

# COOPERATION IS PROVABLY REQUIRED IN A VERSION OF THE NOISY ITERATED PRISONER'S DILEMMA

ARVID LUNNEMARK.

SUPERVISED BY: MICHAEL SIPSER

## 1. INTRODUCTION

The prisoner's dilemma is a classic symmetric two-player game, where both players acting in their own best interests leads to an outcome that is not optimal for either of them. There are two actions, cooperate ( $C$ ) and defect ( $D$ ), with the payoffs represented by the matrix

	$C$	$D$
$C$	$(R, R)$	$(S, T)$
$D$	$(T, S)$	$(P, P)$

where  $T > R > P > S$  (the row player gets the first in each pair of numbers). Given the other player's action, a player always maximizes their payoff by choosing  $D$ , but with both doing so, the players end up in the  $(D, D)$  state, which is worse for both than the  $(C, C)$  state since  $R > P$ . The game can be used to model many real-world situations, including countries failing to act to stop climate change, doping in sport, and economic competition.

To make the game more interesting, one can consider playing it multiple times in a row with multiple players. One might loosely connect the repeated multiplayer game to evolution: why have humans evolved to cooperate with each other? In a seminal paper in 1980, Axelrod presented evidence that in a repeated setting with multiple players facing off in a tournament, cooperation can arise as the strategy of choice even for a selfish player [1]. In particular, the mostly cooperating Tit-for-tat strategy, depicted in fig. 1, won his tournament. The idea is that even though Tit-for-tat loses when played against defecting strategies, that is outweighed by it being heavily rewarded when cooperating with other cooperating strategies.

Clearly, however, the winning strategy in a tournament depends on the composition of strategies in that tournament. If all strategies had been of the type to always defect, Tit-for-tat would not have won the Axelrod tournament. Therefore, follow-up tournaments and simulations have been run, and perhaps surprisingly, most provide further evidence that cooperation is the prevailing strategy, as Axelrod summarizes in a 1981 book [2]. There,

copy the table  
from axelrod's  
book because  
it's much bet-  
ter and much  
clearer!!!

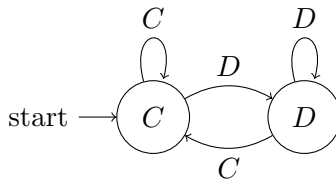


FIGURE 1. The Tit-for-tat strategy. At any one point, it is in one of two states, taking the action corresponding to the label of the state. Upon perceiving its opponent’s move, it decides to switch state if the opponent does something different.

he also introduces an evolutionary model that can be used for analyzing the game theoretically: strategies are considered to exist in a population that evolves over time in an evolutionary way (more successful strategies reproduce, and mutations can happen). In that context, a strategy is said to be evolutionarily stable if, supposing it controls a large share of the population, it resists being overtaken by any other strategy entering the population in small numbers. That is, populations of evolutionarily stable strategies form the equilibrium states of the evolutionary process, and thus, one may say that evolutionarily stable strategies are the “best” strategies.

In light of Axelrod’s tournaments highlighting the effectiveness of cooperation, it has long been a goal to prove that a cooperating strategy like Tit-for-tat is evolutionarily stable. A few variants of this has been shown: First, Nowak has shown that in the finitely repeated game, a strategy that always defects is sometimes not evolutionarily stable [5]. Second, Binmore has shown that in the infinitely repeated game where strategies are modeled as finite automata where having more states comes at a cost, a strategy needs to cooperate with itself to be stable [3]. And third, Fudenberg and Maskin have shown that a strategy needs to cooperate with itself to be stable also in the infinitely repeated game in the presence of a certain notion of infinitesimally small noise [4].

Notably, all these results impose additional restrictions on the setup, and as noted by Fudenberg and Maskin, one has to do that: in the deterministic infinitely repeated game, one could create a strategy that can self-identify — if the opponent ever deviates from the pattern, the strategy resorts to defecting for the rest of the game. This strategy will be evolutionarily stable, but is not cooperative.

In this paper, we, like Fudenberg and Maskin, choose the addition of noise as our restriction. In contrast to them, however, we present a model where, at every step, each strategy has a tiny probability  $p$  of doing the wrong thing, which is arguably the most natural way of modeling noise. In section 2, we define the setup of our version of the problem in detail. Then, in section 3, we state the two main results: that a strategy needs to be

cooperative to be evolutionarily stable, and that such evolutionarily stable strategies exist. In section 4, we prove our results. Finally, in section 5 we briefly discuss other potential ways of modeling the problem.

## 2. SETUP

In this section we define our setup. In summary, we consider the infinitely repeated prisoner's dilemma played by finite automata in infinite populations, evolving evolutionarily in the presence of noise.

