CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

Evolution and Information in a Prisoner's Dilemma Game[1]

By Phillip Johnson, David K. Levine and Wolfgang Pesendorfer[2]

First Version: October 25, 1996

This Version: February 17, 1998

Abstract: In an environment of anonymous random matching, Kandori [1992] showed that with a sufficiently rich class of simple information systems the folk theorem holds. We specialize to the Prisoner's Dilemma and examine the stochastic stability of a process of learning and evolution in this setting. If the benefit of future cooperation is too small, then there is no cooperation. When the benefit of cooperation is large then only cooperation will survive in the very long run.

[1] We are grateful to financial support from National Science Foundation Grant SBR-93-20695 and the UCLA Academic Senate.

[2]Centro de Investigación Economía, ITAM, Department of Economics, UCLA, and Department of Economics, Princeton University.

1. Introduction

This paper is about the emergence of cooperative behavior as the unique long-run result of learning in a repeated Prisoners' Dilemma setting. There is a long-standing tension between the theory of repeated games, for which the folk theorem asserts that when players are patient all conceivable payoffs are possible in equilibrium, and common experience (supported by experimental research), which suggests that repeated Prisoners' Dilemma games typically result in cooperative behavior. Work by Young [1993] and Kandori, Mailath and Rob [1993] suggests that evolutionary forces can lead to unique outcomes in the long run, even in a setting where there are multiple equilibria. The goal of this paper is to apply that theory in the context of the repeated Prisoners' Dilemma game.

Evolution and learning are most easily studied in a setting of repeated interaction within a large population; this avoids complications due to off-path beliefs that occur in a repeated setting with a fixed set of players. There are two basic ways of incorporating a repeated Prisoners' Dilemma in such a setting: one is to study players who are matched to play infinitely repeated Prisoners' Dilemma games. This runs into difficulties with the finite horizon, as well as the size of the strategy space over which evolution or learning is taking place.[3] We instead adopt the framework of Kandori [1992]: here players are matched to play games with opponents for whom limited past play information is available. Economic examples of this sort abound. For example, in purchasing a home, renting an apartment or buying a car, an individual may carry out several transactions over his lifetime, but not with the same partner. Still, some information is available about the past performance of the current partner. For example, it may be possible to find out if someone has cheated in recent interactions. In the terminology of Kandori this

[3] Young and Foster [1991] study players matched to play infinitely repeated Prisoners' Dilemma games. However, they restrict the set of available strategies to "always cooperate", "always defect", and "tit-for-tat".

information about past play is distributed by "information systems." The central result of Kandori's paper is that, like in the purely repeated game setting, the folk theorem holds when players are sufficiently patient and have sufficient information.[4]

To prove a precise theorem about the emergence of cooperation, we make a number of specialized assumptions. We examine the model for a particular range of discount factors and payoffs. In particular, we assume that players' discount factors are such that although the following period is important, the effect of all later periods is small. It is possible to expand the parameter range for which our results are valid by restricting the strategy space. We discuss this issue in more detail in the conclusion.

Our model of learning is based on fictitious play. Because decisions have consequences that span more than one period, we must provide a model of belief formation that also spans multiple periods. We make the fictitious-play assumption of stationarity: players believe that opponents will not change their strategies, at least not in the relevant future. Players base their beliefs on private and public observations of past play. The assumption that all players have access to a common pool of observations of past play is made for tractability. By assuming that the common pool is larger than the private pool, to a good approximation, all players share the same beliefs, so the fictitious play dynamics resemble those of continuous time best-response, which is the model usually studied in the evolution literature.[5]

To the model of fictitious play, we add a stochastic error: players choose optimal strategies with probability less than one. This is similar to the stochastic fictitious play studied by Fudenberg and Kreps [1990] or Fudenberg and Levine [1995]. The stochastic element in the response serves the same role as "mutations" in evolutionary theory.

---

[4] This model applies even in populations too large and to players too impatient to admit the types of contagion effects studied by Ellison [1994].

