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

ARVID LUNNEMARK.

SUPERVISED BY: MICHAEL SIPSER

1. Introduction

Introduce the prisoner's dilemma, the iterated prisoner's dilemma and the goal of finding out how cooperation can appear. Introduce Axelrod along with some other previous results.

Should some of the definitions be in the introduction?

clean up the julia notebook and refer to it for people who want to play with the setup

2. Setup

In this section we thoroughly define our problem setup.

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 $2R > T + S$.

A common choice in simulations of the iterated prisoner's dilemma is $T = 5$, $R = 3$, $P = 1$ and $S = 0$.

We want to study the *iterated* prisoner's dilemma, for which we can define strategies that determine their next move based on the history of previous moves. As discussed previously, we want to restrict ourselves to strategies with finite memory.

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

Notation-wise, we will use c to denote states in s, $G_s(c)$ to denote the output at state c, and $T_s(c, a)$ to denote the state that c transitions to upon receiving input a. For simplicity, we will also define the $\bar{}$ operator such that $\bar{C} = D$ and $\bar{D} = C$.

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 of the values of R, S, T, P.

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 x is the vector (c_1, c_2) where c_1 is a state in s_1 and c_2 is a state in s_2. The transition probabilities are defined in the obvious way, using the error probability p.

We use the notation $G_{s_1,s_2}(c_1, c_2)$ to refer to the vector $(G_{s_1}(c_1), G_{s_2}(c_2))$, and we use S_{s_1,s_2} to refer to the set of all states in the Markov chain.

Definition 2.4. Let X_t be the random variable designating which state the s_1-s_2 Markov chain is in at time t. The payoff of strategy s_1 when played against strategy s_2 is

$$v_{s_1}(s_2) = E\left[\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} r(G_{s_1,s_2}(X_t)) \;\middle|\; X_0 = (c_{\text{start}}(s_1), c_{\text{start}}(s_2))\right]$$

That is, when s_1 plays against s_2, we define its payoff to be the average payoff over all possible infinite sequences of moves. Note that the expectation is taken over the infinite sequence $(X_0, X_1, \ldots)$. The limit inside is thus simply a normal time average limit of bounded real numbers, which clearly exists.

We will now introduce the notion of a time average distribution which will lead us to a second way of defining the payoff $v_{s_1}(s_2)$.

is this really necessary?? should i pick one or the other?? should i move one of them to the next section? which one is easier to understand? do they trivially say the same thing?

Definition 2.5. 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

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

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

We will use π to refer to $\pi^{(c_{\text{start}}(s_1), c_{\text{start}}(s_2))}$.

Lemma 2.6. The payoff when s_1 plays against s_2 is

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

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

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

I think the best way to handle this is to remove the expected value definition altogether. just do the time average

Proof of lemma 2.6. The key idea is that a time average sum where each element is one of finitely many values can be written as a frequency-weighted finite sum instead. Let $I_{c_1,c_2,t}$ be the indicator variable that is 1 if $G_{s_1,s_2}(X_t) = (c_1, c_2)$ and 0 otherwise. Then, we can write

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} r(G_{s_1,s_2}(X_t)) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1, c_2) \cdot I_{c_1,c_2,t}$$

We may now exchange the order of summation and move the finite sum out of the limit, to get

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1, c_2)\cdot I_{c_1,c_2,t} = \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1, c_2)\cdot \lim_{T\to\infty} \sum_{t=0}^{T} \frac{I_{c_1,c_2,t}}{T}$$

We can now use definition 2.4 and linearity of expectation to find that

$$v_{s_1}(s_2) = \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1, c_2) E\left[\lim_{T\to\infty} \sum_{t=0}^{T} \frac{I_{c_1,c_2,t}}{T} \mid X_0 = (c_{\text{start}}(s_1), c_{\text{start}}(s_2))\right]$$

Finally, we note that this is exactly the statement of lemma 2.6, which proves our lemma. □

Appendix A contains more details on time average distributions. In particular, if a unique stationary distribution exists, it is equal to the time-average distribution, which enables us to quickly find the time-average distribution in many cases.

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

Definition 2.7. A *population* of strategies $P = (S, f)$ is a set S of strategies and a function $f : S \to (0, 1]$ such that $\sum_{s\in S} f(s) = 1$, representing the frequency of each strategy in the population.

Definition 2.8. 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, and that they all interact with everyone else. 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, that means that the frequency of s_1 will increase on 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.