**2.1. Formal Definition.** First, we define the reward function.

**Definition 2.1.** The *prisoner's dilemma* is a symmetric two-player game with two actions, cooperate ( $C$ ) and defect ( $D$ ), where, if player 1 selects action  $a$  and player 2 selects action  $b$ , player 1 gets the reward

$$r(a, b) = \begin{cases} R & \text{if } a = C, b = C \\ T & \text{if } a = D, b = C \\ S & \text{if } a = C, b = D \\ P & \text{if } a = D, b = D \end{cases}$$

We require  $T > R > P > S$  and  $2R > T + S$ .

When we study the *iterated* prisoner's dilemma, we want to look at strategies that determine their next move based on the history of previous moves. We restrict ourselves to strategies that can be implemented on a computer with finite memory.

**Definition 2.2.** A *strategy*  $s$  is a Moore machine (finite automaton with outputs) over the input and output alphabet  $\{C, D\}$ .

We will consider strategies in the presence of noise. To model that, we will assume that a strategy has a probability  $1 - p$  of following the correct transition and a probability  $p$  of following the incorrect transition, at every step. Note that this models noise in *perception*. One could also imagine modeling noise in *action taken*, but it is easy to see that the two are equivalent up to a change in the values of  $R, T, S$  and  $P$ .

We can now begin to define the outcome of two strategies playing against each other in the infinitely repeated game. We do this using Markov chains.

**Definition 2.3.** Suppose that strategy  $s_1$  plays against strategy  $s_2$ . This defines an  $s_1$ - $s_2$  *Markov chain* where each state is a tuple  $(c_1, c_2)$  with  $c_1$  being a state in  $s_1$  and  $c_2$  a state in  $s_2$ . The transition probabilities are defined in the obvious way:

$$P_{(c_1, c_2) \rightarrow (c'_1, c'_2)} = p_{c_1 \rightarrow c'_1} \cdot p_{c_2 \rightarrow c'_2},$$

where  $p_{c_1 \rightarrow c'_1}$  is  $1 - p$  if the output of  $s_2$  at state  $c_2$  causes  $c_1$  to transition to  $c'_1$ , and is  $p$  if the opposite of that output causes that transition;  $p_{c_2 \rightarrow c'_2}$  is defined similarly.

Markov chains naturally leads themselves to the study of limiting cases.

**Definition 2.4.** The *time average distribution* of the  $s_1$ - $s_2$  Markov chain given the start state  $(a, b)$ , denoted  $\pi^{(a,b)}$ , is the distribution such that

$$\pi_{c_1, c_2}^{(a,b)} = E[\text{fraction of time in state } (c_1, c_2) \mid \text{initial state is } (a, b)]$$

where the expected fraction of time is taken over the infinite sequence  $(X_0, X_1, \dots)$ .

We will often use  $\pi$  to refer to  $\pi^{(a_0, b_0)}$  where  $a_0$  is the initial state of  $s_1$  and  $b_0$  is the initial state of  $s_2$ .

**Definition 2.5.** The payoff that  $s_1$  gets when playing against  $s_2$  is

$$v_{s_1}(s_2) = \sum_{(c_1, c_2)} \pi_{c_1, c_2} \cdot r(G_{s_1}(c_1), G_{s_2}(c_2))$$

where  $G_s(c)$  is the output of strategy  $s$  at state  $c$ . The sum is taken over all states  $(c_1, c_2)$  in the  $s_1$ - $s_2$  Markov chain.

For notational convenience, we may also make  $r(G_{s_1}(c_1), G_{s_2}(c_2))$  into a vector, denoted by  $r$ , and write this as the dot product

$$v_{s_1}(s_2) = \pi \cdot r.$$

That is, the payoff that  $s_1$  gets when playing against  $s_2$  is simply an average of the reward it gets in each possible state of the Markov chain weighted by the fraction of time that is spent there.

We're now ready to look at how strategies interact.

**Definition 2.6.** A *population* of strategies  $P = (S, f)$  is a set  $S$  of strategies and a function  $f : S \rightarrow (0, 1]$  such that  $\sum_{s \in S} f(s) = 1$ , representing the frequency of each strategy in the population.

**Definition 2.7.** The *fitness* of a strategy  $s$  in a population  $P = (S, f)$  is

$$F(s) = \sum_{s' \in S} f(s') v_s(s').$$

One can think of this as saying that we have infinitely many members of the population, interacting with each other evenly, and that the fitness of a strategy is its expected payoff. Having infinitely many interactions like this justifies the usage of expectation when defining  $v_{s_1}(s_2)$ .