[5] Many other authors have also pointed out the similarity of fictitious play to the continuous time best-response. See for example Fudenberg and Levine [1998].

In addition to optimization errors, we assume that the information systems reporting on players past play also make errors. These errors, which are assumed to be more prevalent than the optimization errors play an essential role in the analysis. While this assumption may be justified on grounds of realism, we make it to avoid the following problem: In a cooperative equilibrium, only cooperation is observed on the equilibrium path. This means that the strategy of always cooperate does as well as the equilibrium strategy, and we would ordinarily think that it is simpler and less costly to operate.[6] This leads players to switch to always cooperating, and, of course this is not an equilibrium at all. Our view is that this is not a problem of practical importance, because in real settings there are always errors and so punishment must be carried out occasionally.

In this basic setting, we study a limited class of "information systems" that are sufficiently rich to allow both cooperative and non-cooperative outcomes as equilibria (without learning or mutations). Applying the methods of Ellison [1995] we find sufficient conditions both for cooperation to emerge in the long run, and for defection to emerge in the long run. Several points deserve emphasis:

- The existence of cooperative equilibria is by itself not sufficient for cooperation to emerge in the long run. For some parameter values there are cooperative equilibria, but defection is nevertheless the long-run outcome. For other parameter values cooperation emerges in the long run.
- We allow for a variety of information systems. Players must choose which information system to consult and hence it is not a priori clear that players would individually choose to collect the appropriate information to support cooperation. We demonstrate that cooperative behavior is indeed associated with one particular information system and hence our results also imply that a unique information system emerges in the long-run to support cooperation.

[6] This is discussed, for example, in the automata models of Rubinstein [1986].

- When cooperation is the unique long run outcome it is supported by a strategy and an information system we call the *team strategy*. This strategy calls upon players to cooperate with members of the same team and punish members of the opposing team. Any player who does this is considered a team member; any player who does not is expelled from the team. The key property of the team strategy is that failure to punish a player is itself punished.
- Our conclusion that cooperation emerges in the long-run stochastically stable distribution does not mean that the first best is obtained. Because there is noise in the process, punishment takes place with positive probability. Consequently the long-run stochastically stable distribution while it involves cooperating "most of the time" is never the less Pareto dominated by the non-equilibrium outcome of always cooperating no matter what.

Our major result is that the team strategy emerges as the "winner" in the long-run when the benefit of cooperation is great, while the strategy of always defect emerges when the benefit of cooperation is low. The intuition is very close to the idea of risk dominance in static evolutionary games. If the benefit of defecting is small relative to the gain from cooperation, then a relatively small portion of the population mutating to team strategies makes it desirable for everyone else to follow, and the long-run outcome is cooperative. Conversely, if the benefit of defecting is too large then a relatively small portion of the population mutating to the strategy of always defecting makes is undesirable for anyone to cooperate.

To understand more clearly why the team strategy emerges in the long-run we can compare it to alternative strategies. First, consider tit-for-tat. This has traditionally been held up as an excellent strategy because it rewards good behavior, it punishes bad behavior, and it is "forgiving." However, the fact that information systems make errors mean that punishments occur a positive fraction of the time in our environment. Tit-for-tat is not robust in these environments because it punishes those who, according to the

strategy, must do the punishing. The team strategy also rewards good behavior, punishes bad behavior, and is "forgiving." But in this case, good behavior includes punishing non-members and bad behavior includes not punishing non-members. Therefore, the team strategy is robust in environments where punishment is actually called for.

As a second comparison, consider a *weak-team-strategy*. This strategy is similar to the team strategy in that members cooperate with other members and punish non-members. While failure to cooperate with members is punished, failure to punish non-members is not. This strategy is very similar to the team strategy since it is a best response (in a population where all players adopt this strategy) to punish non-members. Why is the team strategy is more successful that the weak team strategy? Consider a situation where some fraction of the population is playing tit-for-tat. In this case, punishment of non-members may be costly since it triggers punishment by players who use tit-for-tat. The weak-team strategy gives its members only a weak incentive to punish non-members whereas the team strategy gives its members a strong incentive to do so. Therefore, the team strategy is much more robust to an invasion of players using tit-for-tat than the weak team strategy.

