Bounded Rationality in Repeated Games*

Julian Romero†
California Institute of Technology

January 23, 2010

Job Market Paper

Abstract:

Models of bounded rationality often lead to sharper predictions about real world outcomes than their full rationality counterparts. Full rationality in repeated interactions allows a plethora of equilibrium outcomes. In this paper, I examine the effect of bounded rationality in infinitely repeated games. In particular, does the introduction of boundedly rational agents lead to a smaller set of outcomes in equilibrium?

I show that the number of equilibrium outcomes is smaller when agents are boundedly rational. Importantly, cooperative outcomes are still possible in equilibrium, even when players can't use sophisticated strategies and are not able to perfectly monitor their opponents. The strategy that leads to cooperation is called "Win-Stay, Lose-Shift". Using this strategy, I show that cooperation is possible in equilibrium for a large class of 2x2 games.

I also give necessary and sufficient conditions on equilibrium structure for two-player $N \times M$ games. These conditions suggest that in equilibrium, players must be able to cooperate without getting caught in long periods of conflict.

JEL classification: *C62,C72,C73*

Keywords: Repeated Games, Finite Automata, Bounded Rationality, Prisoner's Dilemma

---

*Many thanks John Ledyard, Leeat Yariv, and Jacob Goeree as well as seminar participants at Caltech, UCLA, and the 2009 SoCal NEGT Symposium for useful comments. Thanks to the Social Information Sciences Laboratory for funding this research. Existing errors are my sole responsibility.

†Email: jnr@hss.caltech.edu. Phone: 626-395-3142. Web: www.hss.caltech.edu/∼jnr. Mailing Address: California Institute of Technology 1200 E. California Blvd, MC 228-77 Pasadena, CA 91125-0077.

1 Introduction

Models of bounded rationality assume that agents have limited ability to process information and solve complex problems (Simon 1957). These models are often able to make sharper predictions than their fully rational counterparts (Conlisk 1996). When players are fully rational and they interact repeatedly, a plethora of equilibrium outcomes are possible. In particular, these games suffer from folk theorems; namely any individually rational and feasible payoff[1] is attainable in equilibrium. This multitude of equilibria suggests further analysis of the equilibrium selection problem is needed. This paper focuses on the question: Does a model of repeated interactions with boundedly rational agents lead to a smaller set of outcomes in equilibrium? A smaller set of outcomes is required to make better predictions about what type of behavior we should expect to see in repeated interactions.

In this paper, players are limited in two ways. First, as in many repeated interactions, players are not able to see the actions of their opponents. Rather, they get an imperfect signal from which the action must be inferred. An example is the "secret price cutting" game (Stigler 1964), in which two competing firms give unobservable price cuts to their customers, which can only be inferred through sales figures. In this paper, each player receives a private signal that correctly conveys the action of their opponent with probability (accuracy) less than one. This builds on the literature that examines imperfect private monitoring in repeated games (Kandori 2002).

The second limitation involves memory constraints. Typical repeated game strategies require that players have perfect memory, and can differentiate between every possible infinitely repeated game history[2]. Due to memory constraints, it is inconceivable that any economic agent could differentiate between every history in this infinite set. Here, I assume that players are able to classify this infinite set of histories into a finite number of groups (referred to as states). This leads to an intuitive class of strategies that capture the simple heuristics used during the infinitely repeated game.

[1] A payoff is feasible if there is some infinite sequence of actions which leads to this payoff. A payoff is individually rational if each player is receiving at least their minmax payoff. The minmax payoff is the payoff that minimizes that maximum possible loss. See Mailath and Samuelson (2006) for more information.

[2] A history refers to a sequence of previously played actions.

By limiting recall to finite states, I can represent players' strategies by finite automata. Intuitively, a finite automaton can be thought of as a set of mental states. Each state represents a different mood (for example good and bad), and therefore may lead to a different behavior (nice and mean). Based on the actions of the other player, the mood might change, in which case the automaton moves (transitions) to a different state. A more sophisticated player may have many states which represent a complex strategy, while a simple player may have only a handful of states. Representing strategies with finite automata was first suggested by Aumann (1981) and has been widely studied since (Mailath and Samuelson 2006).

Using agents that are limited in the ways described above, I examine a model of repeated games, where it is possible for players to attain cooperative relationships without using contracts. The main insight from this paper is that in equilibrium players must play strategies that attain cooperation, and are forgiving enough to avoid long conflicts if cooperation breaks down. Conflicts between players are suboptimal, so the time spent in these conflicts has to be short in relation to the time spent in cooperation. I show that if players spend long periods of time in conflict, then it is possible for them to switch their strategy to something that avoids conflict.

I first consider the case where players select automata with no more than two states. In this case, the set of equilibrium strategies is small. For a class of infinitely repeated prisoner's dilemma games, there are at most two types of equilibria when signal accuracy is less than one (Theorem 4.3). In the first type of equilibrium strategy, a fixed sequence of actions is played regardless of the action of the opponent. The other type of equilibrium strategy follows the simple heuristic: if the other player cooperates, continue playing the same action; if the other player defects, switch actions. This simple strategy, introduced in the theoretical biology literature, has been coined "win-stay, lose-shift" (WSLS) (Nowak and Sigmund 1993). If both players play WSLS, then high levels of cooperation are attained. WSLS is special because it is forgiving, and allows for quick recoordination after cooperation breaks down. I also give sufficient conditions on stage-game payoffs which guarantee that both players playing WSLS is an equilibrium when signal accuracy is sufficiently close to one (Theorem 4.4). These sufficient conditions hold for a large class of 2-by-2 games, suggesting that WSLS is a useful strategy in a wide variety of settings. Finally, experiments run by Wedekind and

Milinski (1996) using human subjects suggest that WSLS is played in repeated prisoner's dilemma games. When players are limited to two-state automata, the number of outcomes is small, and the predictions are supported by experimental evidence.

Next, I examine a more general model in which players' strategies may be any finite-state automaton. In this case, I give necessary and sufficient conditions for the structure of equilibrium strategies when signal accuracy is close to one (Theorems 5.6 and 5.8). These conditions formalize the insight that players must spend almost all the time cooperating. To prove these conditions, I show that if players are not cooperating most of the time, then there exist better strategies which allow players to avoid long periods of conflict and spend almost all the time cooperating. These results show that the benefits of WSLS (cooperation and recoordination) are still required in equilibrium in a more general model.

The analysis presented here is most similar to Compte and Postlewaite (2009). They examine an infinitely repeated prisoner's dilemma game where players have imperfect private monitoring and choose among two-state automata. They show that cooperation is possible for a large region of accuracy and payoff combinations. They focus on a specific class of strategies: two-state automata with fixed transitions. Players choose only the action to be played in each state. In addition they allow for a common knowledge public signal (generalized to almost public signal) which allows the players to coordinate their actions.

This results considered in this paper extend those in Compte and Postlewaite (2009) in several ways. First, the two-state results consider the entire set of two-state automata. This means players select their transition functions as well as an action to be played in each state. I find that WSLS is able to attain high levels of cooperation even when there is no public signal to aid recoordination. I also show that WSLS remains an equilibrium over a large class of two-by-two games. In addition, I examine a more general model which allows for any finite-state automata.

This paper proceeds as follows. In Section 2, I give a motivating example, which highlights the problems of imperfect monitoring. Then, in Section 3, I present the model of boundedly rational agents and define the equilibrium concept. Next I give the results of the paper. First, in Section 4, I consider the restricted case where players only choose among two-state automata. Then in Section

5, I consider the case where players can choose among any finite-state automata. In Section 6.1, I provide a brief review of some related literature. Finally, I conclude and provide extensions in Section 7.

2 Motivating Example

In 1980, Robert Axelrod invited a number of top scholars to submit programs to compete in an iterated Prisoner's Dilemma tournament. The strategy that fared best was tit-for-tat, which simply repeats the play of the opponent in the previous round (Axelrod 1980a, Axelrod 1980b). In later work, Axelrod suggested that players may not perfectly perceive their opponents actions. To further examine the effect of misperceptions, he ran simulations where players had a 1 percent chance of seeing the incorrect action of their opponent. Not surprisingly, he found that these misperceptions led to lower levels of cooperation. However, tit-for-tat was still the dominant strategy in the tournament. Axelrod notes,

> "[TIT FOR TAT] got into a lot of trouble when a single misunderstanding led to a long echo of alternating retaliations, it could often end the echo with another misperception. Many other rules were less forgiving, so that once they got into trouble, they less often got out of it. TIT FOR TAT did well in the face of misperception of the past because it could readily forgive and thereby have a chance to reestablish mutual cooperation."
>
> -Axelrod (2006)

This excerpt captures one of the main insights of this paper: Players do not want to play strategies which get stuck in suboptimal periods. Axelrod states that tit-for-tat was successful because it was forgiving enough to be able to avoid these suboptimal periods more than most strategies. However, the following example shows that these suboptimal periods can be detrimental to payoffs, even when both players play tit-for-tat.

Consider two players playing the infinitely repeated prisoner's dilemma game displayed in Figure 1. Each player plays the tit-for-tat strategy. Each player starts by cooperating, and then repeats

	C	D
C	1,1	-L,1+L
D	1+L,-L	0,0

Figure 1: Game PD

their opponents play from the previous round. If players' signals are perfect, they continue to play C throughout the remainder of the repeated game. Based on the payoff table, this leads to an average payoff of 1 per round.

Now, suppose that players receive an imperfect signal about their opponents action. The players start by cooperating. They continue to cooperate as long as the signals are correct. Suppose that in round r, both players play C, but player 1 gets an incorrect signal that player 2 played D. In round $r+1$ player 1 plays D because of the incorrect signal, and player 2 continues to play C. If both players correctly receive correct signals in round $r+2$, then player 1 plays C and player 2 plays D. The players continue to "echo" each others action until another incorrect signal is received. While stuck in the period of alternations, the average payoff for each player is $1/2$, lower than the payoff when cooperating.

If during this period of alternations, one player receives a signal that C was played when actually D was played, then both players perceive the actions as C, and hence both cooperate in the following round. This cooperation continues until another incorrect signal is received. However, if one of the player's receives a signal that D was played when actually C was played, both players perceive action D, and both with defect in the following round. This mutual defection continues until at least one player receives an incorrect signal. The average payoff per round when both players are defecting is 0.

When both players play tit-for-tat, there are three periods the system can get stuck in: always play C, echo alternations, always play D. The only way to get out of one of these periods is if one of the players receives an erroneous signal. Suppose the signal is correct with probability $1-\varepsilon$ and incorrect with probability ε. Over the course of the infinitely repeated game, for all $\varepsilon > 0$, the frequency of time spent in the cooperate and defect periods is $1/4$ and the alternating period is $1/2$. Therefore, the frequency of each of the four possible action combinations is equal in the infinitely repeated game. So each player gets average payoff $1/4 \left[u_i(C,C) + u_i(C,D), u_i(D,C) + u_i(D,D)\right] = \frac{1+(1+L)-L}{4} = \frac{1}{2}$ in every round. Both players would receive higher payoffs if they played cooperate all the time.

In Section 4 , I show that in contrast to tit-for-tat, when both players play "win-stay, lose-shift", the system does not get caught in these suboptimal periods. When an incorrect signal is received, the strategies are able to recoordinate quickly without incurring large losses. Then, in Section 5, I show that in a more general model, players still do not play strategies that get stuck in suboptimal periods in equilibrium. Before the results, I first introduce the formal model and some notation.

3 Model

Two players, $\mathcal{I} = \{1,2\}$, play the supergame G. In every round, the players play the stage game $g = \{S_1, S_2, u_1, u_2\}$. In the stage game, each player has $|S_i|$ pure strategies. The stage-game payoff function is $u_i : S_1 \times S_2 \to \mathbb{R}$. The stage-game payoffs for player i can be represented by a payoff matrix $P_i \in \mathbb{R}^{|S_1| \times |S_2|}$. In the supergame G, the agents play stage game g for an infinite number of rounds $t = 1, 2, 3, \ldots$.

3.1 Imperfect Monitoring

After both players have their actions in round t of the supergame, each player receives a private signal which conveys the true action of their opponent with probability less than one. More formally, with probability $r_i(s_1, s_2, \varepsilon)$ player i receives a signal that the other player played action s_2 when the other player actually played s_1. The signals functions have a common rate of error, $\varepsilon \in [0, .5]$.

For example, if $S_1 = S_2 = \{C, D\}$, one possible signal function is

$$\begin{aligned} r_i(C, C, \varepsilon) = r_i(D, D, \varepsilon) = 1 - \varepsilon \\ r_i(C, D, \varepsilon) = r_i(D, C, \varepsilon) = \varepsilon. \end{aligned} \tag{1}$$

In words, the signal is correct with probability $1 - \varepsilon$ and incorrect with probability ε. This signal function is referred to as the simple signal function, r_i^S, and is used for many examples and results in this paper.

3.2 Imperfect Memory

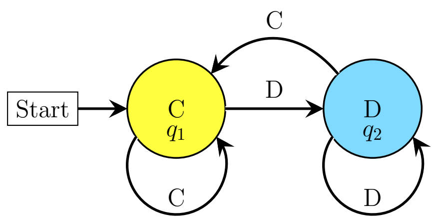
Players have bounds on their ability to differentiate between infinitely repeated game histories. Players are only able to classify this infinite set of histories into a finite number of states. This restriction yields a simple set of strategies which can be represented with finite-state automata.

A finite automaton is defined as a quadruple, $M = \left(Q_i, q_i^0, f_i, \tau_i\right)$. Here, Q_i is the finite set of states for player i and q_i^0 is the initial state. In each state, the automaton prescribes a pure action, which is determined by the output function $f_i : Q_i \to S_i$. Finally, the transition function determines which state to transition to based on the current state and the action of the other player, $\tau_i : Q_i \times S_{-i} \to Q_i$. Since the output function depends on S_i and the transition function depends on S_{-i}, if players have different action sets, then each player selects from a different set of automata. The set of all finite automata for player i is denoted by $\mathcal{M}_i$.

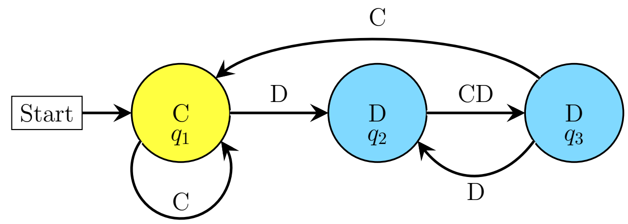
At the beginning of the supergame, each player chooses a finite-state automaton. After each history, this automaton is in a certain state, and plays the action corresponding to that state. So a finite automaton prescribes a stage-game action for every possible outcome.

Examples of Automata

Some example automata are displayed in Figure 2. The first automaton represents the tit-for-tat strategy. There are two states; the automaton cooperates in the first state, and defects in the second state. When the other player plays C, this automata goes to the first state, and hence this player cooperates. When the other player plays D, this automata leads to the second state,



(a) Tit-For-Tat



(b) Punish Twice

Figure 2: Examples of Automata

so this player defects. The second automaton represents a punish twice strategy. The first state of this automaton is a cooperative state. The automaton remains in this state as long as the other automaton is playing C. When the other automaton defects, this automaton goes into a two-state punishment phase. In the first state of the punishment phase, the automaton plays D and regardless of the other play goes to the third state. In the third state, the automaton plays D, and returns to the cooperative state only if the other automaton plays C. More complex strategies, such as N-period action sampling (Selten and Chmura 2008) can also be represented with a finite automaton.

3.3 Payoffs and Equilibria

When choosing automata, the players try to maximize the non-discounted limit of means. For a given pair of signal functions, the payoff is determined by the choice of automata from each player, and the level of error in signal function ε, $U_i : \mathcal{M}_1 \times \mathcal{M}_2 \times [0, .5] \to \mathbb{R}$. Given the signal function, error level, and automata, there is some infinite sequence of realized joint actions, $x^0, x^1, \ldots$. The players payoff is the average payoff per round over this infinite sequence of joint actions.

$$U_i\left(M_1, M_2, \varepsilon\right) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} u_i\left(x^t\right),$$

where $u_i\left(x^t\right)$ is the payoff for player i when joint action x^t is played.

In this paper I assume non-discounted payoffs.[3] This allows me to focus on long run equilibrium

[3] Here I assume that players payoff is determined by the limit of means. Limit of means can sometime be problematic because the limit may cease to exist in some cases, leading to an incomplete preference order. This however is not a problem here as displayed in Lemma 5.3.

rules-of-thumb rather than strategies where players deviate in the beginning because they are impatient.

Definition 3.1 (Best Response)**.** *Player i's best response function* $BR_i : \mathcal{M}_{-i} \times [0, .5] \to \mathcal{M}_i$

$$U_i\left(BR_i(M,\varepsilon), M, \varepsilon\right) \geq U_i\left(M', M, \varepsilon\right) \text{ for all } M' \in \mathcal{M}.$$

Definition 3.2 (Nash Equilibrium)**.** *For fixed signal functions* r_i *and error level* ε*, a pair of automata,* (M_1, M_2)*, is an equilibrium of the supergame* G *if and only if* $M_i = BR_i\left(M_{-i}, \varepsilon\right)$ *for* $i = 1, 2$*.*

A Nash equilibrium pair of automata is referred to as an equilibrium.