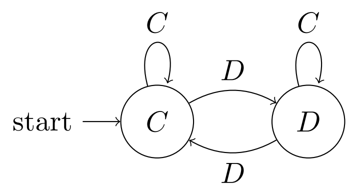

FIGURE 1. Pavlov.

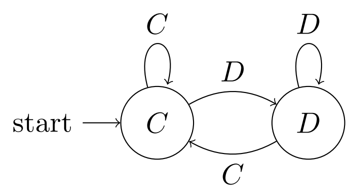

FIGURE 2. Tit for tat.

Definition 2.9. A strategy s_1 is *ϵ-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 ϵ-invadable, there exists a strategy s_2 that can start as only a tiny fraction ϵ of the total population, and consistently have higher fitness than s_1, eventually causing overtaking s_1 completely. We are now finally ready to state our main definition.

Definition 2.10. A strategy s_1 is *evolutionarily stable* if there exists parameters p_0 and ϵ_0, both in $(0, 1)$, such that for all $p < p_0$, and all $\epsilon < \epsilon_0$, s_1 is not ϵ-invadable.

That is, a strategy s_1 is evolutionarily stable if it can withstand invasion attempts from any strategy that starts off in low numbers, as the probability of noise tends to 0.

3. RESULTS

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

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

Conjecture 3.2. Suppose $2R > T + P$. Then, the Pavlov strategy is evolutionarily stable.

Remark. Tit-for-tat, displayed in section 3, is not evolutionarily stable. It has the stationary distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in its own Markov chain, which has a payoff that is significantly smaller than R. This should be intuitive: if tit-for-tat makes one mistake, it goes into a defection cycle that it doesn't break out of until it makes a second mistake.

4. Proofs

4.1. **The time average distribution.** Before we prove theorem 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

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

exists, for any strategies s_1 and s_2.

Proof. By definition,

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

By lemma 4.1, π is a stationary distribution. In particular, if M is the transition matrix for the s_1-s_2 Markov chain, then π is an eigenvector of M with eigenvalue 1. This implies that π is in the nullspace of $M - I$. Thus, we can find π 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 π 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 π_{c_1,c_2} and π_{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.**

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

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

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

$$\gamma < R.$$

We want to prove that s_1 is not evolutionarily stable. This would prove the theorem.

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 ϵ-invadable. We choose $\epsilon = \epsilon_0/2$, and present a strategy s_2 that can invade s_1 for sufficiently small p.

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 $v_{s_1}(s_2) = v_{s_2}(s_1) = v_{s_1}(s_1)$, 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 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 construct C_2 as follows. First, it finds out in finite time what action A that s_1 will take at some large finite future time T (with probability $1-O(p)$). After that, it responds with the opposite action (which we will denote by $\bar{A}$) right after time T. If s_2 subsequently perceives that its opponent takes action A, it will transition into a copy of C_1, the absorbing strongly connected component that C_2 replaces. On the other hand, if s_2 perceives its opponent taking action $\bar{A}$, then it will transition into an absorbing strongly connected component that always cooperates. A schematic view of C_2 is presented in figure ????. This concludes the construction of s_2.

THIS DOESN'T WORK. The problem is the circulation states: if s_2 goes into an absorbing SCC before s_1 does, our modified C_2 might affect the probabilities of s_1 going into certain absorbing SCCs. I think a way to fix that is as follows: You take C_1, and replace each of its individual states by an SCC, targeting a particular non-absorbing SCC of s_1. Make sure

We now want to prove that s_2 can in fact invade s_1 for small p. First, we will assume that C_2 can find out the action A that s_1 would take at time T, and finish the proof of our theorem using this assumption. After that, we will prove that that is indeed the case.

Claim 4.9. Given the above construction of s_2, the payoffs are as follows.

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

Proof. By definition, $v_{s_1}(s_1) = \gamma$. By lemma 4.5, the time average distribution of the s_1-s_1 Markov chain π is

$$\pi = \sum_{(C_1, C_1') \in \mathcal{C}_1 \times \mathcal{C}_1} p_{(C_1, C_1')} \cdot \pi^{(C_1, C_1')}$$