An important ingredient of our analysis is a combination of restrictive assumptions to ensure that stage game strategies can be inferred from observations about actions and states. In particular, we assume that

- The costs of consulting information services are such that each player consults at most one service.
- Each information system sends two messages.
- There are two actions.
- Players believe that their opponents do not use strictly dominated strategies.

When the information system can send more than two messages, stage-game strategies cannot be inferred from observable information. In this case, our analysis fails to extend. We discuss this issue in the conclusion of the paper: both why the inability of players to

infer strategies makes such a large difference to the analysis, and why in practice it may not make so much difference.

2. The Model

In this section we describe a model of the evolution of strategies in a large population of players, randomly matched to play Prisoners' Dilemma games. The model is one of inter-temporally optimizing players who base their beliefs about the current and future play of opponents on information about past play.

Two different types of information about this past play are important in our analysis. First, players have access to specific information about their current opponent's history. This is essential if there is to be any possibility of cooperation in the absence of contagion effects. Second, since players are patient, their play depends on beliefs about the play of opponents they will meet in the future. These beliefs depend on information about the past play of other players, including the current one. It is useful for us to distinguish explicitly between information about the history of the current opponent, which we assume takes the form of "messages," and broader information about the past play of the population, which we refer to as "observations."

Specifically, when a player is matched with an opponent, he receives a "message" that provides information about the history of that opponent. This message is provided by an "information system." In addition, each player has access to a pool of "observations" about the results of various matches (including his own) that are used to draw inferences about the population from which the current and future opponents are drawn. A basic assumption we make is that players base their beliefs on the conjecture that opponents' strategies will not change over time.

2.1. The Stage-Game

There is a single population of $2n$ players who are randomly matched to play a Prisoners' Dilemma stage game. This stage game has two actions denoted C (cooperate) and D (defect). The payoff to player *i* when he plays a^i and his opponent *j* plays a^j are $u(a^i)+v(a^j)$ where $u(C)=0,\ u(D)=1,\ v(C)=x>1$, and $v(D)=0$. The corresponding normal form is

	C	D
C	x, x	$0, x+1$
D	$x+1, 0$	$1, 1$

Notice that the benefit of defecting is independent of the opponent's action, a useful simplification that we discuss later[7]. The parameter *x* measures the benefit from a cooperative opponent relative to the gain from defecting.

Each period, players are randomly matched in pairs. Players have a discount factor δ. We will focus primarily on the case in which δx is large and $\delta^2 x$ is small. In other words, we assume that the payoffs and discount factor are such that players care whether their opponents cooperate next period, but do not care about the more distant future.

2.2. Information Systems

When a player is matched with an opponent he receives a message from an information system about his current opponent's history. We assume that each information system can send only two messages, a "red flag" or a "green flag." Let $\{r, g\}$ be the set of messages. The message sent by an information system is Markov

[7] Note that our results do not depend on this assumption. It is made for convenience only.

meaning that the message sent to player *i*'s opponent in period *t* depends only on the actions taken and messages received by player *i* and his previous opponent in period *t*-1. Therefore, an information system is a map $\eta:\{C,D\}^2\times\{r,g\}^2\to\Delta(\{r,g\})$, with the interpretation that $\eta(a^i_{t-1},a^j_{t-1},\beta^i_{t-1},\beta^j_{t-1})[\beta^i_t]$ is the probability that the message provided to player *i*'s opponent at *t* is β^i_t.