4 Two-State Automata

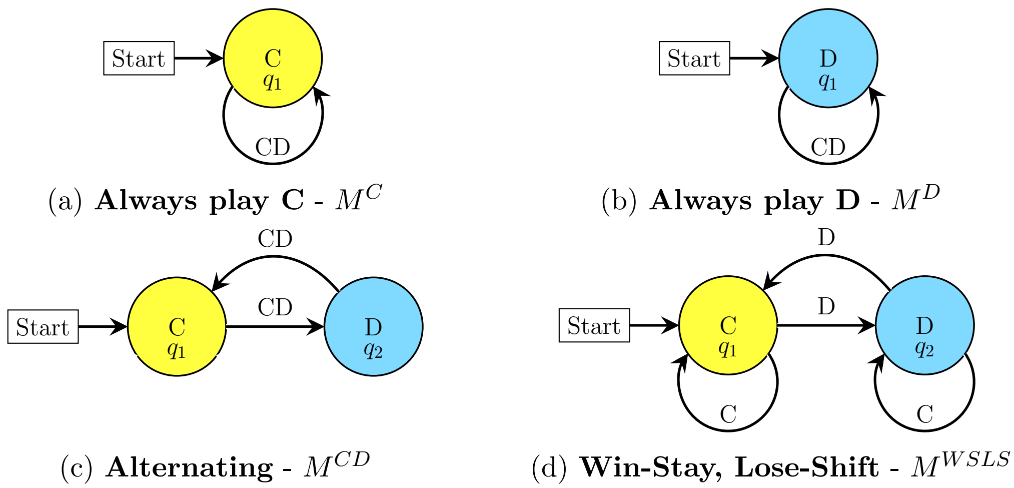
In this section I analyze the set of equilibria when players strategies are restricted to two-state automata. First, I introduce some important automata. I then show that for a class of infinitely repeated prisoner's dilemma games, there are at most two types of equilibria for any parameter pair. I then give sufficient conditions on stage-game strategies that ensure that WSLS is an equilibrium for all small error levels. Finally, I discuss some previous work done on WSLS, including some experiments from Wedekind and Milinski (1996) which show that human subjects play these strategies in the laboratory.

4.1 Important Two-State Automata

The restricted set of automata, $\mathcal{M}^2$, consists of only two-state automata. For notational simplicity, automata are represented by a tuple with the starting points omitted,

$$M = \left(\left\{f\left(q_1\right), \ldots, f\left(q_n\right)\right\}, \left\{\tau\left(q_1, C\right), \ldots, \tau\left(q_n, C\right)\right\}, \left\{\tau\left(q_2, D\right), \ldots, \tau\left(q_n, D\right)\right\}\right).$$

The starting points are mentioned when relevant. Before giving a characterization of the two-state equilibria, I first need to introduce some automata.



(a) **Always play C** - M^C (b) **Always play D** - M^D

(c) **Alternating** - M^{CD} (d) **Win-Stay, Lose-Shift** - M^{WSLS}

Figure 3: Important Two-State Automata

- **Always play C** - $M^C = (\{C\}, \{q_1\}, \{q_1\})$
- **Always play D** - $M^D = (\{D\}, \{q_1\}, \{q_1\})$
- **Alternating** - $M^{CD} = (\{C, D\}, \{q_2, q_1\}, \{q_2, q_1\})$
- **Win-Stay, Lose-Shift** - $M^{WSLS} = (\{C, D\}, \{q_1, q_2\} \{q_2, q_1\})$

Automata "always play C" and "always play D" play the same action regardless of the signal. The alternating automaton always alternates between C and D regardless of the signal. The "win-stay, lose-shift" automaton follows the simple rule: if I get a signal that you cooperated, then I play the same action; if I get a signal that you defected, I switch actions. These automata are displayed in Figure 3.

It is important to point out the differences between M^{WSLS}, and the well studied tit-for-tat automaton $M^{TFT} = (\{C, D\}, \{q_1, q_1\}, \{q_2, q_2\})$. Both automata yield output C in state q_1 and D in state q_2. The transitions from the first state of these two automata are the same as well. Both remain in the state if the signal is C, and change if the signal is D. The difference occurs in the second state transitions. The second state of M^{TFT} returns with a D and changes with a C while the second state of M^{WSLS} returns with a C and switches with a D.

4.2 Characterization of Equilibria

I give a characterization of the equilibria when players face stage-game payoffs presented in Figure 1. This game is a prisoner's dilemma when $L > 0$ with unique Nash equilibrium (D, D). When $L < 0$, the unique Nash equilibrium is (C, C), and it is no longer a prisoner's dilemma game.

I am interested in equilibria which are not heavily tied to the parameters of the game. I focus on robust equilibria. Say that $G(P_1, P_2)$ is the supergame where player i is subject to payoff matrix $P_i \in \mathbb{R}^{|S_1| \times |S_2|}$.

Definition 4.1 (Robust Equilibrium)**.** *Suppose two players play supergame $G(P_1, P_2)$ and have fixed signal functions r_i and error level ε. A pair of automata, (M_1, M_2), is a robust equilibrium of the supergame $G(P_1, P_2)$ if and only if there exists some $\mu > 0$ such that (M_1, M_2) is an equilibrium of all supergames $G(P_1', P_2')$ such that $\max_{s_i \in S_i, s_{-i} \in S_{-i}} |P_i'(s_i, s_{-i}) - P_i(s_i, s_{-i})| < \mu$.*

This equilibrium concept is a refinement of the Nash equilibrium concept defined in Definition 3.2. So every robust equilibrium is also a Nash equilibrium. The types of Nash equilibria that are not robust are only equilibria for a set of measure zero in the payoff space, and are therefore heavily tied to the parameters of the game. Robust equilibria are more universal than non-robust equilibria because they hold for a larger class of games. Therefore, they remain equilibria under small changes in the parameters. In the infinitely repeated PD game, there are at most 2 types of robust equilibria at any parameter pair.

Definition 4.2 (Payoff Equivalent Automata)**.** *Automata M_1 and M_2 are said to be payoff equivalent over set $\mathcal{M}$ if and only if,*

$$U_i(M_1, A, \varepsilon) = U_i(M_2, A, \varepsilon) \text{ for all } A \in \mathcal{M}, \text{ and all } \varepsilon \in (0, .5].$$

Automata M_1 and M_2 are said to be payoff equivalent only if they yield the same payoff against any other automata.

Theorem 4.3. *In the infinitely repeated PD game, when players have the simple signal function r_i^S and choose among the set of two-state automata, $\mathcal{M}^2$, there are only three types of robust equilibria:*

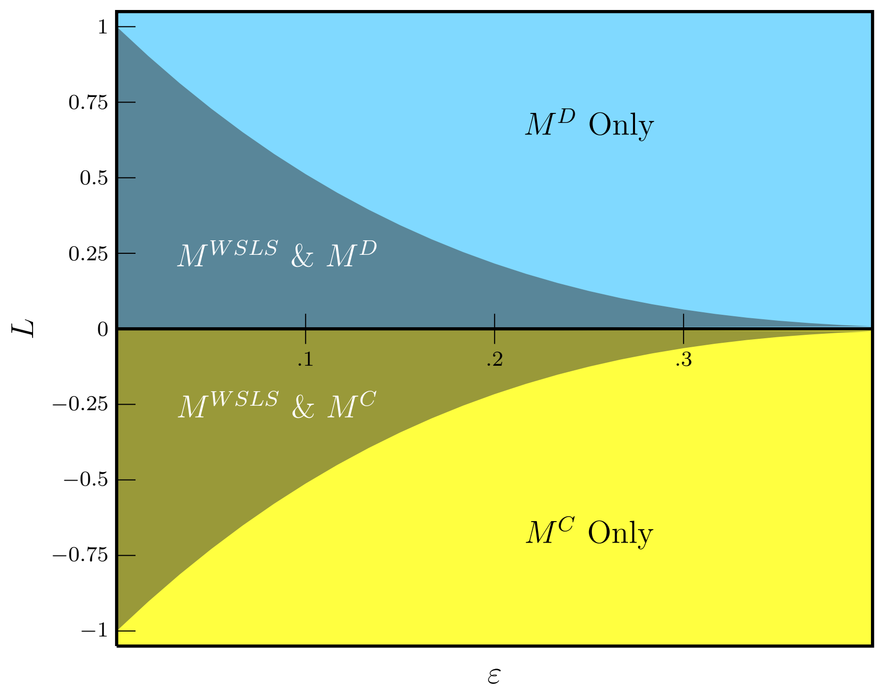


Figure 4: Equilibrium Regions from Theorem 4.3.

1. *$L < 0$ and M_i is payoff equivalent to M^C for $i = 1, 2$.*

2. *$L > 0$ and M_i is payoff equivalent to M^D for $i = 1, 2$.*

3. *$-(1 - 2\varepsilon)^3 < L < (1 - 2\varepsilon)^3$ and $M_i = M^{WSLS}$ for $i = 1, 2$.*

The proof of this result is left to the appendix. The equilibrium regions are displayed in Figure 4. When both players play M^{WSLS}, high levels of cooperation occur in equilibrium. The dark shaded region in Figure 4 represents the area where both players playing M^{WSLS} is an equilibrium. Whenever the payoff parameter and error level is in this region, then both players playing M^{WSLS} is an equilibrium, and therefore high levels of cooperation are attainable in equilibrium.

When players play M^C or M^D, their strategies are unresponsive to the signals they receive. The only robust equilibrium where players play strategies are responsive to their signals is when both players play M^{WSLS}. What makes M^{WSLS} so special? When players are trying to cooperate, they must punish their opponents to deter deviations. When an incorrect signal is received, they may start to punish each other repeatedly. In order to sustain cooperation, they must somehow recoordinate their actions to start cooperating again after an incorrect signal has been received. Since recoordination is typically inefficient, players want to recoordinate as quickly as possible after an incorrect signal is received. If both players play M^{WSLS}, this recoordination is efficient. Consider the following example which describes how this recoordination occurs. Suppose both players are playing M^{WSLS}, the sequence of states and signals is displayed in Table 1.

- **First Round** - Both players start in state q_1, both play C, both receive correct signals, and both transition back to state q_1 in round 2.

- **Second Round** - Both players again play C. Now, player 1 receives an incorrect signal that player 2 played D (error denoted with box around signal). Player 1 thinks player 2 played D and therefore moves to q_2. Player 1 received a correct signal, and therefore returns to state q_1.

- **Third Round** - Player 1 plays D and Player 2 plays C. Both receive correct signals. Player 1 sees that player 2 played C, so player 1 remains in q_2. Player 2 sees that player 1 played D, so player 2 switches to q_2.

- **Fourth Round** - Both players are in state q_2, both play D, both receive correct signals, and both switch states, and transition back to state q_1.

- **Fifth Round** - Both players are back in q_1, and continue to cooperate until another incorrect signal is received.

Round	1	2	3	4	5
Current State of M_1	q_1	q_1	q_2	q_2	q_1
Signal from M_2	C	$\boxed{\text{D}}$	C	D	C
Signal from M_1	C	C	D	D	C
Current State of M_2	q_1	q_1	q_1	q_2	q_1

Table 1: Recoordination of M^{WSLS}.

So in this example, the players start by cooperating in round one. The incorrect signal was received in round two. Player 2 plays D to confirm that an incorrect signal was received in round three. Both players play D in round four. In round five, both players are cooperating again. So after the incorrect signal is received, it only takes two rounds to recoordinate if no other incorrect signals are received. This efficient recoordination is one of the reasons why M^{WSLS} is an equilibrium strategy.

Another reason why M^{WSLS} is special is because it is not dominated by M^C or M^D for large regions of the parameter space. When players play $\left(M^{WSLS}, M^{WSLS}\right)$, the action pair (C, C) is played most of the time, so the players receive close to the cooperative payoff. In the system $\left(M^{WSLS}, M^C\right)$, the action pairs (C, C) and (D, C) are each played half the time. This is bad for player 2, because $u_2(D, C) = -L < u_2(C, C) = 1$ when $L > -1$. Playing M^C is only good for player 2 when L is sufficiently negative. In the system $\left(M^{WSLS}, M^D\right)$, action pairs (C, D) and (D, D) are each played half the time. Again this is not good for player 2 because $\frac{u_2(C,D)+u_2(D,D)}{2} = \frac{1+L}{2} \leq u_2(C, C)$ when $L < 1$. Playing M^D is only profitable for player 2 if L is sufficiently high. For medium ranges of L, M^{WSLS} is the best response to itself, because it receives the cooperative payoff most of the time.

This results does not depend on the prisoner's dilemma game. Similar results hold for a class of battle-of-the-sexes games as well as a class of minimum-effort coordination games. In both of these cases, the only types of equilibria either are unresponsive to the signal of the other players action, or similar to M^{WSLS}. These results (Theorems B.18 and B.19) are left to the appendix.

4.3 General Two-by-Two Games

In the previous section, I showed that both players playing M^{WSLS} is an equilibrium for a large region of the parameter space when players play an infinitely repeated prisoner's dilemma game. In this section, I give conditions on stage-game payoffs, which ensure that $\left(M^{WSLS}, M^{WSLS}\right)$ is an equilibrium.

Theorem 4.4. *Suppose both players have simple signal functions r_i^S. If for $i = 1, 2$,*

1. *$u_i(C, C) > u_i(C, D)$, and*

2. $u_i(C,C) > \frac{u_i(D,C)+u_i(D,D)}{2}$*;*

then there exists some $\bar{\varepsilon} > 0$ *such that* $\left(M^{WSLS}, M^{WSLS}\right)$ *is an equilibrium for all* $\varepsilon \in (0, \bar{\varepsilon})$*.*

This results suggests that when errors are small $\left(M^{WSLS}, M^{WSLS}\right)$ is an equilibrium for a wide range of games. Figure 5 displays four 2-by-2 that satisfy the desired properties.

- Figure 5(a) is a stag-hunt game with Pareto ranked pure strategy Nash equilibria (C,C) and (D,D). Both players playing M^{WSLS} leads to high levels of the Pareto superior equilibrium.

- Figure 5(b) is a chicken game with two pure strategy Nash equilibria (C,D) and (D,C), one preferred by each player. If both play M^{WSLS}, then the cooperative outcome (C,C) is possible, even though it is not one of the pure strategy Nash equilibria.

- Figure 5(c) is a battle of the sexes game with two pure strategy Nash equilibria (C,C) and (D,D). If both players play M^{WSLS} then the outcome (C,C) is frequently attained. Also consider, $M^{LSWS} = (\{C,D\}, \{2,1\}, \{1,2\})$. This strategy is the opposite of M^{WSLS} in that is stays in the same state when the other player plays D, and switches when the other player plays C. The theorem also confirms that both players playing M^{LSWS} is also an equilibrium in this case.

- Figure 5(d) is a game with no pure strategy equilibrium. However, both players playing M^{WSLS} leads to high levels of (C,C) in equilibrium.

So the simple strategy M^{WSLS} is an equilibrium for a variety of 2-by-2 games when errors are small.

4.4 Other Support

The strategy represented by automaton M^{WSLS} has been studied before. The majority of work done on this strategy focuses on biological applications. Nowak and Sigmund (1993) run evolutionary simulations on probabilistic memory one strategies. Memory one strategies are those which only respond to the previous period of play, similar to the two-state case. Their simulations are more

	C	D
C	4,4	1,3
D	3,1	3,3

(a) Stag-Hunt

	C	D
C	4,4	2,5
D	5,2	0,0

(b) Chicken

	C	D
C	4,3	0,0
D	0,0	3,4

(c) Battle of the Sexes

	C	D
C	4,3	2,4
D	1,2	3,1

(d) No Pure Equilibrium

Figure 5: 2x2 Games

general than my two-state result because they allow for probabilistic transitions. Nevertheless, the prevailing strategy in their simulation is the deterministic M^{WSLS} strategy.

More recently, Imhof, Fudenberg, and Nowak (2007) use stochastic evolutionary game dynamics to study the evolution of four strategies, $M^C, M^D, M^{TFT}, M^{WSLS}$. When only M^C, M^D, and M^{WSLS} are considered, they find some payoff threshold which determines which strategy is selected. Below this threshold M^D is selected while above this threshold M^{WSLS} is selected. When M^{TFT} is added to the three other strategies, they again find a threshold, but this time it is lower, meaning that M^{TFT} strengthens M^{WSLS}.

The prediction from my two-state model is that the only equilibria in the infinitely repeated prisoners dilemma game ($L > 0$) are M^D or M^{WSLS}. Experiments with human subjects playing repeated prisoners dilemma games have often tried to identify subjects playing tit-for-tat (Dal Bó and Fréchette 2008). Tit-for-tat typically fits the data well. One of the reasons why tit-for-tat fits the data well is that human subjects tend to always play C or always play D, both of which are supported by tit-for-tat. The predictions of my model also support this behavior. However, there is one key difference between M^{TFT} and M^{WSLS} or M^D that allows us to identify which strategies the subjects are playing.

To identify a strategy, look at the play of both players in round t, and then see the responses in $t+1$. If player 1 is playing M^{TFT} and both players play C in round t, then player 1 should play C in round $t+1$. M^{WSLS} provides the same prediction, that both players playing C leads to player 1 playing C. M^{TFT} and M^{WSLS} again share a prediction if player 1 plays C and player 2 plays D in round t. Both predict that D will be played in round $t+1$. If both players play D in round t, then M^{TFT} and M^D both predict that D is played in the following round. So far the predictions of M^{TFT} have matched the prediction of M^{WSLS} or M^D. The final combination is where they differ. If player 1 plays D and player 2 plays C in round t, then M^{TFT} predicts that player 1 will play C in the next round. Conversely, both M^D and M^{WSLS} predict that player 1 continues to play D in the next round. This provides a testable prediction: if player 1 plays D and player 2 plays C, then player 1 will play C in the next round if he is using tit-for-tat, and will play D in the next round if he is using M^{WSLS} or M^D.

Wedekind and Milinski (1996) run experiments that examine whether players play M^{TFT} or M^{WSLS}. They find that 70% of players can be classified a playing M^{WSLS} in a variety of treatments of repeated prisoner's dilemma game. Their experiments use psuedoplayers which use predetermined strategies. This allows them to focus on the situation of interest. To classify the strategies of players, they focus on the situation where player 1 plays D and player 2 plays C in round t. If player 1 plays C more in round $t+1$, then he is classified as playing M^{TFT}. If player 1 plays D more in round $t+1$, then he is classified as playing M^{WSLS}. These experimental results suggest that players are playing M^{WSLS}.

5 Unrestricted Automata

In this section, I examine the case where players can select any automata with any finite number of states to represent their strategies. I first introduce absorbing classes and communicating classes which are necessary to understand the equilibrium characterization. Then, I reanalyze the motivating example which shows the importance of absorbing and communicating classes. Finally, I give the necessary and sufficient conditions on equilibrium structure for small error levels.