We can now use the fitness of a strategy to compare it with other strategies in the same population. If a strategy  $s_1$  has a higher fitness than another strategy  $s_2$ , we say that the frequency of  $s_1$  will increase at the expense of the frequency of  $s_2$ , in the next step of the evolutionary process. This is getting us close to how we want to define stable strategies; our next move is looking not only at a single evolutionary step, but the entire evolutionary process.

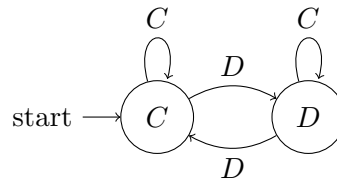


FIGURE 2. The Pavlov strategy. Similar to but not the same as the Tit-for-tat strategy.

**Definition 2.8.** A strategy  $s_1$  is  $\epsilon$ -invadable if there exists a strategy  $s_2$  such that in all populations  $P$  with  $S = \{s_1, s_2\}$  and  $f(s_2) \geq \epsilon$ , we have

$$F(s_2) > F(s_1)$$

That is, if  $s_1$  is  $\epsilon$ -invadable, there exists a strategy  $s_2$  that can start as only a tiny fraction  $\epsilon$  of the total population, and consistently have higher fitness than  $s_1$ , eventually causing  $s_2$  to overtake and then eliminate  $s_1$  completely. We are now finally ready to define evolutionary stability.

**Definition 2.9.** A strategy  $s_1$  is *evolutionarily stable* if there exists parameters  $p_0$  and  $\epsilon_0$  with  $0 < p_0, \epsilon_0 < 1$ , such that for all  $p < p_0$ , and all  $\epsilon < \epsilon_0$ ,  $s_1$  is not  $\epsilon$ -invadable.

That is, a strategy  $s_1$  is evolutionarily stable if, as the noise probability  $p$  goes to 0, it can withstand invasion attempts from any strategy that starts off as a tiny fraction of the population.

**2.2. Interpretation.** Having concluded our formal definition of evolutionary stability, it might be helpful to see examples of concrete setups in which this definition makes sense.

Consider the following world.

### 3. RESULTS

We can now state our results! Together, the following two conjectures imply that in the setup described here, mutual cooperation arises as the only stable choice.

**Conjecture 3.1.** Suppose that a strategy  $s_1$  is evolutionarily stable. Then  $\lim_{p \rightarrow 0} v_{s_1}(s_1) = R$ .

**Conjecture 3.2.** Suppose  $2R > T + P$ . Then, the Pavlov strategy, depicted in section 3, is evolutionarily stable.

*Remark.* Tit-for-tat, depicted in fig. 1, is not evolutionarily stable. When played against itself, it ends up spending just as much time in the defection state as in the cooperation state, because as soon as one mistake is made, it goes into a mutual defection cycle with its clone that is only broken by an additional mistake. Thus,  $v_s(s)$  where  $s$  is Tit-for-tat is much smaller than  $R$ , which means that it is not evolutionarily stable by conjecture 3.1.

separate this definition out into a pure evolutionarily stable definition, and then an evolutionarily stable for infinitesimally small probabilities definition?? would make interpretation much clearer.

WRITE THIS SECTION?? I'M NOT SURE THAT MY MODEL MAKES MUCH SENSE ANYMORE THOUGH, BECAUSE IT REQUIRES PEOPLE TO PLAY AGAINST THEMSELVES WHICH IS A NONO. also the probability going to 0 does

#### 4. PROOFS

In this section, we prove our results from the previous section.

**4.1. The time average distribution.** Before we prove conjecture 3.1, we need to understand what the payoff  $v_{s_1}(s_2)$  really means. In this subsection, we prove a series of lemmas that characterize the time average distribution, and consequently  $v_{s_1}(s_2)$ .

**Lemma 4.1.** The time average distribution  $\pi^{(a,b)}$ , for any starting state  $(a, b)$ , is a stationary distribution of the Markov chain.

We state this lemma without proof, as it is a fairly standard result. Important to note is that the time average distribution is not necessarily a *unique* stationary distribution, as we make no assumptions that our Markov chain be ergodic.

**Definition 4.2.** A *strongly connected component* of a directed graph is a subgraph where there is a path from every node to every other node.

**Definition 4.3.** An *absorbing component* of a directed graph is a subgraph where there are no edges from vertices inside the component to vertices outside it.

We may also put both of the terms together and talk about absorbing strongly connected components, which, as shown by the next few lemmas, are useful.

**Lemma 4.4.** An absorbing strongly connected component has a unique time average distribution, i.e., the time average distribution does not depend on the start state.