We assume that there is a finite set *N* of available information systems. We let b^i_t denote the vector of messages sent by the different information systems in *N* about player *i* at time *t*. We also write $b^i_t(\eta)$ for the message corresponding to information system $\eta\in N$. We assume that information systems are noisy. Specifically, we fix a small positive number $\omega>0$ and assume that $\eta(\beta)\in\{\omega,1-\omega\}$. We take *N* to be the set of all such maps for a given ω. The probability of a flag vector for all information systems is

$$\prod_{\eta\in N}\eta(a^i_{t-1},a^j_{t-1},b^i_{t-1}(\eta),b^j_{t-1}(\eta))[b^i_t(\eta)]$$

Players may base their play on messages provided by information systems. However, we assume it is costly to acquire (or interpret) these messages. There is a small cost of picking one system[8] and a prohibitively large cost of picking two or more information systems. Therefore, each player picks at most one system from which to receive information about one player, and does so only if he intends to make use of the information. This information may either be one of the flags about himself or one of the flags about his opponent. In addition, we assume that players know that their opponents also face these costs and that they know that their opponents do not use dominated strategies. As we shall see below, this assumption plays a crucial role in the analysis.

A *stage game strategy* is a choice of a player to observe, an information system with which to observe the player, and the assignment of an action to the message received

[8] This assumption means that a player will not use an information system unless there is a strict gain in utility from the use of some information system. Conversely, if there is such a strict gain an information system will be used.

from that information system. Formally, we let $s^i = (k, \eta, a(\beta)), \beta \in \{r, g\}$ denote a stage game strategy where $k \in \{i, j\}$ is either the player himself or his opponent. We also allow that the player chooses no information system and represent that choice by $\eta = \varnothing$. In that case a must be independent of β.

We assume that player *i* does not automatically know the realization of his own flags. Only if the information system the player decides to consult reports on himself, can he learn the value of one of his own flags. However, we assume that a player learns all flags (b_t^i, b_t^j) at the end of period *t*. Since a player knows last period's realization of all his flags and the values of all the variables that determine transition probabilities of his information system he can form a forecast of his own flags at the beginning of the following period. This assumption captures the idea that while a player has a very good idea of his own flags he is never exactly sure what his current flags are.

2.3. Observations and the Observability of Stage Game Strategies

At the end of each match, we assume that the play of both players, the information system they consult, and all of their flags are potentially observable. Below we describe precisely who observes them. An *observation* is a vector $\phi_t^i = (a_t^i, \eta_t^i, b_t^i; a_t^j, \eta_t^j, b_t^j)$, where *j* is the opponent of *i* in period *t*. An observation does not include the names of the players who are matched.[9] The finite set of possible observations is denoted by Φ. These observations are used to form beliefs about the current and future play of opposing players.

We have not assumed that strategies are directly observable. But in effect we have. Recall that we assumed that players know their opponents' have a cost of using information systems and that their opponents do not use dominated strategies. Suppose a

[9] A message, on the other hand, does include this information. The assumption that observations are anonymous is a convenient simplification. If the population is large relative to the number of observations, it is unlikely that a player will meet an opponent for whom an observation is available.

player observes a non-null information system, a flag, and an action. To deduce the stage game strategy he needs to know what action would have been used if the flag had been the opposite color. He knows that if the flag had been the opposite color then the action would have been the opposite action since otherwise the strategy of using the null information system would have dominated. In this way, every observation yields a unique stage-game strategy. Note that the observability of strategies only follows because there are two flags and two actions. As we discuss below it is important to our results that players can infer strategies from observations about play.

2.4. Available Observations

We assume that individual players and society are limited in their ability to record and remember observations. Players have access to two pools of observations, both of fixed size. All players have access to a common public pool of observations, and to a private pool. The number of common observations is large relative to the number of private observations, so that all players have similar although not identical beliefs. Each player has access to K total observations: $(1-\xi)K$ in the common pool and ξK in the private pool. In other words, the pool of common observations is a fixed length vector $\theta_t \in \Phi^{(1-\xi)K}$, while player i's private observations are a fixed length vector $\theta_t^i \in \Phi^{\xi K}$. Private observations are updated each period. That is: θ_t^i and θ_{t-1}^i differ in exactly one component, which in θ_t^i is the observation of the most recent match. We assume the particular component replaced is drawn randomly.