5.1 Absorbing Classes

An absorbing class of an automaton M is a region that the automaton can get stuck in when the opponent plays a sequence of actions repeatedly. If player 1 plays automaton M_1, then player 2 wants to play a strategy which ensures that M_1 gets stuck in a high payoff absorbing class rather than a low payoff absorbing class. Formally,

Definition 5.1 (Absorbing Class)**.** *Given automaton $M = (Q, q^0, f, \tau)$, an absorbing class, denoted by $a(M) = \{\boldsymbol{q}, \boldsymbol{s}\}$, where $\boldsymbol{q} = q_1, \ldots, q_n$ is a sequence of states, and $\boldsymbol{s} = s_1, \ldots, s_n$ is a sequence of signals, such that*

$$\tau(q_k, s_k) = \begin{cases} q_{k+1} & k < n \\ q_1 & k = n. \end{cases}$$

So when automaton M is in state q_1, and the opponent plays sequence $\boldsymbol{s}$ repeatedly, then M will loop through the sequence of states $\boldsymbol{q}$ repeatedly. The length of an absorbing class is the length of the sequences of actions and states, $|a| = n$. When automaton M is looping through states $\boldsymbol{q}$ and the opponent is playing actions $\boldsymbol{s}$, then the players are playing a fixed sequence of joint actions repeatedly. The payoff for an absorbing class is the average payoff per round while in this absorbing class,

$$U_i^{AC}(a) = \frac{1}{|a|} \sum_{k=1}^{|a|} u_i\left(f\left(q_k\right), s_k\right).$$

The set of all possible absorbing classes for automaton $M = \left(Q, q^0, f, \tau\right)$ is infinite. However there exists a payoff-maximal absorbing class for player i, denoted by $a_i^*(M)$, with $|a_i^*(M)| \leq |Q|$. This result, presented in Lemmas B.2 and B.3, is left to the appendix. The idea for the proof is that if a payoff-optimal absorbing class travels through the same state twice, then there must be a smaller payoff-optimal absorbing class. Therefore, given any payoff-optimal absorbing class, the length can be reduced by eliminating states that appear more than once, until it has length less than or equal to $|Q|$. This finite length optimal absorbing class is used to construct a best response automaton. In equilibrium, each player must spend almost all of the repeated game in the optimal absorbing class of their opponents automaton.

5.2 Communicating Class

Once both players have selected automata $M_1 = \left(Q_1, q_1^0, f_1, \tau_1\right)$ and $M_2 = \left(Q_2, q_2^0, f_2, \tau_2\right)$, the pair of automata (M_1, M_2) forms a *system* which can be represented with a finite Markov chain $X\left(M_1, M_2, \varepsilon\right)$. Each state of the Markov chain corresponds to a pair of automaton-states, one from each automaton. For example, the situation where M_1 is in state q_1 and M_2 is in state q_2 is represented by one state of the Markov chain. The starting state of the Markov chain represents the situation where both automata are in their initial states, M_1 in q_1^0 and M_2 in q_2^0. Based on the signal functions r_i, the Markov chain has $n \leq |Q_1|\,|Q_2|$ states, one corresponding to each pair of automaton states that are reachable from the initial states with positive probability for any $\varepsilon > 0$. These states are denoted by $x_1, \ldots, x_n$.

Let $q_i(x)$ be the current state of automaton M_i when the Markov chain is in state x. Automaton M_i moves from state $q_i(x_a)$ to $q_i(x_b)$ with probability,

$$\mathbb{P}(M_i, q_i(x_a), q_i(x_b), \varepsilon) = \sum_{s_i | \tau(q_i(x_a), s_i) = q_i(x_b)} r_i(s_i, f_{-i}(q_{-i}(x_a)), \varepsilon).$$

In words, the term inside the sum is the probability that player i receives a signal that the other player played action s_i when the other player actually played action $f_{-i}(q_{-i}(x_a))$. This term is then summed over all actions s_i which take automaton M_i from state $q_i(x_a)$ to $q_i(x_b)$. The Markov chain is therefore defined by the probability that M_1 moves from $q_1(x_a)$ to $q_1(x_b)$ and M_2 moves from $q_2(x_a)$ to $q_2(x_b)$,

$$X(M_1, M_2, \varepsilon)(x_a, x_b) = \mathbb{P}(M_1, q_1(x_a), q_1(x_b), \varepsilon)\, \mathbb{P}(M_2, q_2(x_a), q_2(x_b), \varepsilon). \tag{2}$$

The starting point of this Markov chain is state x^0 such that $q_1(x^0) = q_1^0$ and $q_2(x^0) = q_2^0$. When the signals are perfect, the Markov chain $X(M_1, M_2, 0)$ is deterministic. Each state leads to another state with probability 1. When the signals are imperfect, the Markov chain $X(M_1, M_2, \varepsilon)$ is not necessarily deterministic and any state may lead to multiple different states with varying probabilities.The realizations of the Markov chain are denoted by $x^1, x^2, \ldots$.

Definition 5.2 (Communicating Class). *A communicating class of the system $X(M_1, M_2, \varepsilon)$ is a set of states $A \subseteq X(M_1, M_2, \varepsilon)$ that satisfies,*

- $(X(M_1, M_2, 0)(x, y))^n = 0$ *for all* $x \in A, y \notin A, n > 0$.
- $(X(M_1, M_2, 0)(x, y))^n > 0$ *for all* $x, y \in A$ *and for some* $n > 0$.

A communicating class is a set of absorbing states. Once the Markov chain enters a communicating class, it can only leave if a player receives an incorrect signal. When no erroneous signals are received, the Markov chain deterministically loops through the states in the communicating class.

The payoff of a communicating class is defined to be the average payoff over this loop,

$$U_i^{CC}\left(A_k\right) = \frac{1}{|A_k|} \sum_{x \in A_k} u_i(x), \tag{3}$$

where $u_i\left(x\right) = u_i\left(f_1\left(q_1\left(x\right)\right), f_2\left(q_2\left(x\right)\right)\right)$ is the payoff for player i in state x. This definition gives the average payoff in the communicating class when signals are correct, and is used when giving necessary and sufficient conditions in the finite-state case.

5.3 Calculating Payoffs

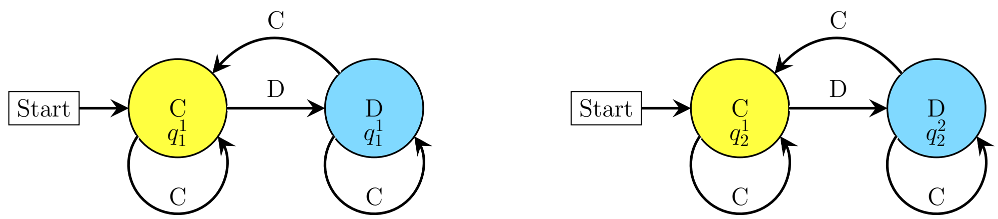
Representing the system as a Markov chain allows me to calculate the payoffs for a given pair of automata using only the stationary distribution of the Markov chain. By Lemma B.1, the Markov chain $X\left(M_1, M_2, \varepsilon\right)$ is irreducible for all $\varepsilon > 0$, and hence has a unique stationary distribution, $\pi\left(M_1, M_2, \varepsilon\right)$.

Lemma 5.3. *Suppose players play automata M_1 and M_2. The average payoff for the infinitely repeated game is equal to,*

$$U_i\left(M_1, M_2, \varepsilon\right) = \sum_{x_k \in X(M_1, M_2, \varepsilon)} \pi\left(M_1, M_2, \varepsilon\right)\left(x_k\right) u_i\left(x_k\right),$$

where $\pi\left(M_1, M_2, \varepsilon\right)\left(x_k\right)$ is the term of the stationary distribution corresponding to state x_k, and $u_i\left(x_k\right)$ is the payoff for player i state x_k.

Lemma 5.3 implies that only the stationary distribution of the system and vector of utilities for the corresponding states are needed to find the limit of means for a pair of automata. The idea behind the proof is that the frequency of time the Markov chain spends in a state converges to the stationary distribution by the law of large numbers. The proof of this lemma is left to the appendix.



(a) Player 1 plays M^{TFT} (b) Player 2 plays M_2^{TFT}

Figure 6: Automata for absorbing class example.

5.4 Tit-For-Tat Absorbing Class Example

To better understand the absorbing classes, communicating classes, and where the analysis is headed, it helps to give an example. Assume the players are playing the PD-game presented in Figure 1. As the motivating example shows, tit-for-tat can get into trouble by getting caught in a suboptimal region when signals are imperfect. Here, I elaborate on the motivating example using the notation introduced above. This example helps to show why these suboptimal regions are a problem, and how to make automata more robust to these imperfect signals.

Suppose player 1 plays automaton M^{TFT} displayed in Figure 6(a). This is a two-state automaton with states q_1 and q_2 which represents the tit-for-tat strategy. The optimal absorbing class of M^{TFT} for player 2 is,

$$a_2^*\left(M^{TFT}\right) = \left\{\left\{q_1\right\}, \left\{C\right\}\right\}.$$

In this absorbing class, M^{TFT} is in state q_1, player 2 plays C, and M_1 returns to q_1. Therefore, the average payoff in this absorbing class for player 2 is $U_2^{AC}(a_2^*(M^{TFT})) = 1$.

Consider two other absorbing classes,

$$a^{CD}(M^{TFT}) = \left\{\left\{q_1, q_2\right\}, \left\{D, C\right\}\right\} \text{ and } a^{D}(M^{TFT}) = \left\{\left\{q_2\right\}, \left\{D\right\}\right\}.$$

The respective payoffs are $U_2^{AC}(a^{CD}) = 1/2$ and $U_2^{AC}\left(a^D\right) = 0$.

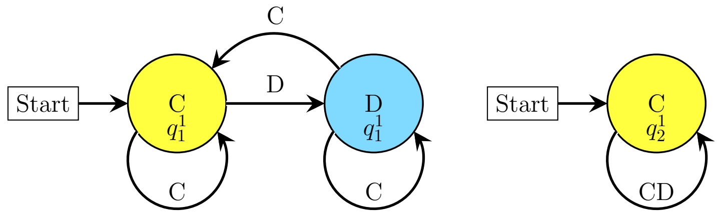
Ideally, player 2 would like to play an automaton which spends most of the supergame in absorbing class $a_2^*(M^{TFT})$, and does not get stuck in $a^{CD}(M^{TFT})$ or $a^D(M^{TFT})$. When an incorrect signal is received, the best response automaton should be able to find its way back to the payoff-optimal absorbing class without getting stuck in a suboptimal absorbing class.

Now, suppose that player 2 also plays M^{TFT} with states q_1 and q_2 as displayed in Figure 6(b). Suppose each player has the simple signal function r_i^S from (1). The Markov chain for the system, $X\left(M^{TFT}, M^{TFT}, \varepsilon\right)$, has states $\left\{x^{CC}, x^{CD}, x^{DC}, x^{DD}\right\} = \{q_1q_1, q_1q_2, q_2q_1, q_2q_2\}$,

$$X\left(M^{TFT}, M^{TFT}, \varepsilon\right) = \begin{array}{c} x^{CC} \\ x^{CD} \\ x^{DC} \\ x^{DD} \end{array} \begin{bmatrix} (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & \varepsilon^2 \\ \varepsilon(1-\varepsilon) & \varepsilon^2 & (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) \\ \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 & \varepsilon^2 & \varepsilon(1-\varepsilon) \\ \varepsilon^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 \end{bmatrix}.$$

When ε is small, and the system is in state x^{CC}, then it returns to x^{CC} with probability $(1-\varepsilon)^2$ (only leaves with an incorrect signal). When the system is in state x^{CD}, then it goes to state x^{DC} and vice-versa, unless an incorrect signal is received. When in x^{DD}, the system remains in x^{DD}, unless an incorrect signal is received. So this system has three communicating classes: $A^C = \left\{x^{CC}\right\}, A^{CD} = \left\{x^{CD}, X^{DC}\right\}$, and $A^D = \left\{x^{DD}\right\}$. Once the system has entered a communicating class, it remains in that class until an incorrect signal is received. Note that $U_2^{CC}\left(A^C\right) = U_2^{AC}\left(a_2^*\left(M_1\right)\right)$, so when the system is in communicating class A^C, automaton M_1 is in player 2's optimal absorbing class. So player 2 wants the system to spend as much time in communicating class A^C as possible.

Since automaton M^{TFT} starts in state q_1, the system starts in state x^{CC}. When $\varepsilon = 0$, the starting point of the system matters, and the stationary distribution is $\pi\left(M^{TFT}, M^{TFT}, 0\right) = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \end{array}\right]$. This means the system always stays in x^{CC}, and both players receive payoffs equal to their optimal absorbing class. Since neither player can do better, this is an equilibrium when $\varepsilon = 0$. For any $\varepsilon > 0$, the starting point no longer matters, and the unique stationary distribution is $\pi\left(M^{TFT}, M^{TFT}, \varepsilon\right) = \left[\begin{array}{cccc} 1/4 & 1/4 & 1/4 & 1/4 \end{array}\right]$. This means for any positive error level, the system



(a) Player 1 plays M^{TFT} (b) Player 2 plays M_2^C

Figure 7: Automata for absorbing class example.

spends one-quarter of the time in A^C and A^D, and one-half the time in A^{CD}, which yields payoff

$$U_2\left(M^{TFT}, M^{TFT}, \varepsilon\right) = \frac{1}{4}U_2^{CC}\left(A^C\right) + \frac{1}{2}U_2^{CC}\left(A^{CD}\right) + \frac{1}{4}U_2^{CC}\left(A^D\right) = \frac{1}{2} < U_2^{AC}\left(a_2^*\left(M^{TFT}\right)\right).$$

Since the system is in suboptimal absorbing classes A^{CD} or A^D three-quarters of the time, the payoff is less than the payoff of the optimal absorbing class. Player 2 spends significant time in a suboptimal absorbing class, even in the limit as the errors go to zero. Player 2 gets higher payoff playing an automaton that doesn't get caught in the suboptimal absorbing classes.

There is a discontinuity in the stationary distribution at $\varepsilon = 0$.

$$\pi\left(M^{TFT}, M^{TFT}, 0\right) = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \end{array}\right] \neq \frac{1}{4}\left[\begin{array}{cccc} 1 & 1 & 1 & 1 \end{array}\right] = \lim_{\varepsilon \to 0} \pi\left(M^{TFT}, M^{TFT}, \varepsilon\right).$$

Since the payoffs for a pair of automata depend on the stationary distribution, a discontinuity like this in the stationary distribution, also leads to a discontinuity in the payoffs,

$$U_i\left(M^{TFT}, M^{TFT}, 0\right) = 1 \neq \frac{1}{2} = \lim_{\varepsilon \to 0} U_i\left(M^{TFT}, M^{TFT}, \varepsilon\right).$$

Because of this discontinuity, some of the equilibria under perfect monitoring may fail to be equilibria when monitoring becomes imperfect. If the stationary distribution does not have this discontinuity, then neither do payoffs.

Next, suppose the player 1 still plays the tit-for-tat automaton M^{TFT}. Player 2 now plays the "always play C" automaton M^C, with state q', as displayed in Figure 7(b). Since M^C only has one state, the system only has two states $\left\{x^{CC}, x^{DC}\right\} = \{q_1 q', q_2 q'\}$. The corresponding Markov chain is,

$$X(M_1, M_2, \varepsilon) = \begin{array}{c} x^{CC} \\ x^{DC} \end{array} \left[\begin{array}{cc} 1-\varepsilon & \varepsilon \\ 1-\varepsilon & \varepsilon \end{array} \right].$$

When this system is in x^{CC}, it returns to x^{CC} unless an incorrect signal is received. State x^{DC} also leads to x^{CC} unless an incorrect signal is received. So now there is only one communicating class in this system, $A^C = \left\{x^{CC}\right\}$. There is also a transient class, $T^C = \left\{x^{DC}\right\}$, which leads to communicating class A^C. As before, the payoff of this communicating class is equal to the payoff for player 2's optimal absorbing class, $U_2^{CC}\left(A^C\right) = U_2^{AC}\left(a_2^*(M_1)\right)$. However, now there are no suboptimal absorbing classes.

When the signals are perfect, the stationary distribution of the system is $\pi(M_1, M_2, 0) = \left[\begin{array}{cc} 1 & 0 \end{array}\right]$. This means that the system spends all of the time in x^{CC}, and yields payoff $U_2\left(M^{TFT}, M^C, 0\right) = 1$. It turns out that M^C is a best response to M^{TFT} when the signals are perfect. When $\varepsilon > 0$, the stationary distribution becomes $\pi(M_1, M_2, \varepsilon) = \left[\begin{array}{cc} 1-\varepsilon & \varepsilon \end{array}\right]$, which yields payoff

$$U_2\left(M^{TFT}, M^C, \varepsilon\right) = (1-\varepsilon) U_2\left(A^C\right) + \varepsilon(-L) = 1 - \varepsilon(1+L).$$

Now player 2 spends $(1-\varepsilon)$ in the optimal absorbing class and only ε in the transient class. Player 2 never gets caught in some suboptimal absorbing class. As ε approaches zero, player 2 is almost always in the optimal absorbing class, and can't do better.

Unlike the previous case, there is no longer a discontinuity in the stationary distributions [4],

$$\pi\left(M^{TFT}, M^C, 0\right) = \left[\begin{array}{cc} 1 & 0 \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \end{array}\right] = \lim_{\varepsilon \to 0} \pi\left(M^{TFT}, M^C, \varepsilon\right).$$

Again, since the payoffs are just a linear function of the stationary distribution, continuous sta-

[4] In the literature on perturbed Markov chains, there are two types of perturbations: *regular perturbations* and *singular perturbations*. In a regular perturbation, the stationary distribution is continuous as the error increases from zero. A singular perturbation is defined as a perturbation which changes the ergodic structure of the matrix, meaning that multiple communicating classes may be combined. Singular perturbations typically do not have continuous stationary distributions when moving the errors away from zero. $X\left(M^{TFT}, M^{TFT}, \varepsilon\right)$ is a singular perturbation while $X\left(M^{TFT}, M^C, \varepsilon\right)$ is a regular perturbation (different than regular perturbation defined in this paper).

tionary distributions lead to continuous payoffs,

$$U_2\left(M^{TFT}, M^C, 0\right) = 1 = 1 = \lim_{\varepsilon \to 0} U_2\left(M^{TFT}, M^C, \varepsilon\right).$$

Since M^C was a best response with perfect signals, and the payoffs are continuous, it remains a best response in the limit as probability of getting an incorrect signal goes to zero.