**Lemma 4.5.** Let  $\mathcal{S}$  be the set of absorbing strongly connected components of the  $s_1$ - $s_2$  Markov chain. For every  $C \in \mathcal{S}$ , let  $\pi^{(C)}$  be its unique time average distribution. Then, for some probabilities  $p_C$  with  $\sum_{C \in \mathcal{S}} p_C = 1$ , depending only on the condensation of the Markov chain, we have

$$\pi = \sum_{C \in \mathcal{S}} p_C \cdot \pi^{(C)}.$$

Finally, we can characterize the absorbing strongly connected components of the  $s_1$ - $s_2$  Markov chain in terms of the absorbing SCCs of  $s_1$  and  $s_2$  separately.

**Lemma 4.6.** Let  $\mathcal{C}_1$  be the set of absorbing SCCs of  $s_1$ , and similarly, let  $\mathcal{C}_2$  be the set of absorbing SCCs of  $s_2$ . If  $\mathcal{C}$  is the set of absorbing SCCs of the  $s_1$ - $s_2$  Markov chain, then it is the cartesian product of the two, i.e.,

$$\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2.$$

**Lemma 4.7.** The limit

$$\lim_{p \rightarrow 0} v_{s_1}(s_2)$$

exists, for any strategies  $s_1$  and  $s_2$ .

*Proof.* By definition,

$$v_{s_1}(s_2) = \pi \cdot r.$$

By lemma 4.1,  $\pi$  is a stationary distribution. In particular, if  $M$  is the transition matrix for the  $s_1$ - $s_2$  Markov chain, then  $\pi$  is an eigenvector of  $M$  with eigenvalue 1. This implies that  $\pi$  is in the nullspace of  $M - I$ . Thus, we can find  $\pi$  by solving for  $X$  in  $(M - I)X = 0$ . If we solve this using Gaussian elimination and back substitution, it is clear that the entries of  $\pi$  will be on the form  $\frac{f(p)}{g(p)}$  where  $f$  and  $g$  are polynomials in  $p$ , since each entry of  $M$  is a second-degree polynomial in  $p$ . This is continuous for all  $p$  where  $g(p) \neq 0$ , and since  $g$  will have a finite degree it is therefore continuous in a small right neighborhood of 0. Finally, note that  $v_{s_1}(s_2)$  is between  $S$  and  $T$ , which in conclusion means that the limit of it as  $p$  goes to 0 tends to a finite number, as desired.  $\square$

**Lemma 4.8.** For any strategy  $s$ ,

$$v_s(s) \leq R$$

*Proof.* For notational simplicity, we will let  $s_1$  and  $s_2$  be two copies of strategy  $s$ . Then,  $v_s(s) = v_{s_1}(s_2) = v_{s_2}(s_1)$ . By definition, we have

$$v_{s_1}(s_2) = \sum \pi_{c_1, c_2} \cdot r(c_1, c_2)$$

and

$$v_{s_2}(s_1) = \sum \pi_{c_2, c_1} \cdot r(c_2, c_1).$$

Note that  $\pi_{c_1, c_2}$  and  $\pi_{c_2, c_1}$  refer to the same state, so we thus have

$$v_{s_1}(s_2) + v_{s_2}(s_1) = \sum \pi_{c_1, c_2} \cdot (r(c_1, c_2) + r(c_2, c_1))$$

which implies that

$$v_s(s) = \sum \left( \pi_{c_1, c_2} \cdot \frac{r(c_1, c_2) + r(c_2, c_1)}{2} \right).$$

Now, note that  $r(c_1, c_2) + r(c_2, c_1) \in \{R + R, S + T, T + S, P + P\}$ . Since  $P < R$  and  $T + S < 2R$ , we thus find that

$$v_s(s) \leq \sum \pi_{c_1, c_2} \cdot R = R \sum \pi_{c_1, c_2} = R,$$

as desired.  $\square$

**4.2. Evolutionary Stability Implies Utilitarianism.** The following proof is relatively involved, and we believe that a simpler proof exists, but haven't been able to find one.

*Proof of conjecture 3.1.* Suppose that the strategy  $s_1$  is such that it is *not* true that

$$\lim_{p \rightarrow 0} v_{s_1}(s_1) = R$$

By lemma 4.7 and lemma 4.8, this assumption implies that the limit is strictly less than  $R$ . Define  $\gamma = \lim_{p \rightarrow 0} v_{s_1}(s_1)$ . Then,

$$\gamma < R.$$

We want to prove that  $s_1$  is not evolutionarily stable.

To do that, we want to prove that for all  $p_0, \epsilon_0 \in (0, 1)$ , there exists  $p < p_0$  and  $\epsilon < \epsilon_0$ , such that  $s_1$  is  $\epsilon$ -invadable. We choose  $\epsilon = \epsilon_0/2$ , and present a strategy  $s_2$  that can invade  $s_1$  for all  $p$  that are sufficiently small.