In a similar way, common observations are augmented each period by randomly replacing some observations with current observations. There are $2n$ possible observations each period; an i.i.d. number of these observations $1 \le m_t \le 2n$ is used to randomly replace existing observations.

2.5. Formation of Beliefs

Our model is based on the fictitious-play like assumption that players believe that they will face the current empirical distribution of opponent strategies in all future periods. Unlike the usual evolutionary setting, beliefs about more than one future period are important because information systems cause current actions to have future consequences. In the standard case where players are myopic, fictitious play is known to have sensible properties as a learning procedure. If players do not frequently change from one strategy to another[10] players receive as much time average utility as if they had known the frequency (but not timing) of opponents' play in advance.[11] We would expect that similar properties hold in this environment.

Specifically, for a given set of observations θ_t, θ_t^i, there corresponds a unique empirical joint frequency distribution of stage game strategies and flags $\vartheta(\theta_t, \theta_t^i)$. At the beginning of each round *t* player *i* believes that last period the distribution of stage game strategies and flags was $\vartheta(\theta_{t-1}, \theta_{t-1}^i)$, and he knows that his own and opponents actions and flags were ϕ_{t-1}^i. To reach an optimal decision, he must form expectations about the joint distribution of stage game strategies and flags at times $t+\ell-1$, $\ell = 0,1,\ldots$. When forming these expectations player i assumes that no other player ever changes stage game strategies.[12] However, he recognizes that his future beliefs about the distribution of flags conditional on stage game strategies will depend upon future observations of opponents flags and actions. Let $\vec{\phi}_t^i(\ell)$ denote the observations acquired by player *i* between period *t* and $t+\ell-1$. The beliefs of player *i* in period *t* about ℓ periods in the future are denoted by $\vartheta_{t-1}^i(\ell, \vec{\phi}_t^i(\ell)), \ell = 0,1,\ldots$. Observe that the process for $\vartheta_{t-1}^i(\ell, \vec{\phi}_t^i(\ell))$, $\ell = 0,1,\ldots$ is determined entirely by the initial condition $\vartheta_{t-1}^i(0, \vec{\phi}_t^i(0)) = \vartheta(\theta_{t-1}, \theta_{t-1}^i)$, the assumption

[10] They do not in the dynamics considered here.
[11] See Fudenberg and Levine [1995] or Monderer, Samet and Sela [1994].
[12] It is important for our results only that opponents are assumed to not vary their stage-game strategies for one period; beliefs about the more distant future do not matter under our assumption about the discount factor. For concreteness, we make this assumption about all future periods as well.

of random matching, and the information systems determining the transition probabilities for flags. We should emphasize the importance of the player's belief that all other players repeat the stage game strategy used in period $t-1$ in every subsequent period.

2.6. Behavior of Individual Players

Player *i*'s *intentional behavior* is given by the solution to the optimization problem of choosing a function $\rho^i_{t+\ell}$ of $\vartheta^i_{t-1}(\ell, \vec{\phi}^i_t(\ell))$ and $\phi^i_{t+\ell-1}$ to maximize

$$E\sum_{\ell=0}^{\infty}\delta^{\ell}\left(u(\rho^i_{t+\ell}(b^j_{t+\ell}, b^i_{t+\ell})) + v(s^j_{t+\ell}(b^j_{t+\ell}, b^i_{t+\ell}))\right)$$

where the evolution of $b^i_{t+\ell}$ is determined by the information systems. We let $\rho^i_t(\theta_{t-1}, \theta^i_{t-1}, \phi^i_{t-1})$ be the intentional behavior at *t*. In case of a tie, a tie-breaking rule that depends only on $\theta_{t-1}, \theta^i_{t-1}, \phi^i_{t-1}$ is used.[13]

We also assume the possibility that players make errors. Specifically we suppose that the probability of the intentional behavior $\rho^i_t(\theta_{t-1}, \theta^i_{t-1}, \phi^i_{t-1})$ is $1-\varepsilon$, and that every other stage-game strategy is chosen with probability $\varepsilon[(\#S)-1]^{-1}$.[14]

2.7. Evolution of the System

The evolution of the entire system is a Markov process *M*, where the state space Θ consists of the set of common observations and the collection of private observation, flag pairs. The Markov process is determined by the assumption that players are equally likely to be matched with any opponent, the rules for updating observations, the information systems governing the dynamics of flags, and by the behavior of individual players described above.