To summarize, if player 1 plays M_1, then player 2 wants to play an automaton such that automaton M_1 spends most of the time in player 2's optimal absorbing class $a_2^*(M_1)$ and does not get caught in suboptimal absorbing classes. Suboptimal absorbing classes can be detrimental to payoffs, even in the limit as the probability of an incorrect signal goes to zero.

5.5 Necessary and Sufficient Conditions

In this section I provide the necessary and sufficient conditions for equilibria in the finite-state case. As the above example shows, in equilibrium, players cannot spend significant amounts of time in communicating classes that do not yield that optimal absorbing class payoff. If all communicating classes yield the same payoff as the optimal absorbing class for each player, then it is an equilibrium. However, it is possible to have communicating classes which don't yield the optimal absorbing class payoff in equilibrium as long as the time spent in these communicating classes is significantly less than time spent in other communicating classes.

To formalize these conditions, we must understand how some communicating classes are more robust to incorrect signals than others. The system may exit some communicating classes with only one incorrect signal, while others may require many more incorrect signals. The system visits those communicating classes that are most robust to incorrect signals almost all the time as probability of error goes to zero.

Definition 5.4 (Prevalent Communicating Class)**.** *A communicating class A of the matrix $X(M_1, M_2, \varepsilon)$ is a prevalent communicating class if*

$$\lim_{\varepsilon \to 0} \pi\left(M_1, M_2, \varepsilon\right)(x) > 0 \textit{ for some } x \in A.$$

A prevalent communicating class is a set of states that the Markov Chain $X(M_1, M_2, \varepsilon)$ visits with positive probability in the limit as the error goes to zero. When ε is small, the system spends almost all the time in the prevalent communicating classes.

Next, these results hold for a wider class of signal functions, as defined below,

Definition 5.5 (Regular Signal Function)**.** *A signal function $r_i : S_{-i} \times S_{-i} \times [0, .5] \to [0, 1]$ is said to be regular if the following conditions hold.*

1. $\lim\limits_{\varepsilon \to 0} r_i(s_i, s_j, \varepsilon) = \begin{cases} 1 & s_i = s_j \\ 0 & s_i \neq s_j \end{cases}$,
2. $r(s_i, s_j, \varepsilon) > 0$ *for all* $\varepsilon \in (0, .5]$ *and all* $s_i, s_j \in S_{-i}$,
3. $\exists n \geq 0$ *such that* $0 < \lim\limits_{\varepsilon \to 0} \varepsilon^{-n} r(s_i, s_j, \varepsilon) < \infty$ *for all* $s_i, s_j \in S_{-i}$.

It is clear that the simple signal function, r_i^S from (1), is a regular signal function. There are also more complex signal functions that satisfy this as well.

For the finite-state results, I restrict the set of finite automata to those which are finite, strongly connected, and reduced. This set is denoted by $\mathcal{M}_i^R$. All equilibria over this set are also equilibria over the set of all finite automata. For more details see Appendix A. With this notation, I introduce the main results for the finite-state case.

Theorem 5.6 (Necessity)**.** *Suppose players play supergame G with regular signal function r_i, and play automata $M_i \in \mathcal{M}_i^R$ represented by Markov chain $X(M_1, M_2, \varepsilon)$. If there exists some $\bar{\varepsilon} > 0$ such that (M_1, M_2) is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$, then for all prevalent communicating classes A_k, $U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i}))$.*

These conditions say that, for small error levels, each prevalent communicating class must yield the optimal absorbing class payoff for each player. Since almost all time is spent in the prevalent communicating classes when the errors are small, the system must spend almost all the time in good regions and not get caught in bad regions in equilibrium.

To prove the necessary conditions, I show that if the necessary conditions are not satisfied, then it is always possible to construct an automaton M_2' such that for some $\bar{\varepsilon} > 0$,

$$U_2(M_1, M_2, \varepsilon) < U_2\left(M_1, M_2', \varepsilon\right) \text{ for all } \varepsilon \in (0, \bar{\varepsilon}).$$

So M_2' is a better response than M_2 to automaton M_1. This means that (M_1, M_2) is not an equilibrium if the desired properties are not satisfied. I show that such an automaton M_2' exists in the following lemma.

Lemma 5.7. *Given automaton $M_1 \in \mathcal{M}^R$ with n states, and any absorbing class $a(M_1)$, there exists automaton M_2 such that for all communicating classes, A_k, of the system $X(M_1, M_2, \varepsilon)$,*

$$U_2^{CC}(A_k) = U_2^{AC}(a(M_1)).$$

The proof of the Lemma is left to the appendix. To prove this I construct automaton M_2 with the desired properties. Automaton M_2 contains three regions. The first region is called the *absorbing region*. As long as no incorrect signals are received, M_2 remains in this region when M_1 is in the desired absorbing class $a(M_1)$. When an incorrect signal is received by either player, there is a chance that automaton M_1 will leave the states of $a(M_1)$. When this happens, player 2 becomes confused about the current state of M_1, and must try to make inferences about current state. Player 2 wants to get back to the states of $a(M_1)$ without getting caught in another suboptimal absorbing class. To do this, player 2 plays what is called a *homing sequence*. This homing sequence is a fixed sequence of actions, which based on the output, determines the current state of M_1. Once the current state of M_1 has been identified, the automaton enters the *resynchronization region*. When entering this region, M_2 knows the current state of M_1. M_2 then plays a sequence of actions which resynchronizes the automata after which automaton M_1 returns to the states of the desired absorbing class, and automaton M_2 returns to the absorbing region. Automaton M_1 remains in the states of the desired absorbing class until an incorrect signal is received. To better understand the construction, I provide an example in appendix C.2.

Next, I give the sufficient conditions for the structure of equilibrium automata. Let $\mathcal{M}^{SPM}(M_i)$ be the set of all automata $M_{-i} \in \mathcal{M}_{-i}^R$ such that all prevalent communicating classes of $X(M_i, M_{-i}, \varepsilon)$, A_k, yield the optimal absorbing class payoff, $U_i^{CC}(A_k) = U_i^{AC}(a_i^*(M_{-i}))$ for $i = 1, 2$. This is the set of all automata that when paired with M_i yield the optimal absorbing class payoff in all prevalent communicating classes.

Theorem 5.8 (Sufficiency)**.** *Suppose players play supergame G with regular signal function r_i, and play automata $M_i \in \mathcal{M}_i^R$ represented by Markov chain $X(M_1, M_2, \varepsilon)$. If*

1. *for all prevalent communicating classes A_k, $U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i}))$, and*
2. $\frac{\partial U_i(M_1,M_2,0)}{\partial \varepsilon} = \sup_{M \in \mathcal{M}^{SPM}(M_{-i})} \frac{\partial U_i(M_i,M,0)}{\partial \varepsilon}$*;*

then there exists some $\bar{\varepsilon} > 0$ such that (M_1, M_2) is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$.

This theorem provides sufficient conditions for equilibrium automaton in the finite-state case when errors are sufficiently small. The first condition requires that all prevalent communicating classes yield the optimal absorbing class for both players. Since the system spends almost all the time in prevalent communicating classes and almost no time in the other states, this formalizes the intuition that the system cannot get stuck in suboptimal regions for long periods of time. The second condition requires that out of all $M \in \mathcal{M}^{SPM}(M_i)$, the player must select the one that yields the highest marginal utility at zero. The two conditions together are the necessary and sufficient for equilibrium for sufficiently small errors. The proof of the sufficient conditions is left to the appendix.

6 Discussion

6.1 Literature Review

There has been a lot of work done examining imperfect monitoring in repeated games. The different models of imperfect monitoring all share the common theme that the players must recoordinate after an error is made. When there is some common knowledge among players, recoordination is relatively easy. Models involving imperfect public monitoring (Fudenberg, Levine, and Maskin 1994) as well as models of imperfect private monitoring with communication (Compte 1998, Kandori and Matsushima 1998, Obara 2009) are able to obtain the folk theorem. This common knowledge allows for relatively easy recoordination.

When there is no common knowledge, as in the imperfect private monitoring case, coordination becomes more difficult because players are not able to condition their strategies on a common knowledge signal and therefore must make inferences about the actions of their opponents. There are two main approaches to the study of equilibria under imperfect private monitoring: the belief-based approach and the belief-free approach. In the belief based approach players must make statistical inferences about the history of actions. The inferences quickly become difficult, and therefore results here typically require signals to be highly correlated (i.e. almost public) in order to obtain meaningful results (Bhaskar and Obara 2002, Mailath and Morris 2002). In the belief-free approach, strategies are constructed to ensure that beliefs are irrelevant. The prevailing equilibrium strategies are complex, and require that players are indifferent over a set of actions (Piccione 2002, Ely, Horner, and Olszewski 2005, Yamamoto 2009). If the payoffs are perturbed slightly, then these strategies fail to remain equilibria (Bhaskar 2000). So previous models of imperfect private monitoring are heavily dependent on either payoffs or signal structure. In this paper, the players use strategies represented by finite automata to represent their strategies which allows for a simpler representation of the beliefs.

There has also been work examining repeated games when players have bounds on memory. Lehrer (1988) and Sabourian (1998) look at models where players have bounded recall and perfect monitoring, while Cole and Kocherlakota (2005) examine a model of bounded recall with imperfect public monitoring. These results typically examine the effect of memory length on possible outcomes. Others have examined models where players select finite automata as their strategies. Using finite automata to represent strategies was first suggested by Aumann (1981). Since then, applications have included looking at finitely repeated games (Neyman 1985), assuming players have some exogenous cost of complexity (more states more costly) on their strategies (Rubinstein 1986, Abreu and Rubinstein 1988), or examining the evolutionary stability of such strategies (Miller 1996, Ioannou 2009). It is important to note that not every finite automaton strategy can be represented with a bounded memory strategy, but every bounded memory strategy can be represented as an automaton (Cole and Kocherlakota 2005).

7 Conclusion

The paper started with the question: Does a model of repeated interactions with boundedly rational agents lead to a smaller set of outcomes in equilibrium? When player's are limited to two-state automata, the number of outcomes in equilibrium is small (Theorem 4.3). Importantly, in an infinitely repeated prisoner's dilemma game, high levels of cooperation are still possible in equilibrium, even when agents can't perfectly monitor their opponents and have no common knowledge public signal with which to recoordinate. The important strategy used is called "Win-Stay, Lose-Shift". If both players play this strategy, when cooperation breaks down, the players are able to quickly recoordinate and get back to cooperation without getting stuck in conflict for long periods. I show that WSLS holds for a variety of 2x2 games as well (Theorem 4.4). So when restricted to two-state automata, the number of equilibrium outcomes is small, and the predictions match the behavior of human subjects in the laboratory.

When I remove the restriction of two-state automata, the analysis becomes more difficult. In this case I am able to provide necessary and sufficient conditions on equilibrium structure for small error levels (Theorems 5.6 and 5.8). The results show that in equilibrium players must play strategies which are able to cooperate without getting stuck in long periods of conflict. However, the implications of these conditions on the set of equilibrium outcomes remains an open question.

There are many extensions for this work. First, a better understanding of the effect of the necessary and sufficient conditions on outcomes. It is possible that for even small errors and finite-state strategies, the set of outcomes could still be small compared to the folk theorem. Also there is more work to be done examining what happens for larger errors when players can use finite-state automata as their strategies. In addition, more experiments with human subjects to further verify that these strategies are actually played in the lab.

There are also some more broad extensions. Assuming that players use finite automata as their strategies is assuming that they are classifying the infinite set of repeated game histories into a finite set of groups. It would be interesting to examine more general classification systems that would allow players to have more general groupings of their histories, rather than just those that can be represented with a finite automaton. Also, this paper only focuses on the equilibrium, but there may be some learning that takes place to get to this equilibrium. If we assume that players use automata to represent their strategies, there are a number of interesting learning dynamics that the players could use to learn to play certain strategies.

Appendix

A Structure of Automata

The set of finite automata contains many automata which are redundant. It simplifies the analysis to eliminate some of these redundant automata, allowing me to focus on a smaller set of automata. Much of the notation from this section is from Kohavi (1978).

A.1 Payoff Equivalent Automata

Definition A.1 (Payoff Equivalent Automata)**.** *Automata M_1 and M_2 are said to be payoff equivalent over set $\mathcal{M}$ if and only if,*

$$U_i\left(M_1, A, \varepsilon\right) = U_i\left(M_2, A, \varepsilon\right) \text{ for all } A \in \mathcal{M}, \text{ and all } \varepsilon \in (0, .5]$$

Two automata are considered payoff equivalent over a set $\mathcal{M}$ if the they yield the same payoff when matched against any automaton from $\mathcal{M}$. For any set of payoff equivalent automata $\mathcal{M}^{PE}$, I only need to consider one automaton $M_1 \in \mathcal{M}^{PE}$ when calculating equilibria. When M_1 is not part of an equilibrium, none of the automata in $\mathcal{M}^{PE}$ are part of an equilibrium. When M_1 forms an equilibrium with M_2, then any automaton from $\mathcal{M}^{PE}$ forms an equilibrium with M_2. When computing equilibrium in my model, I can without loss of generality search over a smaller set of automata where any set of payoff equivalent automata is represented by a single automaton.

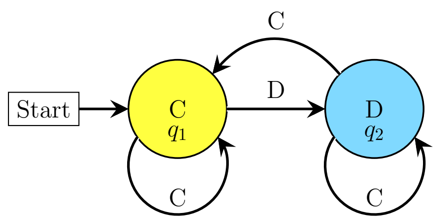
A.2 Reduced Automata

Next, I introduce the concept of a reduced automaton. Any non-reduced automaton is payoff equivalent to some reduced automaton. Therefore, I am able to only focus on the set of reduced automata without loss of generality.

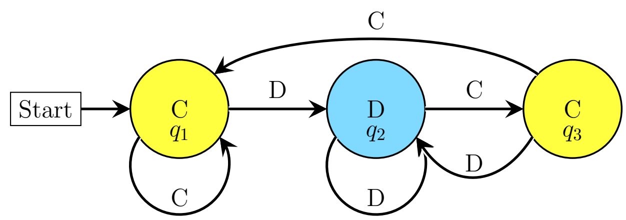
Definition A.2 (Equivalent States)**.** *States s_i and s_j are said to be equivalent if and only if, for every possible input sequence, the same output sequence is produced, regardless of whether s_i or s_j is the initial state.*

Definition A.3 (Equivalent Automata)**.** *Two automata, M_1 and M_2, are said to be equivalent if and only if, for every state in M_1, there is a corresponding equivalent state in M_2, and vice versa.*

If two automata are equal, then they must be equivalent. However, if two automata are equivalent they need not be equal. Each of the automata in Figure 8 represent the tit-for-tat strategy. Figure 8(a) is a two-state automaton which represents tit-for-tat, while Figure 8(b) is three-state automaton which represents tit-for-tat. Both q_1 and q_3 from 8(b) are equivalent to q_1 from 8(a), and state q_2 in 8(b) is equivalent to q_2 in 8(a), so these automata are equivalent but not equal.



(a) 2-state Tit-for-Tat



(b) 3-state Tit-for-Tat

Figure 8: Example of equivalent but not equal automata.

Definition A.4 (Reduced Automaton)**.** *An automaton M is reduced if and only if it contains no equivalent states.*

Every non-reduced automaton has a corresponding reduced automaton, where equivalent states are combined into a single state. The non-reduced automata and the corresponding reduced automata are payoff equivalent over the set of finite automata, because they produce the same output for all sequences of input. I am therefore able to restrict the set of automata from all finite automata to reduced automata without loss of generality.

A.3 Strongly Connected Automata

Next, I introduce the notion of a strongly connected component, an absorbing region of the automaton.

Definition A.5 (Reachable State)**.** *Given automaton $M = (Q, q^0, f, \tau)$, state $q_m \in Q$ is reachable from $q_1 \in Q$ if there exists some sequence of signals, $\boldsymbol{r} = \{r_1, \ldots, r_m\}$ such that,*

$$\tau\left(q_k, r_k\right) = q_{k+1} \textit{ for all } 1 \leq k \leq m-1$$

where states $q_2, \ldots, q_{m-1}$ are defined recursively.

Definition A.6 (Strongly Connected Subset)**.** *Given automaton $M = (Q, q^0, f, \tau)$, a subset of states $Q^{SCS} \subseteq Q$ is said to be strongly connected if for every pair of states $q_i, q_j \in Q^{SCS}$, q_i is reachable from q_j.*

Definition A.7 (Strongly Connected Component)**.** *Given automaton $M = (Q, q^0, f, \tau)$, a subset of states $Q^{SCC} \subseteq Q$ is said to be strongly connected component (SCC) if Q^{SCC} is strongly connected and there is no state $q \in Q \backslash Q^{SCC}$ such that $Q^{SCC} \cup q$ is strongly connected.*

A strongly connected component is a region of the automaton that cannot be left once it has been reached regardless of the future signal sequence. All states in a strongly connected component are reachable from all other states in the SCC. Therefore, once the automaton enters one of these SCCs, all other states of the automaton become irrelevant.

Definition A.8 (Strongly Connected Automaton)**.** *Automaton $M = (Q, q^0, f, \tau)$ strongly connected if Q is a strongly connected component.*

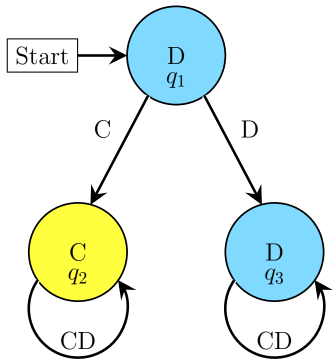


Figure 9: Non-strongly-connected automaton.

An example of an automaton that is not strongly connected is displayed in Figure 9. This automaton has three states, $Q = \{q_1, q_2, q_3\}$. It is clear by definition that this automaton has two strongly connected components; $Q_1^{SCC} = \{q_2\}$ and $Q_2^{SCC} = \{q_3\}$, and therefore is not a strongly connected automaton. The automaton starts in state q_1. If it receives a C signal in the first round, then it enters q_2 and always plays C. If it receives a D signal in the first round, then it enters q_3 and always plays D. So with certain probability this automata always plays C, otherwise it always plays D.