The underlying idea is to create  $s_2$  such that  $s_1$  can see no difference between itself and  $s_2$ , while  $s_2$ , on the other hand, can. In that case, we would have that  $v_{s_1}(s_2) = v_{s_2}(s_1) = v_{s_1}(s_1)$ , and could construct  $s_2$  such that it always cooperates when it recognizes itself, thereby yielding  $v_{s_2}(s_2) = R$ . This would give  $s_2$  a higher fitness than  $s_1$ . The rest of this proof executes this plan in detail.

We create the strategy  $s_2$  as follows. First, we copy all of  $s_1$  into  $s_2$ . Let  $\mathcal{C}_1$  be the set of absorbing strongly connected components of  $s_1$ . The key idea, now, is to replace each original absorbing strongly connected component  $C_1 \in \mathcal{C}_1$  with a new absorbing component  $C_2$ , which has the capability of self identifying.

We now construct  $C_2$ , given  $C_1$ . Figure 3 shows the high-level construction of  $C_2$ . All of the steps succeed with probability  $1 - O(p)$ . The first step makes sure that  $C_2$  does not deviate from  $s_1$  at all, until  $s_1$  has reached an absorbing strongly connected component. Once  $s_1$  is in an absorbing strongly connected component, we already know that the time average distribution does not depend on the particular state in  $s_1$ , which means that even if  $s_1$  now “figures out” that it is not playing against itself anymore, there is nothing it can do about it. The next step makes sure that, if  $s_2$  is playing against itself, it is in sync; i.e., with probability  $1 - O(p)$  they will both copies of  $s_2$  will be in the same state after this step. The third step now figures out what action  $s_1$  would take at some large finite time  $T$ , and then outputs the exact opposite action at time  $T$ . At that point, if its opponent outputs  $s_1$ ’s expected action, it transitions into the absorbing strongly connected component  $C_1$ , but if its opponent outputs the opposite of that action, it has successfully recognized itself and enters an always cooperating state.

This concludes the construction of  $s_2$ . We now want to prove that this  $s_2$  can in fact invade  $s_1$ . To do that, we will first assume that the below 3 claims are correct, i.e., that the construction is possible and works as specified, and finish the proof of our theorem using those assumptions. Then, we will prove the 3 claims.

**Claim 4.9.** We can construct a non-absorbing component with indistinguishable output from the component  $C_1$ , such that when we leave the component,  $s_1$  will be in an absorbing strongly connected component with probability  $1 - O(p)$ .



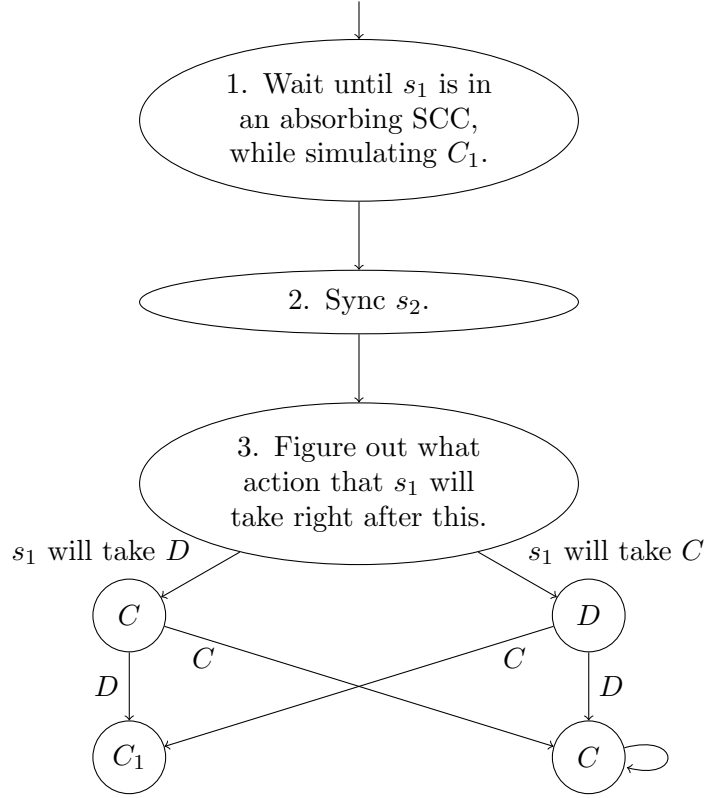


FIGURE 3. High level construction of  $C_2$ , the absorbing component replacing  $C_1$  when creating  $s_2$  in the proof of conjecture 3.1. Note that  $C_2$  contains exactly 2 absorbing strongly connected components.