When creating s_2, we replaced each C_1 by a C_2, containing two absorbing strongly connected components: a copy of C_1, which we call $C_{1,\text{copy}}$ and the always cooperating C_c. This means that the condensation of the s_1-s_2 Markov chain is the same as the condensation of the s_1-s_1 Markov chain, except that each absorbing strongly connected component has been replaced by two. That is, the absorbing strongly connected component (C_1, C_1') of the s_1-s_1 chain has become both $(C_1, C_{1,\text{copy}})$ and (C_1, C_c) in the s_1-s_2 chain. Then, we find by our construction, that

$$p_{(C_1, C_{1,\text{copy}})} = (1 - O(p)) p_{(C_1, C_1')} \tag{1}$$
$$p_{(C_1, C_c)} = O(p) p_{(C_1, C_1')} \tag{2}$$

This equation is true because when s_2 plays against s_1, it identifies the action A that s_1 will take at time A with probability $(1 - O(p))$, and upon seeing that transitions to $C_{1,\text{copy}}$ with probability $(1 - p)$.

Now, we can calculate τ, the time average distribution of the s_1-s_2 Markov chain:

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

we're kinda abusing notation here

where the second-to-last equality follows by eq. (1). 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, note that at time T, s_2 and its copy s_2 will be in the exact same state with probability at least $(1 - p)^T = 1 - O(p)$, since if there are no mistakes they will both go to the exact same states. Thus, when s_2 outputs $\bar{A}$ to identify s_1, both of them will do that, and thus transition to an 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 4.9, 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 ϵ-invadable for this value of p. In conclusion, then, s_1 is not evolutionarily stable, which concludes the proof of theorem 3.1. □

We now return to the part we left out: how to construct C_2 such that it can find out which action that s_1 would take at time T.

Claim 4.10. 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 $1 - O(p)$.

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

(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. □

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-

Proof.

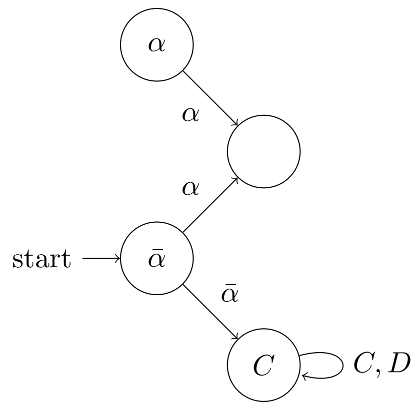

FIGURE 3. Constructon of invasion strategy, used in the proof of theorem 3.1

Claim 4.11. 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 α, and that the state s goes to upon perceiving the opponent move α is $c' = T(c, \alpha)$. Now, create two new states: c_0 and c_1. Define the transitions as

$$\begin{aligned} T(c_0, \alpha) &= c' \\ T(c_0, \bar{\alpha}) &= c_1 \\ T(c_1, \cdot) &= c_1 \end{aligned}$$

and the outputs as

$$\begin{aligned} G(c_0) &= \neg G(c_2) \\ G(c_1) &= C. \end{aligned}$$

Let the start state of s_2 be c_0.

Claim 4.12. Given the above construction of s_2, the following inequalities hold:

$$\begin{aligned} v_{s_1}(s_1) &\leq (1-p)^2\gamma + 2(1-p)pR + p^2 R \\ v_{s_1}(s_2) &\leq (1-p)\gamma + pT \\ v_{s_2}(s_1) &\geq (1-p)\gamma + pS \\ v_{s_2}(s_2) &\geq (1-p)^2 R + 2(1-p)p(\tfrac{S+T}{2}) + p^2\gamma. \end{aligned}$$

Before proving this claim, we will use it to finish our proof of theorem 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 ϵ 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 ϵ-invadable for this value of p. In conclusion, then s_1 is not evolutionarily stable, which concludes the proof of theorem 3.1.

□

Proof of claim 4.9. We can prove this using either of the two definitions. □

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 $\lim_{p \to 0} v_{s_2}(s_1) = T + P$ 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: Time average distributions

We might have periodicity, but for our purposes, we might as well extend the definition and look at periodic distributions as stationary too. The following two lemmas help with that.

Lemma 6.1. Given a starting distribution v and a Markov matrix M, for every $\epsilon > 0$, there will exist a k such that $|vP^{nk} - vP^{mk}| < \epsilon$ for all n and $m > 0$.

This proves that a Markov chain will always reach a periodic state.

Lemma 6.2. Suppose distributions form a chain $p_1 \to p_2 \to \dots \to p_n \to p_1$. Then $\pi = \frac{p_1 + \dots p_n}{n}$ is stationary.

This proves that we're able to talk about stationary distributions even when they don't really actually exist.

7. Appendix: Probablistic Automata

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.