Every automaton has at least one strongly connected component. When signal are imperfect, the automaton reaches a SCC with probability one, and remains in that SCC for the remainder of the supergame. Since I am focusing on the long run behavior of the automata, I restrict the set of automata to only strongly connected automata. It is important to note that if player one is plays strongly connected automaton M_1, then player two is at least weakly better playing a strongly connected automaton as well.

Lemma A.9. *For $M_1 \in \mathcal{M}^{SCC}$ and $M_2 \in \mathcal{M}\backslash\mathcal{M}^{SCC}$ and any $\varepsilon \in (0, .5]$, there exists $M_2' \in \mathcal{M}^{SCC}$ such that $U_2(M_1, M_2', \varepsilon) \geq U_2(M_1, M_2, \varepsilon)$.*

Therefore, any equilibrium over the set of strongly connected automata is also an equilibrium over the set of all finite automata. However, there may be equilibria that contain one or more automata which are not strongly connected.

The idea for this proof is as follows. Suppose player one plays a strongly connected automaton. If player two plays an automaton with more than one strongly connected component, then depending on the starting state, the system may enter any one of the strongly connected components with positive probability. If different strongly connected components yield different payoffs, player 2 is better playing the automaton with only the strongly connected component with the highest payoff.

To summarize, for the N-state analysis, I restrict the set of finite automata to those which are finite, strongly connected, and reduced. This set is denoted by $\mathcal{M}_i^R$. All equilibrium over this set are also equilibrium over the set of all finite automata. However, there may be additional equilibrium consisting of one or more non-strongly connected automata.

Proof of Lemma A.9

Since $M_2 = \left(Q_2, q_2^0, f_i, \tau_i\right)$ is not a strongly connected automaton, then the states can be divided up into strongly connected components and transient classes. Let $Q_1^{SCC}, \ldots, Q_n^{SCC} \subset Q_2$ be the strongly connected components of automaton M_2.

First, consider the trivial case that automaton starts in a strongly connected component, $q^0 \in Q_k^{SCC}$. Then let automaton M_2' have the states Q_k^{SCC} and the corresponding output function, transition function, and starting points from M_2. Since Q_k^{SCC} is strongly connected component this automaton is well defined. It is clear by the definition of strongly connected components that M_2 and M_2' yield the same payoff against M_1.

Next, consider the situation where M_2 does not start in a strongly connected component, $q_2^0 \notin Q_k^{SCC}$ for any $k = 1, \ldots, n$. Given the starting point x^0 corresponding to q_1^0 and q_2^0, the system $X\left(M_1, M_2, \varepsilon\right)$ has a unique stationary distribution $\pi\left(M_1, M_2, \varepsilon\right)\left(x^0\right)$. This stationary distribution is the convex combination of stationary distributions,

$$\pi\left(M_1, M_2, \varepsilon\right)\left(x^0\right) = \sum_{k=1}^{n} \beta_k \pi_k$$

Where β_k is the probability that starting at x^0 the system gets absorbed to Q_k^{SCC}, and π_k is the stationary distribution of the system when M_2 starts in Q_k^{SCC}. The payoff is therefore written as,

$$U_2\left(M_1, M_2, \varepsilon\right) = \sum_{k=1}^{n} \beta_k U_2\left(M_1, M_k^{SCC}, \varepsilon\right)$$

where M_k^{SCC} is the automaton composed of the states Q_k^{SCC}. Let M_2' be the automaton M_k^{SCC} which yields the highest payoff against M_1 and has positive probability of being reached, $\beta_k > 0$. Then this automaton yields at least weakly higher payoffs against M_1 than M_2. ■

B Proofs

I present the finite-state results first, as some of these are used in the two-state results.

B.1 Finite-State Results

Lemma B.1. *Given $M_1 \in \mathcal{M}_1^R$ and $M_2 \in \mathcal{M}_2^R$ and regular signal functions $r_i\left(s_i, s_j, \varepsilon\right)$, then the Markov chain $X\left(M_1, M_2, \varepsilon\right)$ is irreducible for all $\varepsilon > 0$.*

Proof of Lemma B.1

The Markov chain starts in state x^0, the state corresponding to the situation where both automata are in their initial state, $q_1\left(x^0\right) = q_1^0$ and $q_2\left(x^0\right) = q_2^0$. By definition, the Markov chain has one state for each automata-state pair that is reachable from x^0 with positive probability. Therefore,

$$\left[X\left(M_1, M_2, \varepsilon\right)\left(x^0, x\right)\right]^N > 0 \text{ for all } x \in X\left(M_1, M_2, \varepsilon\right), \text{ all } \varepsilon > 0, \text{ some } N \geq 0$$

So every state is reachable from x^0. Next, I show that x^0 is reachable from every state $x \in X(M_1, M_2, \varepsilon)$. By definition, an automaton $M_i = \left(Q_i, q_i^0, f_i, \tau_i\right)$ is strongly connected if Q_i is a strongly connected component. This means that every state in Q_i is reachable from every other state. Therefore, there is some sequence of actions, $\boldsymbol{s}_i(q_1, q_2)$, which takes M_i from state $q_1 \in Q_i$ to state $q_2 \in Q_i$. By the second condition of regular signal function, for all $\varepsilon > 0$, the probability that player i see sequence of signals $\boldsymbol{s}_i(q_1, q_2)$ is greater than 0.

I want to show that it is possible to get from any state $x \in X(M_1, M_2, \varepsilon)$ to state x^0. Let $q_i(x)$ be the state of M_i when $X(M_1, M_2, \varepsilon)$ is in state x. Then there exists sequences of actions $\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)$, $\boldsymbol{s}_1\left(q_1(x), q_1(x^0)\right)$, $\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)$, and $\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)$. Since x is reachable from x^0, then there exists sequences $\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)$ and $\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)$ of equal length, $\left|\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)\right| = \left|\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)\right|$. The length of the other sequences may not be equal.

If player 2 plays sequences $\boldsymbol{s}_1\left(q_1(x), q_1(x^0)\right)$ and $\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)$ repeatedly $\left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| + \left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| - 1$ times, and then $\boldsymbol{s}_1\left(q_1(x), q_1(x^0)\right)$ is played one more time, then M_1 goes from $q_1(x)$ to $q_1(x^0)$ in

$$\left(\left|\boldsymbol{s}_1\left(q_1(x), q_1(x^0)\right)\right| + \left|\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)\right|\right)\left(\left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| + \left|\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)\right|\right) - \left|\boldsymbol{s}_1\left(q_1(x^0) - q_1(x)\right)\right|$$

moves. Similarly, if player 1 plays sequences $\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)$ and $\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)$ repeatedly $\left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| + \left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| - 1$ times, and then $\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)$ is played one more time, then M_2 goes from $q_2(x)$ to $q_2(x^0)$ in

$$\left(\left|\boldsymbol{s}_1\left(q_1(x), q_1(x^0)\right)\right| + \left|\boldsymbol{s}_1\left(q_1(x^0), q_1(x)\right)\right|\right)\left(\left|\boldsymbol{s}_2\left(q_2(x), q_2(x^0)\right)\right| + \left|\boldsymbol{s}_2\left(q_2(x^0), q_2(x)\right)\right|\right) - \left|\boldsymbol{s}_2\left(q_2(x^0) - q_2(x)\right)\right|$$

moves. The length of these sequences are the same. So each automaton goes from $q_i(x)$ to $q_i(x^0)$, meaning the system goes from x to x^0 with positive probability. So the Markov chain is irreducible.■

Lemma 5.3. *Suppose players play automata M_1 and M_2. The average payoff for the infinitely repeated game is equal to,*

$$U_i(M_1, M_2, \varepsilon) = \sum_{x_k \in X(M_1, M_2, \varepsilon)} \pi(M_1, M_2, \varepsilon)(x_k)\, u_i(x_k)$$

where $\pi(M_1, M_2, \varepsilon)(x_k)$ is the term of the stationary distribution corresponding to state x_k, and $u_i(x_k)$ is the payoff for player i state x_k.

Proof of Lemma 5.3

By Lemma B.1, $X(M_1, M_2, \varepsilon)$ is irreducible and hence has a unique stationary distribution $\pi(M_1, M_2, \varepsilon)$ for all $\varepsilon > 0$. Let $H(x_i, T) = \frac{1}{T}\sum_{t=0}^{T} I\left\{x^t = x_i\right\}$ be the number of times that $X(M_1, M_2, \varepsilon)$ has visited state x_i in T rounds. By the law of large numbers for irreducible Markov chains (Theorem 11.12 p.439 (Grinstead and Snell 1997)), for all starting states,

$$\lim_{T \to \infty} H(x_k, T) = \pi(M_1, M_2, \varepsilon)(x_k)$$

where $\pi(M_1, M_2, \varepsilon)(x_k)$ is the term of $\pi(M_1, M_2, \varepsilon)$ corresponding to state x_k. The payoff for the first T rounds can be rewritten as,

$$U_i^T(M_1, M_2, \varepsilon) = \sum_{x_k \in X(M_1, M_2, \varepsilon)} H(x_k, T)\, u_i(x_k)$$

Therefore,

$$U_i(M_1, M_2, \varepsilon) = \lim_{T\to\infty} \sum_{x_k \in X(M_1, M_2, \varepsilon)} H(x_k, T)\, u_i(x_k) = \sum_{x_k \in X(M_1, M_2, \varepsilon)} \pi(M_1, M_2, \varepsilon)(x_k)\, u_i(x_k)$$

■

The infinite set of all possible absorbing classes of automaton M is denoted by $AC(M)$. The set of payoff-maximal absorbing states for player i is,

$$AC_i^*(M) = \{a | U_i(a) \geq U_i(b) \text{ for all } b \in C(M).\}$$

Lemma B.2. *If $a \in AC_i^*(M)$ and $q_j, q_k \in c$ such that $q_j = q_k$, then there exists $a' \in AC_i^*(M)$ such that $|a'| < |a|$.*

Proof of Lemma B.2

Consider absorbing classes $a \in AC_i^*(M)$ with $q_j, q_k \in c$ such that $q_j = q_k$. Then consider the two absorbing classes,

$$a_1 = (\{q_1, \ldots, q_{j-1}, q_j, q_{k+1}, q_n\}, \{s_1, \ldots, s_{j-1}, s_k, s_{k+1}, \ldots, s_n\})$$

$$a_2 = (\{q_{j+1}, \ldots, q_k\}, \{s_{j+1}, \ldots, s_{k-1}, s_j\})$$

Both of these satisfy the conditions for an absorbing class, because,

$$\tau(q_j, s_k) = \tau(q_k, s_k) = q_{k+1}$$

and,

$$\tau(q_k, s_j) = \tau(q_j, s_j) = q_{j+1}$$

The payoff of absorbing class a is,

$$\begin{aligned}
U_i^{AC}(a) &= \frac{1}{n} \sum_{l=1}^{n} u_i(f(q_l), s_l) \\
&= \frac{1}{n} \left[(n-k+j) \left(\frac{1}{n-k+j} \sum_{l=1,k+1}^{j,n} u_i(f(q_l), s_l) \right) + (k-j) \left(\frac{1}{k-j} \sum_{l=j+1}^{k} u_i(f(q_l), s_l) \right) \right] \\
&= \left(\frac{n-k+j}{n} \right) U_i^{AC}(a_1) + \left(\frac{k-j}{n} \right) U_i^{AC}(a_2)
\end{aligned}$$

Since $a \in AC_i^*(M)$, it must be that $U_i^{AC}(a) = U_i^{AC}(a_1) = U_i^{AC}(a_2)$, or else either a_1 or a_2 would have higher payoff than a. Since $0 < j < k < n$, $|a_1| = n-k+j < n = |a|$ and $|a_2| = k-j < n = |c|$.

So for all payoff-maximal absorbing classes with multiple visits to one state, there exists a smaller payoff-maximal absorbing class. ■

The set of absorbing classes which contain all unique states for player i is denoted by,

$$AC_i^U(M) = \{a | q_i \neq q_j \text{ for all } q_i, q_j \in a\}$$

Lemma B.3. *At least one of the unique state absorbing classes is payoff-maximal.*

$$AC_i^*(M) \cap AC_i^U(M) \neq \emptyset.$$

Proof of Lemma B.3

Select any absorbing class $a \in AC_i^*(M)$. By Lemma B.2, for each payoff-maximal absorbing class that visits state q_j more than once, there exists another payoff-maximal absorbing class that visits q_i strictly less. This process can be repeated until the new absorbing class visits state q_j only once. This can be done for each state in a. Then end result is a payoff-maximal absorbing class in the set of unique state absorbing classes. ■

Lemma B.3 suggests that there is a payoff-maximal absorbing class with weakly fewer states than automaton M, which means that it is finite. Player i's payoff-maximal absorbing class is denoted by $a_i^*(M)$.

Lemma B.4. $U_i(M_1, M_2, \varepsilon) \leq U_i^{AC}(a_i^*(M_{-i}))$ *for all* $\varepsilon \in [0, .5]$ *and all* $M_i \in \mathcal{M}$.

Proof

Suppose by means of contradiction that for some $\varepsilon \in [0, .5]$, $U_2(M_1, M_2, \varepsilon) > U_2^{AC}(a_2^*(M_1))$. The Markov chain $X(M_1, M_2, \varepsilon)$ yields a sequence of automaton-state pairs $x^0, x^1, x^2, \ldots$. By definition,

$$U_2(M_1, M_2, \varepsilon) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} u_2\left(x^t\right)$$

where $u_2\left(x^t\right)$ is the payoff to player 2 for the automaton-state profile x^t. For every finite integer K, there must be some sequence of length K of automaton-state pairs $y^1, \ldots, y^K$ such that,

$$\frac{1}{K} \sum_{k=1}^{K} u_2\left(y^k\right) \geq U_2(M_1, M_2, \varepsilon) \tag{4}$$

Let $|Q_1|$ be the number of states in M_1, and let $\underline{u}_2$ be the lowest possible stage-game payoff for player 2. Set $\bar{K}$ to be a sufficiently high integer such that,

$$U_2(M_1, M_2, \varepsilon) - U_2(a_2^*(M_1)) > \frac{j\left(U_2(M_1, M_2, \varepsilon) - \underline{u}_2\right)}{\bar{K} + j} \tag{5}$$

holds for all $j = 1, \ldots, |Q_1|$. From (5), we get,

$$\begin{aligned} U_2\left(a_2^*\left(M_1\right)\right) &< U_2\left(M_1, M_2, \varepsilon\right) - \frac{j\left(U_2\left(M_1, M_2, \varepsilon\right) - \underline{u}_2\right)}{\bar{K} + j} \\ &= \frac{\bar{K} U_2\left(M_1, M_2, \varepsilon\right) - j \underline{u}_2}{\bar{K} + j} \end{aligned} \tag{6}$$

Fix sequence of automaton-state pairs $y^1, \ldots, y^{\bar{K}+1}$ such that,

$$\frac{1}{\bar{K}} \sum_{k=1}^{\bar{K}} u_2\left(y^k\right) \geq U_2\left(M_1, M_2, \varepsilon\right)$$

Let $q_i^1, q_i^2, \ldots, q_i^{\bar{K}+1}$ be the sequence of states for automaton i from the sequence of automaton-state pairs $y^1, \ldots, y^{\bar{K}+1}$. Automaton M_1 starts in state q_i^1 and ends in state $q_i^{\bar{K}+1}$. Because M_1 is strongly connected, there exists some sequence of actions $s_2^1, s_2^2, \ldots, s_2^j \in S_2$, that moves automaton M_1 from state $q_i^{\bar{K}+1}$ to state q_i^1. Let $p_1^1 = q_1^{\bar{K}+1}$ and $p_1^l = \tau_2\left(p_1^{l-1}, s_2^{l-1}\right)$ for $l = 2, \ldots j$. By construction it must be that, $\tau_2\left(p_1^j, s_2^j\right) = q_1^1$. Since M_1 has $|Q_1|$ states, then $\left|\left\{s_2^1, s_2^2, \ldots, s_2^j\right\}\right| \leq |Q_1|$. Define the absorbing class

$$a_2'\left(M_1\right) = \left(\left\{q_1^1, q_1^1, \ldots, q_1^{\bar{K}}, p_1^1, p_1^2, \ldots, p_1^j\right\}, \left\{f_2\left(q_2^1\right), f_2\left(q_2^2\right), \ldots, f_2\left(q_2^{\bar{K}}\right), s_2^1, s_2^2, \ldots, s_2^j\right\}\right)$$

This is a well defined absorbing class. Note that $u_2\left(f_1\left(p_1^k\right), s_1^k\right) \geq \underline{u}_2$ for $k = 1, \ldots, j$. Therefore,

$$\begin{aligned} U_2^{AC}\left(a_2'\left(M_1\right)\right) &= \frac{1}{\bar{K} + j}\left[\sum_{k=1}^{\bar{K}} u_2\left(y^k\right) + \sum_{k=1}^{j} u_2\left(f_1\left(p_1^k\right), s_2^k\right)\right] \\ &\geq \frac{1}{\bar{K} + j}\left[\sum_{k=1}^{\bar{K}} u_2\left(y^k\right) + j \underline{u}_2\right] && \text{(By minimality of } \underline{u}_2.\text{)} \\ &\geq \frac{\left[\bar{K} U_2\left(M_1, M_2, r_1(\varepsilon), r_2(\varepsilon)\right) + j \underline{u}_2\right]}{\bar{K} + j} && \text{(From (4))} \\ &> U_2\left(a_2^*\left(M_1\right)\right) && \text{(From (6))} \end{aligned}$$

This contradicts the maximality of $a_2^*\left(M_1\right)$. ■

Lemma B.5. *Given regular signal functions r_i, if $X\left(M_1, M_2, \varepsilon\right)$ has communicating classes $A_1, \ldots, A_m$, then*

$$\lim_{\varepsilon \to 0} U_i\left(M_1, M_2, \varepsilon\right) = \sum_{A_k | \gamma(A_k) = \gamma^*} \beta(A_k) U_i^{CC}\left(A_k\right)$$

with $\sum_{A_k | \gamma(A_k) = \gamma^*} \beta(A_k) = 1$ and $\beta\left(A_k\right) > 0$ for all A_k such that $\gamma\left(A_k\right) = \gamma^*$