**Claim 4.10.** We can construct a non-absorbing component placed after the first component, such that if  $s_2$  plays against itself, both clones transition out of the component at the exact same time with probability  $1 - O(p)$ .

**Claim 4.11.** We can construct a non-absorbing component with a finite number of maximum steps that, supposing that  $s_2$  plays against  $s_1$ , can figure out what action  $s_1$  will take at the exact timestep following the transition out of the component.

We now finish our proof.

**Claim 4.12.** Given the above construction of  $s_2$ , the payoffs are as follows.

$$\begin{aligned}
 v_{s_1}(s_1) &= \gamma + O(p) \\
 v_{s_1}(s_2) &= \gamma + O(p) \\
 v_{s_2}(s_1) &= \gamma + O(p) \\
 v_{s_2}(s_2) &= R + O(p)
 \end{aligned}$$

*Proof.* That  $v_{s_1}(s_1) = \gamma + O(p)$  follows by our definition of  $\gamma$ .

Consider first the case of  $s_1$  playing against  $s_2$ . By claim 4.9, with probability  $1 - O(p)$ ,  $s_2$  is indistinguishable from  $s_1$  until  $s_1$  reaches an absorbing strongly connected component  $C'_1$ . This means that the probability for reaching the absorbing component  $(C'_1, C_2)$  in the  $s_1$ - $s_2$  Markov chain, is up to a factor of  $1 - O(p)$  the same as if we hadn't replaced  $C_1$  by  $C_2$ , which is just the probability of reaching the absorbing component  $(C'_1, C_1)$  in the  $s_1$ - $s_1$  Markov chain. Thus,

$$\begin{aligned} P(s_1\text{-}s_2 \text{ chain ends up in } (C'_1, C_2)) \\ = (1 - O(p))P(s_1\text{-}s_1 \text{ chain ends up in } (C'_1, C_1)) \end{aligned}$$

Now, by claim 4.11, the probability that  $s_2$  enters the absorbing strongly connected component  $C_1$  after having gotten to  $C_2$  is  $1 - O(p)$ , and that it enters the always cooperating component  $C_c$  is  $O(p)$ . Thus,

$$\begin{aligned} P(s_1\text{-}s_2 \text{ chain ends up in } (C'_1, C_1)) &= P(s_1\text{-}s_1 \text{ chain ends up in } (C'_1, C_1)) + O(p) \\ P(s_1\text{-}s_2 \text{ chain ends up in } (C'_1, C_c)) &= O(p) \end{aligned}$$

That is, when  $s_1$  plays against  $s_2$ , the absorbing strongly connected components are mostly the same, with mostly the same probabilities. Now, we can calculate  $\tau$ , the time average distribution of the  $s_1$ - $s_2$  Markov chain:

$$\begin{aligned} \tau &= \sum_{(C_1, C') \in \mathcal{C}_1 \times \mathcal{C}_2} p_{(C_1, C')} \cdot \pi^{(C_1, C')} \\ &= \sum_{(C_1, C_{1, \text{copy}})} p_{(C_1, C_{1, \text{copy}})} \cdot \pi^{(C_1, C_{1, \text{copy}})} + \sum_{(C_1, C_c)} p_{(C_1, C_c)} \cdot \pi^{(C_1, C_c)} \\ &= \sum_{(C_1, C'_1) \in \mathcal{C}_1 \times \mathcal{C}_1} p_{(C_1, C'_1)} \cdot \pi^{(C_1, C'_1)} + O(p) \\ &= \pi + O(p) \end{aligned}$$

this is nonsensical here!! i should really try to explain this better!!!!

we're kinda abusing notation here

where the second-to-last equality follows by ??. This proves that  $v_{s_1}(s_2) = \gamma + O(p)$  and  $v_{s_2}(s_1) = \gamma + O(p)$ .

For the last part of the claim, consider the  $s_2$ - $s_2$  Markov chain. By claim 4.10, both clones will be in the exact same position after the sync step. The third step takes only a finite number of steps by claim 4.11. Thus, after leaving the third step of  $C_2$ , if no mistakes have happened since leaving state 2, both clones of  $s_2$  will be in the exact same state. This happens with probability  $(1 - p)^N$  where  $N$  is the finite number of states in step 3, and thus with probability  $1 - O(p)$ . Thus, in the identifying stage, both clones will output the same thing, and as we can see in fig. 3, they will then both transition to the always cooperating state with probability  $1 - O(p)$ . Therefore,  $v_{s_2}(s_2) = R - O(p)$ , which finishes the proof of the claim. ■

Given claim 4.12, we simply compute  $F(s_2) - F(s_1)$ , which we want to show is greater than 0.