---

[13] The particular tie-breaking rule is irrelevant to our analysis. Notice that we are not allowing players to play mixed strategies: because we are dealing with a large population $2n$, mixed strategies can be purified as in Harsanyi [1973].

[14] Note that the assumption that alternative strategies are chosen with equal probability is not essential. It is essential that the ratio between the probabilities of alternative strategies not go to 0 as ε goes to 0. For a discussion of the problems that occur when this assumption fails, see Bergin and Lipman [1995].

Not all combinations of observations are possible. For example, when two players are matched they must add the same observation to their private pool; since at least one observation is added to the common pool it must also be added to at least two of the private pools. We denote the set of feasible observations by Θ^f and note that *M* is also a Markov process on Θ^f.

To analyze the long-run dynamics of *M* on Θ^f, note that it takes no more than $(1-\xi)K$ periods to replace all the observations in the common pool and no more than ξK periods to replace all observations in all $2n$ private pools. Since we assume that $(1-\xi)K > \xi K$, the positive probability of behavioral and flag errors imply that M^T is strictly positive for $T \geq (1-\xi)K$. It follows that the process *M* is ergodic with a unique stationary distribution μ^ε. Moreover, because of the behavioral errors, the transition probabilities are polynomials in ε. Consequently we can apply Theorem 4 from Young [1993]and conclude that $\lim_{\varepsilon \to 0} \mu^\varepsilon$ exists. We denote this limit as μ and refer to it as the *stochastically stable distribution*. From Young's theorem, this distribution places weight only on states that have positive weight in stationary distributions of the transition matrix for $\varepsilon = 0$. Our goal is to characterize the stochastically stable distribution for several special cases using methods developed by Ellison [1995].

3. The Main Theorems

Our main results characterize the stochastically stable distribution for particular parameter ranges. First we give conditions under which players always defect in the stochastically stable distribution. Let $\Theta^D \subset \Theta^f$ denote the set of states where all players have samples consisting of all players playing the stage game strategies of always defect. Proposition 1 says that if the gains to cooperation are small, if the size of the private samples is small compared to that of the public sample, and if players update their beliefs slowly, then the stochastically stable distribution places all its weight on states in Θ^D. Recall that ω is the probability of a erroneous message, that ξ is the fraction of total

observations which are private, $2n$ is the number of players, x is the utility from cooperating, and K is the total number of observations available to an individual player.

Proposition 1: If $\delta x/(1-\delta) < 2(1-\xi)/(1-2\omega)$ then $\mu(\Theta^D) = 1$.

Remark: If ξ, ω are small, the condition is to a good approximation $\delta x/(1-\delta) < 2$. If $\delta x/(1-\delta) < 1$, then the present value benefit of changing the play of all future opponents from defecting to cooperating is less than the gain from defecting. In this case, the stochastically stable distribution obviously has all players defecting. The statement of the theorem shows that even if $\delta x/(1-\delta)$ is greater than the gain from defecting there will be stochastically stable distributions in which all players defect.

Proof: Each observation has information about two strategies. It is useful to consider the induced *empirical frequency of strategies*. Let $\Theta^{D/2}$ denote the set of states where the public sample has an empirical frequency of strategies of at least ½ always defect. Notice that individual samples must have an empirical frequency of strategies of at least $(1-\xi)/2$ always defect. From Ellison [1995] it is sufficient to show that at $\Theta^{D/2}$ the intentional play of all players is to defect. At $\Theta^{D/2}$ defect earns an immediate payoff of 1. Cooperation at best gains an expected present value of

$$\frac{(1-2\omega)\delta x}{2(1-\xi)(1-\delta)}$$

over all subsequent periods. If this is less than one, the then the intentional behavior must be to defect (for all players).