Proof

By Lemma B.1, the Markov chain $X(M_1, M_2, \varepsilon)$ is irreducible and has a unique stationary distribution $\pi(M_1, M_2, \varepsilon)$. Let $\pi(M_1, M_2, \varepsilon)(x)$ denote the term of the stationary distribution corresponding to state $x \in X(M_1, M_2, \varepsilon)$. By Theorem B.13, if a communicating class doesn't minimize stochastic potential, $\gamma(A) > \gamma^*$, then,

$$\lim_{\varepsilon \to 0} \pi(M_1, M_2, \varepsilon)(x) = 0 \text{ for all } x \in A. \tag{7}$$

If a communicating class A does minimize stochastic potential, $\gamma(A) = \gamma^*$, then,

$$\lim_{\varepsilon \to 0} \pi(M_1, M_2, \varepsilon)(y) > 0 \text{ for all } y \in A. \tag{8}$$

For each communicating class A_k, there exists some constant, $\alpha(A_k)$ such that,

$$\lim_{\varepsilon \to 0} \sum_{x \in A_k} \pi(M_1, M_2, \varepsilon)(x) u_i(x) = \alpha(A_k) \sum_{x \in A_k} u_i(x) \tag{9}$$

Then,

$$\begin{aligned}
\lim_{\varepsilon \to 0} U_i(M_1, M_2, \varepsilon) &= \lim_{\varepsilon \to 0} \sum_{x \in X(M_1, M_2, \varepsilon)} \pi(M_1, M_2, \varepsilon)(x) u_i(x) && \text{(by Lemma 5.3)} \\
&= \lim_{\varepsilon \to 0} \sum_{A_k | \gamma(A_k) = \gamma^*} \sum_{x \in A_k} \pi(M_1, M_2, \varepsilon)(x) u_i(x) && \text{(by (7))} \\
&= \sum_{A_k | \gamma(A_k) = \gamma^*} \alpha(A_k) \sum_{x \in A_k} u_i(x) && \text{(by (9))} \\
&= \sum_{A_k | \gamma(A_k) = \gamma^*} \alpha(A_k) |A_k| U_i^{CC}(A_k) && \text{(by def. of } U_i^{CC})
\end{aligned}$$

Set $\beta(A_k) = \alpha(A_k)|A_k|$, then $\sum_{A_k|\gamma(A_k)=\gamma^*} \beta(A_k) = \sum_{x \in X(M_1, M_2, \varepsilon)} \pi(M_1, M_2, \varepsilon)(x) = 1$ and $\beta(A_k) > 0$ for all A_k such that $\gamma(A_k) = \gamma^*$ by (8). ■

Definition B.6 (Homing Sequence)**.** *Given automaton $M = (Q, q^0, f, \tau)$, the action sequence $h \in S^n$ is a homing sequence if and only if,*

$$\forall q_1, q_2 \in Q \text{ and } q_1\langle h\rangle = q_2\langle h\rangle \Rightarrow q_1 h = q_2 h$$

Where $q\langle h\rangle \in f\left(S^{n+1}\right)$ is the output of M starting at state q when the sequence h is played, and qh is the end state of M when h is played.

This means that when h is played, the output of M allows us to determine the current state of M.

Theorem B.7 (Kohavi (1978))**.** *A preset homing sequence, whose length is at most $(n-1)^2$, exists for every reduced, strongly connected n-state machine M.*

Lemma 5.7. *Given automaton $M_1 \in \mathcal{M}^R$ with n states, and any absorbing class $a(M_1)$, there* *exists automaton M_2 such that for all communicating classes, A_k, of the system $X(M_1, M_2, \varepsilon)$,*

$$U_2^{CC}(A_k) = U_2^{AC}(a(M_1)).$$

Proof of Lemma 5.7

I construct automaton $M_2 = (Q_2, q_2^0, f_2, \tau_2)$ which yields the desired properties. Consider automaton M_1 with n states and absorbing class

$$a = (\{q_1, \ldots, q_m\}, \{s_1, \ldots, s_m\})$$

with $m \leq n$ states.

The automaton will be made up of three main parts. The first part is the absorbing class. This section of the automaton will keep the system in the desired absorbing state when reached. Then second part of the automaton is the homing region. In this region, the automaton plays the homing sequence. Based on the response from M_1, the current state of M_1 is determined. The goal of the homing region is to determine the current state of automaton M_1 after an error has been made. Once the state of automaton M_1 is known, it will be possible to move it back into the absorbing class. The third part allows the two automata to resynchronize, transitioning from the homing region back to the absorbing class.

Start constructing M_2 by creating states and transitions such that the absorbing class is maintained. That is for each state q_j in the absorbing class a of M_1, create corresponding state p_j in automaton M_2 that satisfies,

$$f_2(p_j) = s_j \text{ and } \tau(p_j, f_1(q_j)) = \begin{cases} p_{j+1} & j < m \\ p_1 & j = m \end{cases}$$

Also, let all incorrect plays in the absorbing class states lead to state p_{m+1},

$$\tau(p_j, s \neq f_1(q_j)) = p_{m+1}$$

The second region of the automaton is the homing region. By Theorem B.7, there exists a homing sequence for automaton M_1, call this $h(M_1) = \{h_1, \ldots, h_l\}$. There is a set of sequences imposed by this homing sequence when started at different states,

$$S(h) = \{q\langle h\rangle | q \in Q\} = \left\{ \begin{array}{c} (s_1^1, \ldots, s_l^1) \\ (s_1^2, \ldots, s_l^2) \\ \vdots \\ (s_1^k, \ldots, s_l^k) \end{array} \right\}$$

There is also a set of states that this homing sequence will lead to,

$$Q(h) = \left\{ q | q'h = q \text{ for some } q' \in Q \right\}$$

Let $S(h,j)$ for $0 < j \leq l$ be the first j terms of these sequences,

$$S(h,j) = \left\{ \begin{array}{c} \left(s_1^1, \ldots, s_j^1\right) \\ \left(s_1^2, \ldots, s_j^2\right) \\ \vdots \\ \left(s_1^k, \ldots, s_j^k\right) \end{array} \right\} = \left\{ \begin{array}{c} \boldsymbol{s}^1(j) \\ \boldsymbol{s}^2(j) \\ \vdots \\ \boldsymbol{s}^k(j) \end{array} \right\}$$

Let $S^U(h,j)$ be the set of unique sequences of length j imposed by h,

$$S^U(h,j) = \left\{ \begin{array}{c} \left(s_1^1, \ldots, s_j^1\right) \\ \left(s_1^2, \ldots, s_j^2\right) \\ \vdots \\ \left(s_1^u, \ldots, s_j^u\right) \end{array} \right\} = \left\{ \begin{array}{c} \boldsymbol{s}^1(j) \\ \boldsymbol{s}^2(j) \\ \vdots \\ \boldsymbol{s}^u(j) \end{array} \right\},$$

where $\left|S^U(h,j)\right| = u(j)$ is the number of unique sequences in $S(h,j)$.

The homing region consists of $l+1$ classes, $P_{m+1}, \ldots, P_{m+l+1}$. Class P_{m+1+j} consists of $u(j)$ states, $p_{m+1+j}\left(\boldsymbol{s}^1(j)\right), \ldots, p_{m+1+j}\left(\boldsymbol{s}^{u(j)}(j)\right)$, one corresponding to each sequence in $S^U(h,j)$. Define $S^U(h,0) = \{\emptyset\}$ and $u(0) = 1$. Automaton M_2 will play the same action in all states of a given class,

$$f_2(p) = h_i \text{ for all } p \in P_{m+i} \text{ for } i \in \{1, \ldots, l\}$$

This choice will correspond to the matching term in the homing sequence.

The transition function for $0 < i \leq l$ is defined as follows.

$$\tau_2\left(p_{m+i}(\boldsymbol{s}), C\right) = \begin{cases} p_{m+i+1}\left(\{\boldsymbol{s}, C\}\right) & \text{if } \{\boldsymbol{s}, C\} \in S^U(h,i) \\ p_{m+i+1}\left(\{\boldsymbol{s}, D\}\right) & \text{if } \{\boldsymbol{s}, C\} \notin S^U(h,i) \end{cases}$$

$$\tau_2\left(p_{m+i}(\boldsymbol{s}), D\right) = \begin{cases} p_{m+i+1}\left(\{\boldsymbol{s}, D\}\right) & \text{if } \{\boldsymbol{s}, D\} \in S^U(h,i) \\ p_{m+i+1}\left(\{\boldsymbol{s}, C\}\right) & \text{if } \{\boldsymbol{s}, D\} \notin S^U(h,i) \end{cases}$$

Finally, the last region of M_2 will resynchronize play, and get the system back to the absorbing class a. There will be k states in class P_{m+l+1}. By definition of the homing sequence, for each state $p_{m+l+1}(q) \in P_{m+l+1}$ there is a corresponding state $q \in M_1$ such that when M_2 is in state $p_{m+l+1}(q)$, M_1 is in state q. Define the resynchronizing sequence $\boldsymbol{t}(q) = \left\{t_1(q), t_2(q), \ldots, t_{r(q)}(q)\right\}$ to be the sequence of plays necessary to get from state q to state q_1 where $r(q) = |\boldsymbol{t}(q)|$. This sequence exists for each state because M_1 is strongly connected. Then for each state $p^1(q) = p_{m+l+1}(q) \in P_{m+l+1}$, for $0 < i < r(q)$.

$$\tau_2\left(p^i(q), \text{C or D}\right) = p^{i+1}(q)$$

and

$$\tau_2\left(p^{r(q)}(q), \text{C or D}\right) = p_1$$

The output function for $0 < i \leq r(q)$ is,

$$f_2\left(p^i(q)\right) = t_i(q)$$

So the system will always end up in state (q_1, p_1) regardless of the starting position. Once the system is in (q_1, p_1), it has entered the communicating class, and will not leave without errors. ■

Definition B.8 (Regular Perturbation). *Given Markov chain X, a perturbation X^ε is called a regular perturbation if the following three conditions hold,*

1. *X^ε is irreducible for all $\varepsilon \in (0, .5]$.*
2. *$\lim\limits_{\varepsilon \to 0} X^\varepsilon(x, y) = X(x, y)$*
3. *$X^\varepsilon(x, y) > 0$ for some ε implies $\exists n \geq 0$ such that $0 < \lim\limits_{\varepsilon \to 0} \varepsilon^{-n} X^\varepsilon(x, y) < \infty$*

Let $A_1, \ldots, A_m$ be the communicating classes of $X(M_1, M_2, \varepsilon)$. To leave a communicating class, there must be at least one incorrect signal.

Definition B.9 (Resistance). *The resistance ρ_{ij} is the order of the probability that the system goes from communicating class A_i to A_j.*

$$\rho_{ij} = \min_{x \in A_i, y \in A_j, n \in \mathbb{N}} \mathcal{O}\left(X\left(M_1, M_2, \varepsilon\right)(x, y)^n\right),$$

where $\mathcal{O}(\cdot)$ is the order of the function. If the probability is 0, then the resistance is defined to be ∞.

Define the graph $\mathcal{G}$, which has one vertex, v_k, for every communicating class A_k. For every vertex pair, $v_i, v_j \in \mathcal{G}$, there is an edge with resistance ρ_{ij}.

Definition B.10 (i-tree). *An i-tree in $\mathcal{G}$ is a spanning tree such that from every vertex $j \neq i$, there is a unique path directed from j to i.*

For each vertex, $\mathcal{T}_i$ is the set of all i-trees on $\mathcal{G}$. The resistance on an i-tree is,

$$\rho(\tau) = \sum_{(i,j) \in \tau} \rho_{ij}$$

Definition B.11 (Stochastic Potential). *The stochastic potential of the communicating class A_i is the least resistance among all i-trees:*

$$\gamma_i = \min_{\tau \in \mathcal{T}_i} \rho(\tau)$$

The stochastic potential measures the likelihood of the system visiting a certain communicating class. Communicating classes that don't have the minimum stochastic potential are at least an order ε less likely to be visited by the system. As the errors approach zero, the system spends non-trivial amounts of the supergame in only the communication classes with minimum stochastic potential. Finally, define the minimum stochastic potential of the system to be,

$$\gamma^* = \min_{i=1,\ldots m} \gamma_i$$

Lemma B.12. *Given automata M_1 and M_2 subject to regular signal functions r_1 and r_2, the perturbed system $X(M_1, M_2, \varepsilon)$ is a regular perturbation.*

Proof

To show that this is true, I must show that the three criteria are satisfied. The system formed by automata M_1 and M_2 and regular signal functions r_1 and r_2 is represented by Markov chain $X(M_1, M_2, \varepsilon)$. By Lemma B.1, $X(M_1, M_2, \varepsilon)$ is always irreducible, so the first criterion is satisfied. By the first part of the definition of regular signal function and (2), the second criterion is satisfied. Finally, it is clear that the third condition of the regular signal function remains under addition and multiplication, so the third criterion also holds by (2). ■

Theorem B.13 (Theorem 4 from Young (1993))**.** *Let X^0 be a stationary Markov process on a finite state space with communicating communication classes $A_1, \ldots, A_m$. Let X^ε be a regular perturbation of X^0, and let π^ε be its unique stationary distribution for every small positive ε. Then:*

1. *as $\varepsilon \to 0$, π^ε converges to a stationary distribution π^0 of X^0.*
2. *x is stochastically stable ($\pi_x^0 > 0$) if and only if x is contained in a communicating class A_j that minimizes γ_j.*

The second part of this theorem implies that a communicating class is prevalent if and only if it minimizes stochastic potential.

Theorem 5.6. *Suppose players play supergame G with regular signal function r_i, and play automata $M_i \in \mathcal{M}_i^R$ represented by Markov chain $X(M_1, M_2, \varepsilon)$. If there exists some $\bar{\varepsilon} > 0$ such that (M_1, M_2) is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$, then for all prevalent communicating classes A_k, $U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i}))$.*

Proof of Theorem 5.6

Suppose that (M_1, M_2) is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$. Suppose by means of contradiction that there exists a communicating class A_k such that $\gamma(A_k) = \gamma^*$ and

$$U_2^{CC}(A_k) < U_2^{AC}(a_2^*(M_1)) \tag{10}$$

Using (10), Lemma B.5 and that fact that a communicating class can never get payoff higher than the optimal absorbing class gives,

$$U_2(M_1, M_2, \varepsilon) < U_2^{AC}(a_2^*(M_1)) \tag{11}$$

By Lemma 5.7, there exists an automaton M_2' such that for all communicating classes A of $X(M_1, M_2, \varepsilon)$, $U_2^{CC}(A) = U_2^{AC}(a_2^*(M_1))$. Therefore, by Lemma B.5,

$$\lim_{\varepsilon \to 0} U_2\left(M_1, M_2', \varepsilon\right) = U_2^{CC}(A) = U_2^{AC}(a_2^*(M_1)) \tag{12}$$

By Lemma 1 from Young (1993), the stationary distribution of $X(M_1, M_2', \varepsilon)$ is continuous at $\varepsilon = 0$. Therefore the payoff must also be continuous at $\varepsilon = 0$. So, for all $\varepsilon \in (0, \bar{\varepsilon})$, there exists some $\delta > 0$

such that,

$$\left|\lim_{\varepsilon\to 0} U_2\left(M_1, M_2', \varepsilon\right) - U_2\left(M_1, M_2', \varepsilon\right)\right| < \delta \tag{13}$$

Set $\bar{\varepsilon}$ sufficiently small so that,

$$\left|\lim_{\varepsilon\to 0} U_2\left(M_1, M_2', \varepsilon\right) - U_2\left(M_1, M_2', \varepsilon\right)\right| < \left|U_2^{AC}\left(a_2^*\left(M_1\right)\right) - U_2\left(M_1, M_2, \varepsilon\right)\right| \tag{14}$$

By (11), (12), and (14) for all $\varepsilon \in (0, \bar{\varepsilon})$,

$$U_2\left(M_1, M_2, \varepsilon\right) < U_2\left(M_1, M_2', \varepsilon\right)$$

So (M_1, M_2) is not an equilibrium for any $\varepsilon \in (0, \bar{\varepsilon})$, which is a contradiction. ■

Fix automaton M_1. This automaton has some optimal absorbing class $a_2^*(M_1)$. Let $\mathcal{M}^{SPM}(M_1)$ be the set of all automata $M_2 \in \mathcal{M}^R$ such that all communicating classes A_k of $X(M_1, M_2, \varepsilon)$ that minimize stochastic potential ($\gamma(A_k) = \gamma^*$) yield the optimal absorbing class payoff, $U_2^{CC}(A_k) = U_2^{AC}(a_2^*(M_1))$.

Theorem 5.8. *Suppose players play supergame G with regular signal function r_i, and play automata $M_i \in \mathcal{M}_i^R$ represented by Markov chain $X(M_1, M_2, \varepsilon)$. If*

1. *for all prevalent communicating classes A_k, $U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i}))$, and*
2. *$\frac{\partial U_i(M_1,M_2,0)}{\partial\varepsilon} = \sup_{M\in\mathcal{M}^{SPM}(M_{-i})} \frac{\partial U_i(M_i,M,0)}{\partial\varepsilon}$;*

then there exists some $\bar{\varepsilon} > 0$ such that (M_1, M_2) is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$.

