooperating forever is preferred to alternating \((D)\) and \((C)\) forever. Otherwise, a one-state solution is selected.

Consider player i’s decision 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. \(\{q_1 \rightarrow C, q_2 \rightarrow D\}\) 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. \(\{q_1 \rightarrow D, q_2 \rightarrow C\}\) yields payoff \(\delta/(1 - \delta) = \delta(1 + \delta)/(1 - \delta^2)\).

If a two-state response is optimal in BR1 it is selected if \(A_3 \leq A_4\). This is equivalent to saying that cooperating forever is preferred to alternating \((D)\) and \((C)\) forever. The two-state response in BR4 is never chosen for the decision period by a cautious player. Otherwise, a one-state solution is selected at the decision period, a contradiction to Lemma 3.

Assume that player i selects the two-state solution. Then this solution can be implemented exactly with the set of automata \(I^s_{i-1}\). Likewise, if player j selects a two-state solution, it can be implemented with the set \(I^s_{j-1}\). Notice that for any pair of automata in \(I^s_i \times I^s_j\) the outcome path is \(\{(D), (C)\}; \{(D), (C)\}; \ldots\}. Hence, any pair of automata in \(I^s_i \times I^s_j\) is a steady state.

**Remark.** Lemmas 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.** The adaptation process under Condition (C) does not have a cycle in two-state strategies with a path \(\pi(M^1_i, M^2_j)\) that consists only of the terms:

\[
\begin{align*}
&b. \text{COOPERATE yields } \{1 \rightarrow C, 2 \rightarrow D\} \text{ yields payoff } A_3 = -l + \delta(1 + g - \delta l)/(1 - \delta^2). \\
&b. \text{COOPERATE with payoff } A_4 = -l + \delta(1 + \delta)/(1 - \delta^2). \\
&b. \text{DEFECT yields payoff } \delta(1 + g)/(1 - \delta^2). \\
&b. \text{DEFECT yields payoff } \delta/(1 - \delta) = \delta(1 + \delta)/(1 - \delta^2). \\
&b. \text{DEFECT with payoff } 1 + g. \\
&b. \{q_1 \rightarrow D, q_2 \rightarrow C\} \text{ with payoff } A_1 = (1 + g - \delta l)/(1 - \delta^2).
\end{align*}
\]
Proof. Similar to Lemmas 12 and 13.

References