☑

Our second result describes a range of parameters for which cooperation will occur in the stochastically stable distribution. Moreover, a particular pair of strategies emerges as the stochastically stable outcomes. We call these strategies *red-team* and the *green-team* strategies.

In the team strategies the color of the flag can be interpreted as a "team" that the player is on. The strategy demands that players cooperate with team members, and defect against non-members. Failure to do so results in expulsion from the team. Conversely, anyone who behaves in accordance with the strategy is admitted to the team. So, for example, a player who follows the green team strategy cooperates with players who have a green flag and defects when facing a player with a red flag. As long as the player follows the green team strategy he receives a green flag with probability $1-\omega$. If he either does not cooperate with a team member (someone who has a green flag) or if he does not defect when facing a player with a red flag then the player receives a red flag with probability $1-\omega$. The red-team strategy is the analogous strategy when the role of the flags is reversed.

Let Θ^R denote the set of states where all players' samples consist of all players playing the red team strategy, and Θ^G where the samples consist of all players playing the green team strategy. Our main result about cooperation is:

Proposition 2: If

$$\frac{1}{2} > \frac{1}{\omega^5(1-2\omega)}\left(\frac{4}{\delta x(1-\delta)} + \frac{4n}{K}(1-2\omega) + 2\xi + \frac{\delta}{1-\delta}\right)$$

then $\mu(\Theta^R) = \mu(\Theta^G) = 1/2$.

We will prove this below. To understand more clearly the implications of these two propositions fix $\omega > 0$. Suppose that δx is large, that n/K is small, that ξ is small and that δ is small. Then the hypothesis of Proposition 2 is satisfied. Having established values of these variables that satisfy the hypothesis, allow *x*, the gain to cooperation, to vary. Then for $x = 1$ there is no benefit from cooperation, and by Proposition 1, the stochastically stable distribution has all players defecting and receiving per period utility of one. As *x* increases up to $\underline{x} = 2(1-\xi)(1-\delta)/\delta(1-2\omega)$ there is benefit to cooperation, but the stochastically stable distribution remains with all players defecting and receiving

per period utility of one. For some intermediate range of x, neither proposition applies. However, as x continues to increase into the range where $\bar{u}/\delta \le x \le \underline{v}/\delta^2$, the stochastically stable distribution is all players observed to be playing the same team strategy. When one of the team strategies is played every player receives x with probability $1-\omega$ and 0 with probability ω. The per-period expected utility of a player is therefore $(1-\omega)x$. For values of x larger than $\underline{v}/\delta^2$, we again do not know the stochastically stable distribution. Figure 1 plots per period utility in the stochastically stable distribution as a function of the benefit of cooperating, x.

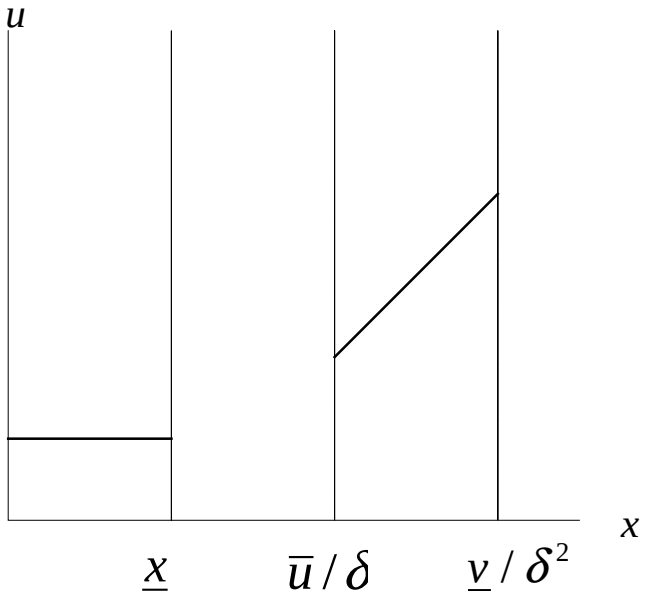

Figure 1 - Utility as a Function of Benefit of Cooperation

What happens when the environment becomes more "cooperative" in the sense that the benefit to cooperation x increases? For small values of x nothing happens. But when x becomes sufficiently large, players begin to cooperate. This increase in x improves welfare for two reasons: first, x itself is larger; second, players actually succeed in cooperating and therefore realize the benefits of the increased x.