$$\begin{aligned} F(s_2) - F(s_1) &= \\ &= (1 - \epsilon) \cdot v_{s_2}(s_1) + \epsilon \cdot v_{s_2}(s_2) - (1 - \epsilon) \cdot v_{s_1}(s_1) - \epsilon \cdot v_{s_1}(s_2) \\ &= (1 - \epsilon)(\gamma - O(p)) + \epsilon(R - O(p)) - (1 - \epsilon)\gamma - \epsilon(\gamma + O(p)) \\ &\leq \epsilon(R - \gamma) + O(p) \end{aligned}$$

We know that  $R - \gamma > 0$  by our initial assumption. For a sufficiently small  $p$ , thus,  $F(s_2) - F(s_1) > 0$ . This proves that  $s_2$  can invade  $s_1$ , and thus, that  $s_1$  is not  $\epsilon$ -invadable for this value of  $p$ . In conclusion, then,  $s_1$  is not evolutionarily stable, which concludes the proof of conjecture 3.1.  $\square$

We now return to the three claims we left out that detail the construction of  $s_2$ .

**Claim 4.13.** There is a  $T$  and a procedure implementable in  $C_2$  that runs in a finite number of steps and can determine the action  $A$  of  $s_1$  at time  $T$  with probability  $1 - O(p)$ .

*Proof.* Let  $\mathcal{C}_1$  be the set of all absorbing strongly connected components in  $s_1$ . Let  $M$  be the set of all tuples of  $(C, c)$  where  $C \in \mathcal{C}_1$  and  $c$  is a possible state of  $C$ . Let  $M'$  be a subset of  $M$  where a tuple  $(C, c)$  has been removed if there is another tuple  $(C', c') \in M'$  produces exactly the same output on all inputs, when only following the  $1 - p$  transitions.

The procedure works as follows:

- (1) Simulate  $C_1$  for  $N$  steps. (This can be done by e.g. duplicating  $C_1$   $N$  times and having all edges advance to corresponding state in the next copy.) Choose  $N$  so that after  $N$  timesteps,  $s_1$  will be in an absorbing strongly connected component with probability  $1 - O(p)$ .
- (2) Iterate over every pair of tuples  $(C, c)$  and  $(C', c')$  in  $M$ :
  - (a) There is a string  $S$  of actions on which  $C$  starting in  $c$  and  $C'$  starting in  $c'$  will produce a different output at a time  $t$ . Now, output the string  $S$ , and determine if the sequence of perceived actions corresponds to  $(C, c)$  or  $(C', c')$ .

Choose  $T = \max t$  over all  $t$ . With probability  $1 - O(p)$ , there will be exactly 1 pair  $(C, c)$  that corresponds to all perceived actions in each of its comparisons. We know what  $C$ , when starting in  $c$ , outputs at time  $T$ . Thus, with probability  $1 - O(p)$ , we know what action  $A$  that  $s_1$  outputs at time  $T$ .

It should be noted that this procedure is easily implementable on a finite automaton, using e.g. a decision tree structure.  $\square$

*Proof.*

move this out of the claim??? but it's annoying and needs a space somewhere

**KEY IMPORTANT DETAIL:** we need to say that at a large finite  $T$  then  $s_1$  will be within an absorbing strongly connected component, proba-

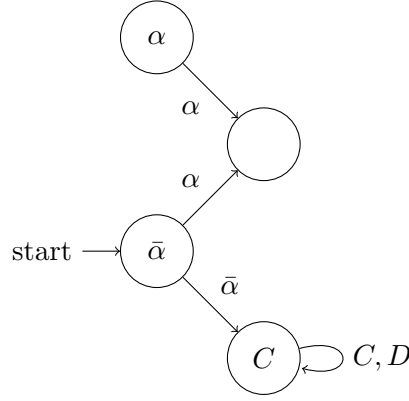


FIGURE 4. Constructon of invasion strategy, used in the proof of conjecture 3.1

**Claim 4.14.** Suppose that there are no mistakes, i.e., that  $p = 0$ . We can create a  $C_2$  such that

First, copy the entire  $s_1$  machine into  $s_2$ . Suppose that the state corresponding to the start state of  $s_1$  is  $c$ , and that the output at  $c$  is  $\alpha$ , and that the state  $s$  goes to upon perceiving the opponent move  $\alpha$  is  $c' = T(c, \alpha)$ . Now, create two new states:  $c_0$  and  $c_1$ . Define the transitions as

$$T(c_0, \alpha) = c'$$

$$T(c_0, \bar{\alpha}) = c_1$$

$$T(c_1, \cdot) = c_1$$

and the outputs as

$$G(c_0) = \neg G(c_2)$$

$$G(c_1) = C.$$

Let the start state of  $s_2$  be  $c_0$ .