Proof of Theorem 5.8

Fix (M_1, M_2) represented by $X(M_1, M_2, \varepsilon)$ such that for all stochastic potential minimizing communicating classes $\gamma(A_k) = \gamma^*$,

$$U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i})) \tag{15}$$

and

$$\frac{\partial U_i(M_1, M_2, \varepsilon)}{\partial \varepsilon} = \sup_{M\in\mathcal{M}^{SPM}(M_{-i})} \frac{\partial U_i(M_1, M, \varepsilon)}{\partial \varepsilon}$$

Without loss of generality, I will show that when these conditions are satisfied, M_2 is a best response to M_1. For all $M_2 \notin \mathcal{M}^{SPM}(M_1)$, there exists a stochastic potential minimizing communicating class such that $U_2^{CC}(A_k) < U_2^{AC}(a^*(M_1))$. By Lemma B.5 and that fact that a communicating class can never get payoff higher than the optimal absorbing class,

$$U_2(M_1, M_2, \varepsilon) < U_2^{AC}(a_2^*(M_1)) \text{ for all } M_2 \notin \mathcal{M}^{SPM}(M_1) \tag{16}$$

For all $M_2 \in \mathcal{M}^{SPM}(M_1)$,

$$\lim_{\varepsilon\to 0} U_2(M_1, M_2, \varepsilon) = U_2^{AC}(a_2^*(M_1))$$

By continuity of U_2, this means that for all $\varepsilon \in (0, \bar{\varepsilon})$ for $\bar{\varepsilon}$ sufficiently small,

$$U_2(M_1, M, \varepsilon) < U_2\left(M_1, M', \varepsilon\right) \text{ for all } M \notin \mathcal{M}^{SPM}, M' \in \mathcal{M}^{SPM}$$

So the best response to M_1 for $\varepsilon \in (0, \bar{\varepsilon})$ must come from the set $\mathcal{M}^{SPM}$. For all $M \in \mathcal{M}^{SPM}(M_1)$,

$$\lim_{\varepsilon \to 0} U_2(M_1, M, \varepsilon) = U_2^{AC}(a^*(M_1))$$

By definition of the derivative, for some $\bar{\varepsilon} > 0$,

$$\left.\frac{\partial U_2(M_1, M, \varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} \leq \left.\frac{\partial U_2(M_1, M', \varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0}$$

$$\Rightarrow U_2(M_1, M, \varepsilon) \leq U_2\left(M_1, M', \varepsilon\right) \text{ for all } \varepsilon \in (0, \bar{\varepsilon})$$

Therefore, if M_2 satisfies,

$$\frac{\partial U_i(M_1, M_2, \varepsilon)}{\partial \varepsilon} = \sup_{M \in \mathcal{M}^{SPM}(M_{-i})} \frac{\partial U_i(M_1, M, \varepsilon)}{\partial \varepsilon}$$

Then it must be that for all $\varepsilon \in (0, \bar{\varepsilon})$,

$$U_2(M_1, M, \varepsilon) \leq U_2(M_1, M_2, \varepsilon) \text{ for all } M \in \mathcal{M}^{SPM}$$

Therefore, M_2 is a best response to M_1. ■

B.2 Two-State Results

Proposition B.14. *Players play supergame G, where each action in stage game g has a unique best response. For any error $\varepsilon \in (0, 1/2]$, both players playing automata equivalent to open-loop finite automata is an equilibrium of the supergame G if and only if they play a Nash equilibrium of the stage game in every round of the supergame.*

If player 1 is playing a open-loop automaton, then it plays a fixed sequence of actions. The best response to this is simply to best respond in every round. If player 2's automaton is not equivalent to an open loop strategy and the chance of misperception is positive, then it is possible that a misperception could lead to a situation where player 2 doesn't best respond to player 1 in a given round.

If players play automata which always play the same action, and this action pair is a Nash equilibrium of the stage game, then this pair of automata always has to be a Nash equilibrium of the supergame. In addition, if the stage game has multiple Nash equilibrium, then any payoff in the convex hull of the Nash equilibrium payoffs is possible in equilibrium.

Proof of Proposition B.14

$\Rightarrow$First suppose that both players play automata equivalent to open loop automata M_1 and M_2. These form the Markov chain $X^{\varepsilon}_{M_1,M_2}$ with n states and all entries either 0 or 1. Depending on x^0, the Markov chain loops through $m \leq n$ states, $x^1, \ldots, x^m$. This yields payoff,

$$U_i(M_1, M_2, \varepsilon) = \frac{1}{m} \sum_{k=1}^{m} u_i\left(x^k\right)$$

Suppose without loss of generality that the actions in state x^j are not a Nash equilibrium of the stage game, because player 2 receives higher payoff from playing s_2^j than $f_2\left(q_2(x^j)\right)$ when player 1 plays $f_1\left(q_1(x^j)\right)$,

$$u_2\left(x^j\right) < u_2\left(f_1\left(q_1(x^j)\right), s_2^j\right)$$

Then player 2 is better playing automaton M' which is the same as M_2 except $f_2\left(q_2(x^j)\right) = s_2^j$,

$$U_2\left(M_1, M_2, \varepsilon\right) = \sum_{k\neq j} u_2(x^k) + u_i(x^j) < \sum_{k\neq j} u_2(x^k) + u_2(f_1\left(q_1(x^j)\right), s_2^j) = U_2\left(M_1, M', \varepsilon\right)$$

So both players playing automata equivalent to open loop automata M_1 and M_2 is an equilibrium only if a Nash equilibrium is played in every round.

$\Leftarrow$ Assume that automata M_1 and M_2 generate a sequence of actions which yield a Nash equilibrium in every stage game. Suppose that M_1 is not equivalent to an open loop automaton. For some state q_1,

$$f_1\left(\tau_1\left(q_1, C\right)\right) \neq f_1\left(\tau_1\left(q_1, D\right)\right)$$

So when M_1 is in q_1, the play in the next round can be either $f_1\left(\tau_1\left(q_1, C\right)\right)$ or $f_1\left(\tau_1\left(q_1, D\right)\right)$. Since $\varepsilon > 0$, either signal is possible with positive probability. Automaton M_2 will play s_2, which has a unique best response. So, with positive probability the system of automata M_1 and M_2 will not play a Nash equilibrium of the stage game. This contradicts the assumption that M_1 is not equivalent to an open loop automaton. A similar argument holds for M_2. Therefore, if M_1 and M_2 generate a sequence of action which yield a Nash equilibrium in every stage game that has unique best responses, the automata must be equivalent to open-loop automata. ■

Definition B.15 (Eventually Always Plays)**.** *An automaton $M_i = (Q_i, q_i^0, f_i, \tau_i)$ eventually always plays action $s_i \in S_i$ if for all strongly connected components $Q_k^{SCC} \subseteq Q_i$,*

$$f_i\left(q\right) = s_i \text{ for all } q \in Q_k^{SCC}$$

Lemma B.16. *When $\varepsilon > 0$, and automaton M which eventually always plays C is payoff equivalent to M^C over the set of automata with only one SCC.*

Proof of Lemma B.16

Assume that player 1 plays $M_1 = M$. Assume that player 2 plays $M_2 = \left(Q_2, q_2^0, f_2, \tau_2\right)$ which has one strongly connected component. Let T_i be the round for which automata M_i reaches a strongly connected component. Since $\varepsilon > 0$, any sequence of signals occurs with positive probability, so $P\left(T_i < \infty\right) = 1$. Let u_i^t be the payoff for player i in round t. Let $T^* = \max(T_1, T_2)$. Then,

$$\begin{aligned} U_i\left(M_1, M_2, \varepsilon\right) &= \lim_{T\to\infty} \frac{1}{T}\left[\sum_{k=0}^{T^*} u_i^k + \sum_{k=T^*+1}^{T} u_i^k\right] \\ &= \lim_{T\to\infty} \frac{1}{T}\left[\sum_{k=T^*+1}^{T} u_i^k\right] \\ &= U_i\left(M', M_2, \varepsilon\right) \end{aligned}$$

So any automaton M with only one SCC is payoff equivalent to the automaton M' consisting only of the states of the SCC over the set of automata with only one strongly connected component. ■

Lemma B.17. *The set of two-state automata, $\mathcal{M}^2$, can be reduced to a smaller set of automata, $\bar{\mathcal{M}}^2$, such that,*

1. *For all $M \in \mathcal{M}^2$, there exists some $M' \in \bar{\mathcal{M}}^2$ such that M and M' are payoff equivalent over $\mathcal{M}^2$.*

2. *For all $M, M' \in \bar{\mathcal{M}}^2$, M and M' are not payoff equivalent over $\mathcal{M}^2$.*

Proof of Lemma B.17

There are $|S_i|^N \left(N^N\right)^{|S_{-i}|}$ total N-state automata when the starting states are omitted. So when both players have two actions, there are 64 two-state automata. Many of these automata are redundant.

First, divide the 64 into four categories, each containing 16 automata:

$$\mathcal{M}_1^2 = \left\{M \in \mathcal{M}^2 | f\left(q_1\right) = C, f\left(q_2\right) = C\right\}$$
$$\mathcal{M}_2^2 = \left\{M \in \mathcal{M}^2 | f\left(q_1\right) = C, f\left(q_2\right) = D\right\}$$
$$\mathcal{M}_3^2 = \left\{M \in \mathcal{M}^2 | f\left(q_1\right) = D, f\left(q_2\right) = C\right\}$$
$$\mathcal{M}_4^2 = \left\{M \in \mathcal{M}^2 | f\left(q_1\right) = D, f\left(q_2\right) = D\right\}$$

The automata in $\mathcal{M}_1^2$ play C regardless of the play of the other automaton. Therefore, these automata are equivalent to M^C, and hence payoff equivalent to M^C over the set $\mathcal{M}^2$. Similarly, the automata in $\mathcal{M}_4^2$ all play D regardless of the play of the other, so they are all payoff equivalent to M^D over $\mathcal{M}^2$.

For every $M_2 \in \mathcal{M}_2^2$, there exists an equivalent $M_3 \in \mathcal{M}_3^2$ (the only difference is that the states are switched). For example, $M_2 = \left(\{C, D\}, \{q_1, q_1\}, \{q_2, q_2\}\right)$ and $M_3 = \left(\{D, C\}, \{q_2, q_2\}, \{q_1, q_1\}\right)$. Both of these automata implement tit-for-tat, so they produce the same output regardless of the input, and hence are payoff equivalent. Without loss of generality, I only consider those automata in $\mathcal{M}_2^2$.

If automaton $M^E = \left(\{C, D\}, \{q_1, q_2\}, \{q_1, q_2\}\right)$ starts in q_1, then regardless of the signals it plays C in every round of the supergame, and hence is equivalent to M^C. If M^E starts in q_2, then regardless of the signals, it plays D in every round, and hence is equivalent to M^D. So depending on the starting point, M^E is equivalent to either M^C or M^D. After equivalent automata have been eliminated, there are 17 remaining automata: M^C, M^D, and the set $\mathcal{M}_2^2 \backslash M^E$.

Note that all two-state automata have only one reachable SCC. For a two-state automaton to have multiple strongly connected components, each state needs to be a strongly connected component. The only two-state automaton which satisfies this is $M^E = \left(\{C, D\}, \{q_1, q_2\}, \{q_1, q_2\}\right)$. If M^E starts in q_k, then only q_k can be reached, so it only has one reachable SCC, regardless of the starting point. Therefore, by Lemma B.16, any automaton which eventually always plays C is payoff equivalent to M^C over the set $\mathcal{M}^2$.

Out of the 17 remaining automata, three eventually always play C, and three eventually always play D ,

Eventually Always Play C	Eventually Always Play D
$(\{C, D\}, \{1,1\}, \{1,2\})$	$(\{C, D\}, \{1,2\}, \{2,2\})$
$(\{C, D\}, \{1,1\}, \{1,2\})$	$(\{C, D\}, \{2,2\}, \{1,2\})$
$(\{C, D\}, \{1,1\}, \{1,1\})$	$(\{C, D\}, \{2,2\}, \{2,2\})$

So by Lemma B.16, these automata are payoff equivalent to M^C and M^D over $\mathcal{M}^2$. The remaining 11 automata for the minimal set $\bar{\mathcal{M}}^2$.

1. M^C	5. $(\{C, D\}, \{1,1\}, \{2,1\})$	9. $(\{C, D\}, \{2,1\}, \{2,2\})$
2. M^D	6. $(\{C, D\}, \{1,1\}, \{2,2\})$	10. $(\{C, D\}, \{2,2\}, \{1,1\})$
3. M^{CD}	7. $(\{C, D\}, \{2,1\}, \{1,1\})$	11. $(\{C, D\}, \{2,2\}, \{2,1\})$
4. M^{WSLS}	8. $(\{C, D\}, \{2,1\}, \{1,2\})$	

■

Theorem 4.3. *In the infinitely repeated PD game, when players have the simple signal function r_i^S and choose among the set of two-state automata, $\mathcal{M}^2$, there are only three types of robust equilibria:*

1. *$L < 0$ and M_i is payoff equivalent to M^C.*
2. *$L > 0$ and M_i is payoff equivalent to M^D.*
3. *$-(1-2\varepsilon)^3 < L < (1-2\varepsilon)^3$ and $M_i = M^{WSLS}$.*

Proof of Theorem 4.3

If M_2 is the best response to M_1, then any automaton which is payoff equivalent to M_2 is also a best response to M_1. Therefore, I only need to consider the automata in reduce payoff equivalent set $\bar{\mathcal{M}}(2)$ from Lemma B.17 when finding equilibrium. However, if one of the automata in $\bar{\mathcal{M}}(2)$ is an equilibrium, then all payoff equivalent automata are also equilibria.

Three of the automata in $\bar{\mathcal{M}}(2)$ are open loop automata: M^D, M^C, M^{CD}. When $L \neq 0$, both players have unique best responses for all strategies in PD, so by Proposition B.14 these are equilibria if and only a Nash equilibrium is played in every stage game. Therefore, when $L > 0$, the unique Nash equilibrium of the stage game PD is for both players to play D. So M^D is an equilibrium when $L > 0$.

There are 8 remaining automata in $\bar{\mathcal{M}}^2$. For each of these automata M, I find the stationary distributions and payoffs when matched with each of the other automata in $\bar{\mathcal{M}}^2$. Using the payoffs, I calculate the best response function for each of the remaining 8 automata over almost all of the parameter space (all but set of measure zero). I find that the only regions which $M_1 = BR_1(M_2)$ and $M_2 = BR_2(M_1)$ are those stated in the theorem. For conciseness, these stationary distributions are not included here, but are available on my website.

The only equilibrium that is supported by a set of positive measure from these remaining 8 automata is M^{WSLS} on the region $-(1-2\varepsilon) < L < (1-2\varepsilon)$. So the three equilibria from $\bar{\mathcal{M}}^2$ are M^C, M^D, and M^{WSLS}.

There are also automata which are payoff equivalent to some of these three automata. By Lemma B.16, every automaton which eventually always plays C is payoff equivalent to M^C. Therefore, any combination of automata which eventually always play C is an equilibrium in the region $L < 0$. Similarly, any pair of automata which eventually play D is an equilibrium in the region $L > 0$. Finally, there are no other two-state automata which are payoff equivalent to M^{WSLS}. Therefore, the set of two-state equilibria that are supported by a region of positive measure is:

1. Both automata eventually always play C is a symmetric equilibrium if and only if $L > 0$.

2. Both automata eventually always play D is a symmetric equilibrium if and only if $L < 0$.

3. M^{WSLS} if and only if $-(1-2\varepsilon)^3 \leq L \leq (1-2\varepsilon)^3$.

■

	C	D
C	1+L,1	0,0
D	0,0	1,1+L

(a) BOS Game

	C	D
C	1+L,1+L	-L,0
D	0,-L	0,0

(b) MECG Game

Theorem B.18. *In the infinitely repeated BOS game, when players have the simple signal function r_i^S and choose among the set of two-state automata, $\mathcal{M}^2$, the only non-open-loop robust robust equilibria are:*

1. $-\frac{(1-2\varepsilon)^2}{2\varepsilon(2-5\varepsilon+4\varepsilon^2)} < L < \frac{(1-2\varepsilon)^2}{1-4\varepsilon+10\varepsilon^2-8\varepsilon^3}$ *and* $M_i = M^{WSLS}$.

Proof

The proof for this Theorem follows the argument of the proof of Theorem 4.3. Details available upon request. ■

Theorem B.19. *In the infinitely repeated MECG game, when players have the simple signal function r_i^S and choose among the set of two-state automata, $\mathcal{M}^2$, the only non-open-loop robust equilibria are:*

1. $L > \frac{1-4\varepsilon+10\varepsilon^2-8\varepsilon^3}{2\varepsilon(1-2\varepsilon)^2}$ *and* $M_i = M^{LSWS}$,

2. $L > -\frac{1-8\varepsilon+14\varepsilon^2-8\varepsilon^3}{2(1-\varepsilon)(1-2\varepsilon)^2}$ *and* $M_i = M^{WSLS}$.

Proof

The proof for this Theorem follows the argument of the proof of Theorem 4.3. Details available upon request. ■

Theorem 4.4. *Suppose both players have simple signal function r_i^S. There exists some $\bar{\varepsilon} > 0$ such that $\left(M^{WSLS}, M^{WSLS}\right)$ is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$ only if for $i = 1, 2$,*

1. $u_i(C, C) > u_i(C, D)$.
2. $u_i(C, C) > \frac{u_i(D,C)+u_i(D,D)}{2}$.

Proof of Theorem 4.4

To prove this theorem, I use the sufficient conditions for equilibria provided in Theorem 5.8. This says that to be an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$,

1. For all communicating classes such that $\gamma(A_k) = \gamma^*$, $U_i^{CC}(A_k) = U_i^{AC}(a^*(M_{-i}))$, and
2. $$\frac{\partial U_i(M_1, M_2, \varepsilon)}{\partial \varepsilon} = \sup_{M \in \mathcal{M}^{SPM}(M_{-i})} \frac{\partial U_i(M_1, M, \varepsilon)}{\partial \varepsilon}$$

First assume that

$$u_i(C, C) > u_i(C, D) \tag{17}$$

and

$$u_i(C, C) > \frac{u_i(D, C) + u_i(D, D)}{2} \tag{18}$$

hold. I then show that the two sufficient conditions are satisfied, meaning $\left(M^{WSLS}, M^{WSLS}\right)$ is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$.

When both players play M^{WSLS}, the Markov chain for the system is,

$$X\left(M^{WSLS}, M^{WSLS}, \varepsilon\right) = \begin{bmatrix} (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & \varepsilon^2 \\ \varepsilon^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 \\ \varepsilon^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 \\ (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & \varepsilon^2 \end{bmatrix}$$

This system has one communicating class, A, consisting of the first state of the Markov chain. Since there is only one communicating class, it trivially minimizes stochastic potential. Therefore, it must be the case that the payoff in this communicating class is equal to the optimal absorbing class payoff. The payoff for the communicating class is,

$$U_i^{CC}(A) = u_i(C, C),$$

which is the stage-game payoff associated with joint action pair (C, C).

Next, I must calculate the optimal absorbing class payoff for M^{WSLS}. There are three extreme absorbing classes, such that any other absorbing class can be written as a convex combination of these extreme absorbing classes. So one of these has to be the optimal absorbing class.

1. $a_1(M^{WSLS}) = (\{q_1\}, \{C\})$ with payoff $u_i\left(a_1(M^{WSLS})\right) = u_i(C, C)$

2. $a_2(M^{WSLS}) = (\{q_1, q_2\}, \{D, D\})$ with payoff $u_i\left(a_2(M^{WSLS})\right) = \frac{u_i(D,C)+u_i(D,D)}{2}$

3. $a_3(M^{WSLS}) = (\{q_2\}, \{D\})$ with payoff $u_i\left(a_3(M^{WSLS})\right) = u_i(C, D)$

By (17) and (18), it is clear that $a_1\left(M^{WSLS}\right)$ is the optimal absorbing class. Therefore $U_i^{CC}(A) = U_i^{AC}\left(a_i^*\left(M^{WSLS}\right)\right)$, so the first condition is satisfied.

Next, I need to show that the marginal utility condition is satisfied. By Lemma B.17, the set of automata can be reduced some minimal payoff equivalent set. There are 11 remaining automata, call this set $\bar{\mathcal{M}}_2$. It can easily be verified that when M^{WSLS} is matched with any automaton $M \in \bar{M}_2$, then all communicating classes minimize stochastic potential.

There are only two automata, such that when paired with M^{WSLS}, all communicating classes yield the optimal absorbing class payoff. These are M^{WSLS} and $M^5 = (\{C, D\}, \{1, 1\}, \{2, 1\})$.

When both play M^{WSLS}, then the stationary distribution is,

$$\pi\left(M^{WSLS}, M^{WSLS}, \varepsilon\right) = \begin{bmatrix} 1 - 4\varepsilon + 7\varepsilon^2 - 4\varepsilon^3 \\ \varepsilon(1-\varepsilon) \\ \varepsilon(1-\varepsilon) \\ \varepsilon\left(2 - 5\varepsilon + 4\varepsilon^2\right) \end{bmatrix}'$$

By Lemma 5.3, the payoff is the stationary distribution dotted with the vector of payoffs,

$$U_i\left(M^{WSLS}, M^{WSLS}, \varepsilon\right) = \pi\left(M^{WSLS}, M^{WSLS}, \varepsilon\right) \cdot \boldsymbol{u},$$

where $\boldsymbol{u}$ is the vector of payoffs. Therefore the marginal utility at $\varepsilon = 0$ is,

$$\frac{\partial U_i\left(M^{WSLS}, M^{WSLS}, 0\right)}{\partial \varepsilon} = -4u_i(C, C) + u_i(C, D) + u_i(D, C) + 2u_i(D, D)$$

When player 1 plays M^{WSLS} and player 2 plays M^5, then the stationary distribution is,

$$\pi\left(M^{WSLS}, M^5, \varepsilon\right) = \frac{1}{1 + 2\varepsilon - 6\varepsilon^2 + 10\varepsilon^3 - 4\varepsilon^4} \begin{bmatrix} 1 - 3\varepsilon + 5\varepsilon^2 - 2\varepsilon^3 \\ \varepsilon\left(1 - 2\varepsilon + 3\varepsilon^2 - 2\varepsilon^3\right) \\ \varepsilon\left(2 - 3\varepsilon + 2\varepsilon^2\right) \\ \varepsilon\left(2 - 6\varepsilon + 7\varepsilon^2 - 2\varepsilon^3\right) \end{bmatrix}$$

Again by Lemma 5.3, the payoff is the dot product,

$$U_i\left(M^{WSLS}, M^5, \varepsilon\right) = \pi\left(M^{WSLS}, M^5, \varepsilon\right) \cdot \boldsymbol{u},$$

This means the marginal utility at $\varepsilon = 0$ is,

$$\frac{\partial U_i\left(M^{WSLS}, M^5, 0\right)}{\partial \varepsilon} = -5u_i(C, C) + u_i(C, D) + 2u_i(D, C) + 2u_i(D, D)$$

So

$$\frac{\partial U_i\left(M^{WSLS}, M^{WSLS}, 0\right)}{\partial \varepsilon} \geq \frac{\partial U_i\left(M^{WSLS}, M^{5}, 0\right)}{\partial \varepsilon} \iff u_i\left(C, C\right) > u_i\left(D, C\right)$$

This clearly holds by the assumption (17), and therefore both conditions are satisfied. So $\left(M^{WSLS}, M^{WSLS}\right)$ is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$ if the two conditions are satisfied. ■

C Examples

C.1 Stochastic Potential Example

To better understand the definitions used for the theorem, I provide a corollary which shows that both players playing tit-for-tat can never be an equilibrium in the finite-state case. Let M^{TFT} be the two-state tit-for-tat automaton from Figure 6(a). Suppose players use the simple signal function r_i^S from (1). Finally suppose that players play supergame G with the prisoner's dilemma stage-game payoffs displayed in Figure 1.

Corollary C.1. *Suppose players play super game G with stage game PD and signal functions r_i^S, there is no $\bar{\varepsilon} > 0$ such that the pair of automata $\left(M^{TFT}, M^{TFT}\right)$ is an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$.*

Proof

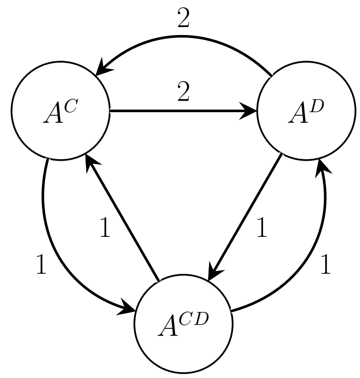
To prove this, I need to show that the necessary conditions from Theorem 5.6 are not satisfied. The Markov chain of this system is,

$$X\left(M_1, M_2, \varepsilon\right) = \begin{array}{c} x^{CC} \\ x^{CD} \\ x^{DC} \\ x^{DD} \end{array} \begin{bmatrix} (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & \varepsilon^2 \\ \varepsilon(1-\varepsilon) & \varepsilon^2 & (1-\varepsilon)^2 & \varepsilon(1-\varepsilon) \\ \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 & \varepsilon^2 & \varepsilon(1-\varepsilon) \\ \varepsilon^2 & \varepsilon(1-\varepsilon) & \varepsilon(1-\varepsilon) & (1-\varepsilon)^2 \end{bmatrix}$$

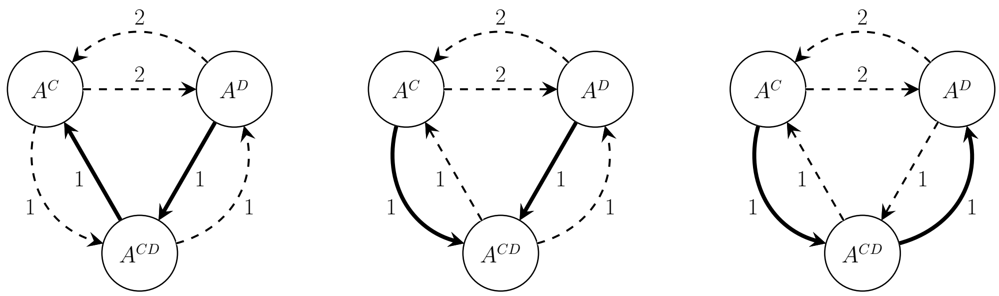
There are three communicating classes: $A^C = \left\{x^{CC}\right\}, A^{CD} = \left\{x^{CD}, x^{DC}\right\}, A^D = \left\{x^{DD}\right\}$. The resistance matrix R which tells the resistance between each communicating class is,

$$R = \begin{array}{c} A^{C} \\ A^{CD} \\ A^{D} \end{array} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

The entry in the first row, third column means that to probability of getting from A^C to A^D is order ε^2. The graph $\mathcal{G}$ with a vertex for each communicating class, and edge weights equal to the resistance between classes is displayed in Figure C.1(a). The optimal i-tree for each communicating class is displayed by the bold lines in Figure C.1(b)-(d). These graphs show that each communicating class has stochastic potential $\gamma_i = 2$. Therefore, the minimum stochastic potential for this system is $\gamma^* = 2$. By Theorem B.13, all communicating classes are prevalent. By Theorem 5.6, all prevalent communicating classes must yield the same payoff as the optimal absorbing class. The optimal absorbing class for each player yields payoff 1. Both players playing



(a) Resistance graph



(b) Optimal i-tree for A^C (c) Optimal i-tree for A^{CD} (d) Optimal i-tree for A^D

Figure 10: Resistance graph and optimal i-trees if both players play M^{TFT}.

M^{TFT} is only satisfies the necessary conditions if all communicating classes yield the same payoff, 1. Since $U_2\left(A^C\right) = 1$ and $U_2\left(A^D\right) = 0$, it is never possible for $\left(M^{TFT}, M^{TFT}\right)$ to be an equilibrium for all $\varepsilon \in (0, \bar{\varepsilon})$. ■

C.2 Constructed Automaton Example

Suppose that player 1 plays the three state automaton displayed in Figure 11. First, player 2 wants to determine the desired absorbing class. For automaton M_1, the optimal absorbing class based on the prisoner's dilemma game from 1 is $a^*(M_1) = \{\{q_1\}, \{C\}\}$. Assume that player 2 wants to create an automaton M_2 which only gets stuck in this absorbing class. This automaton has three regions as described above, and is displayed in Figure 12. First the absorbing region is simple, it consists of one state, q_1, which plays C and returns when M_1 plays C. It is clear that when M_2 is in q_1, and M_1 is in q_1, player 2 is in his optimal absorbing class. If there is an incorrect signal while in the absorbing region, player 2 loses track of the current state of M_2, and therefore moves to the homing region to determine the current state.

The homing sequence for this automaton is $h = C, D$. To see why this is a homing sequence, suppose automaton M_1 starts in state q_1. Player 2 is trying to determine the current state by playing the homing sequence. In the first period, M_1 plays C and player 2 plays C. Automaton M_1 returns to state q_1. In the second period M_1 plays C again and player 2 play D. So the output from automaton M_1 from the homing sequence is C, C. The other sequences of plays for the other

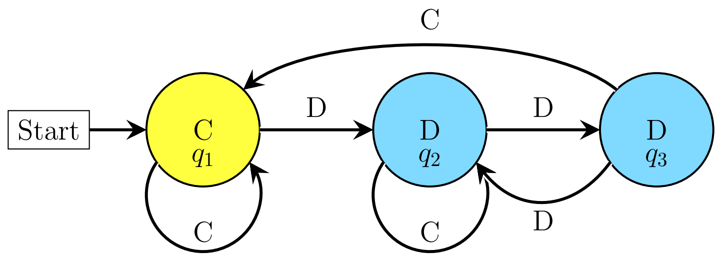


Figure 11: Homing sequence example: automaton M_1.

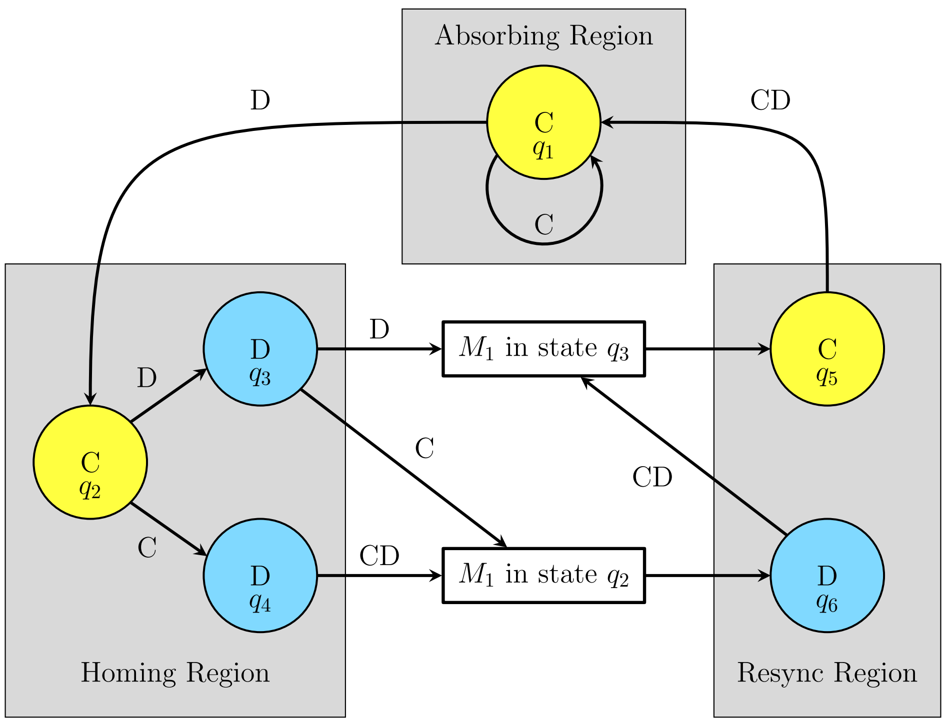


Figure 12: Homing sequence example: constructed automaton M_2.

starting states is displayed in the following table:

Starting State	**First Play**	**Second Play**	**Final State**
q_1	C	C	q_2
q_2	D	D	q_3
q_3	D	C	q_2

When player 2 plays the homing sequence and sees output C, C or D, C, M_1 must be in state q_2. When the output is D, D, M_1 must be in q_3. So based on this output, player 2 knows the current state of M_1. The second region of M_2 is the homing region. In the homing region, M_2 always plays the homing sequence, and leaves the homing sequence after it has played this sequence. The homing region in Figure 12 consists of states q_2, q_3, and q_4. In state q_2 the first term of the homing sequence is played, then depending on the output, M_2 moves to either state q_3 or q_4 where the second term of the homing sequence is played. The response from automaton M_1 after the homing region allows player 2 to know the current state of M_1. In this example, M_1 is either in state q_2 or q_3 after the homing region.

Finally, once the state has been determined, the automaton M_2 simply has to resynchronize the two automata back to the desired absorbing class $a_2^*(M_1)$. The resynchronization region consists of states q_5 and q_6. If M_1 is in state q_2, then automaton M_2 goes to state q_6. If automaton M_1 is in state q_3, then automaton M_2 goes to state q_5. After the resynchronization region, both automata are in state q_1, and they remain here until an incorrect signal is received.

References

Abreu, D., and A. Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56(6), 1259–1281.

Aumann, R. (1981): *Essays in game theory and mathematical economics in honor of Oskar Morgenstern*. Bibliographisches Institut.

Axelrod, R. (1980a): "Effective Choice in the Prisoner's Dilemma," *The Journal of Conflict Resolution*, 24(1), 3–25.

——— (1980b): "More Effective Choice in the Prisoner's Dilemma," *The Journal of Conflict Resolution*, 24(3), 379–403.

——— (2006): *The evolution of cooperation*. Basic Books, revised edn.

Bhaskar, V. (2000): "The robustness of repeated game equilibria to incomplete payoff information," *University of Essex*.

Bhaskar, V., and I. Obara (2002): "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, 102(1), 40 – 69.

Cole, H. L., and N. R. Kocherlakota (2005): "Finite memory and imperfect monitoring," *Games and Economic Behavior*, 53(1), 59 – 72.

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66(3), 597–626.

Compte, O., and A. Postlewaite (2009): "Effecting Cooperation," .

Conlisk, J. (1996): "Why Bounded Rationality?," *Journal of Economic Literature*, 34(2), 669–700.

Dal Bó, P., and G. R. Fréchette (2008): "The evolution of cooperation in infinitely repeated games: Experimental evidence," *Unpublished manuscript. Brown University and New . . . .*

Ely, J. C., J. Horner, and W. Olszewski (2005): "Belief-Free Equilibria in Repeated Games," *Econometrica*, 73(2), 377–415.

Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62(5), 997–1039.

Grinstead, C. M., and J. L. Snell (1997): *Introduction to probability*. American Mathematical Society, second revised edn.

Imhof, L. A., D. Fudenberg, and M. A. Nowak (2007): "Tit-for-tat or win-stay, lose-shift?," *Journal of Theoretical Biology*, 247(3), 574 – 580.

Ioannou, C. (2009): "Algorithmic Bounded Rationality, Noise And Optimality," .

Kandori, M. (2002): "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory*, 102(1), 1 – 15.

Kandori, M., and H. Matsushima (1998): "Private Observation, Communication and Collusion," *Econometrica*, 66(3), 627–652.

Kohavi, Z. (1978): *Switching and Finite Automata Thoery*. McGraw-Hill Inc., 2nd edn.

Lehrer, E. (1988): "Repeated games with stationary bounded recall strategies," *Journal of Economic Theory*, 46(1), 130 – 144.

Mailath, G., and L. Samuelson (2006): *Repeated games and reputations: long-run relationships*. Oxford University Press, USA.

Mailath, G. J., and S. Morris (2002): "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, 102(1), 189 – 228.

Miller, J. (1996): "The coevolution of automata in the repeated Prisoner's Dilemma," *Journal of Economic Behavior & Organization*.

Neyman, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economics Letters*, 19, 227–229.

Nowak, M., and K. Sigmund (1993): "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, 364(6432), 56–58.

Obara, I. (2009): "Folk theorem with communication," *Journal of Economic Theory*, 144(1), 120 – 134.

Piccione, M. (2002): "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, 102(1), 70 – 83.

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

Sabourian, H. (1998): "Repeated games with M-period bounded memory (pure strategies)," *Journal of Mathematical Economics*, 30(1), 1 – 35.

Selten, R., and T. Chmura (2008): "Stationary Concepts for Experimental 2x2-Games," *The American Economic Review*, 98(3).

Simon, H. (1957): *Models of man: Social and rational*. John Wiley.

Stigler, G. J. (1964): "A Theory of Oligopoly," *The Journal of Political Economy*, 72(1), 44–61.

Wedekind, C., and M. Milinski (1996): "Human Cooperation in the Simultaneous and the Alternating Prisoner's Dilemma: Pavlov Versus Generous Tit-For-Tat," *Proceedings of the National Academy of Sciences of the United States of America*, 93(7), 2686–2689.

Yamamoto, Y. (2009): "A limit characterization of belief-free equilibrium payoffs in repeated games," *Journal of Economic Theory*, 144(2), 802 – 824.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.