**Claim 4.15.** Given the above construction of  $s_2$ , the following inequalities hold:

$$v_{s_1}(s_1) \leq (1-p)^2\gamma + 2(1-p)pR + p^2R$$

$$v_{s_1}(s_2) \leq (1-p)\gamma + pT$$

$$v_{s_2}(s_1) \geq (1-p)\gamma + pS$$

$$v_{s_2}(s_2) \geq (1-p)^2R + 2(1-p)p(\frac{S+T}{2}) + p^2\gamma.$$

Before proving this claim, we will use it to finish our proof of conjecture 3.1.

Now, we simply compute  $F(s_2) - F(s_1)$ , which we want to show is greater than 0.

$$\begin{aligned} F(s_2) - F(s_1) &= \\ &= (1 - \epsilon) \cdot v_{s_2}(s_1) + \epsilon \cdot v_{s_2}(s_2) - (1 - \epsilon) \cdot v_{s_1}(s_1) - \epsilon \cdot v_{s_1}(s_2) \\ &= (1 - \epsilon)(\gamma + p(\dots)) + \epsilon(R + p(\dots)) - (1 - \epsilon)(\gamma + p(\dots)) - \epsilon(\gamma + p(\dots)) \\ &= \epsilon(R - \gamma) + p(\dots) \end{aligned}$$

We know that  $R - \gamma > 0$  by our initial assumption. Clearly, since  $(\dots)$  is some polynomial in  $p$ , given an  $\epsilon$  we can find a sufficiently small  $p$  such that the full expression is positive. This proves that  $s_2$  can invade  $s_1$ , and thus, that  $s_1$  is not  $\epsilon$ -invadable for this value of  $p$ . In conclusion, then  $s_1$  is not evolutionarily stable, which concludes the proof of conjecture 3.1.  $\square$

*Proof of claim 4.12.* We can prove this using either of the two definitions.  $\square$

**4.3. Evolutionarily Stable Strategies Exist.** Unfortunately, we have no proof of conjecture 3.2.

We note that the  $2R > T + P$  condition is necessary. Otherwise, the AllD strategy would be able to invade Pavlov. We see this by noting that  $\lim_{p \rightarrow 0} v_{s_2}(s_1) = T + P$  if  $s_2$  is AllD and  $s_1$  is Pavlov, and that AllD is better against itself than Pavlov is against it.

## 5. DISCUSSION OF MODEL

**5.1. Other Potential Models.** Right now we have only modeled noise in perception. One could think of another possible kind of noise: a “failure of the mind,” which perhaps could be modeled instead by a probability  $p$  of being transported to any random state, instead. This would create ergodicity which is nice.

## 6. APPENDIX: PROBABLISTIC AUTOMATA

OHHHHHHH. THIS WILL HELP WITH MY PROOF????? THIS IS EXACTLY WHAT I'M TRYING TO DO IN MY PROOF RIGHT????? YEAHHHH, NO.

In this paper, we have considered strategies that make a deterministic move based on what they perceive. One could also imagine strategies that attaches a certain probability distribution to a perceived input, and chooses their next action based on that. In this appendix we show that these can be reduced to the deterministic ones, and thus that all results for the deterministic ones also hold for the probabilistic ones.

Proof idea: we can use cycles in the Markov chain with  $n$  total outputs,  $x$  of which are to state 1, to model getting to state 1 with probability  $x / n$ . this assumes that the outputs are of low enough probability, which can be achieved by chaining together lots of  $(1-p)$  transitions, which go to 0.

the hard part of this is showing that modifying the finite automaton like this won't hurt us. in fact, it would certainly not hurt us if not every state had to give an output. but that doesn't work for our model i think. so there are certainly things to think about here.

#### REFERENCES

- [1] AXELROD, R. Effective choice in the prisoner's dilemma. *Journal of conflict resolution* 24, 1 (1980), 3–25.
- [2] AXELROD, R., AND HAMILTON, W. D. The evolution of cooperation. *science* 211, 4489 (1981), 1390–1396.
- [3] BINMORE, K. G., AND SAMUELSON, L. Evolutionary stability in repeated games played by finite automata. *Journal of economic theory* 57, 2 (1992), 278–305.
- [4] FUNDENBERG, D., AND MASKIN, E. Evolution and cooperation in noisy repeated games. *The American Economic Review* 80, 2 (1990), 274–279.
- [5] NOWAK, M. A., SASAKI, A., TAYLOR, C., AND FUDENBERG, D. Emergence of co-operation and evolutionary stability in finite populations. *Nature* 428, 6983 (2004), 646–650.