Proposition 2 requires that the gains from cooperation are sufficiently large and that the discount factor δ is sufficiently small as compared to the flag noise ω. At first sight it may seem puzzling that impatience should be a condition required to sustain cooperation. This condition arises because we need to ensure that players' payoffs two and more periods in the future do not affect current behavior. Thus, our condition on δ is an expression of the assumption that only the current and the following period matter for the agents. Alternatively, we could restrict the map that determines the transition probabilities of flags and eliminate the restriction on the discount factor. Specifically, consider the case where the flag distribution only depends on the player's action and his opponent's flag. Then the player's action has consequences for his flag in following period only. Therefore, when a player chooses an optimal strategy it is sufficient to consider a two-period horizon even if the assumption on δ in Proposition 2 does not hold.

4. About Stage Game Strategies

Before proving Proposition 2, we begin with a general discussion of different types of stage game strategies. We define, from the perspective of the opponent of a player *j* in period t (denoted by player *i*), a map that represents the choice of a stage game strategy of player *i*. More precisely, assume that player *i* meets player *j* in period t and player *j* met player *k* in period $t-1$. For a particular s_t^i, b_t^i we define the *reaction* of player *i*, denoted by $\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i, b_t^i)[a_t^i] \in \Delta(\{C, D\})$, to be the probability that *i* takes a particular action as a function of the behavior and flags of *j* and *j*'s opponent in the previous period.

We call a reaction, $\bar{s}^i(\cdot; s_t^i, b_t^i)$, *responsive* if there exists some previous action of player *k*, a_{t-1}^k, and a state of information about players *j* and *k*, (b_{t-1}^j, b_{t-1}^k) such that $\bar{s}^i$ responds to the previous action of *j*, a_{t-1}^j. More precisely, $\bar{s}^i$ is responsive if there exists some $(a_{t-1}^k, b_{t-1}^j, b_{t-1}^k)$ such that if $a_{t-1}^j \neq \hat{a}_{t-1}^j$ then

$$\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i, b_t^i)[a_t^i] \neq \bar{s}^i(\hat{a}_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i, b_t^i)[a_t^i].$$

Note that for a responsive reaction the own flag of player *i* in the vector (s_t^i, b_t^i) does not affect the value of $\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i, b_t^i)[a_t^i]$ since a necessary condition for responsiveness is that the player looks at a flag of his opponent (and hence by assumption does not look at his own flag). Therefore we may write $\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i)[a_t^i]$ instead of $\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i, b_t^i)[a_t^i]$ whenever the reaction is responsive. Consequently the set of stage game strategies that give rise to responsive reactions is well defined. We call those the responsive stage game strategies and denote them by S^R. A strategy that is not responsive is called *unresponsive*. We say that the stage game strategy is *always responsive* if

$$\bar{s}^i(a_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i)[a_t^i] \neq \bar{s}^i(\hat{a}_{t-1}^j, a_{t-1}^k, b_{t-1}^j, b_{t-1}^k; s_t^i)[a_t^i]$$

regardless of the value of $a_{t-1}^k, b_{t-1}^j, b_{t-1}^k$. The two team strategies are always responsive, while the always defect and always cooperate stage game strategies are unresponsive.

There are several other specific strategies that are of interest. Other always responsive stage game strategies are *tit-for-tat* and *tat-for-tit*. These two strategies use an information system that assigns flags based only on how the player himself plays. For example, consider the information system that gives a $1-\omega$ chance of a green flag for cooperating and a $1-\omega$ chance of a red flag for cheating. Tit-for-tat then cooperates on green and defects on red; tat-for-tit defects on green and cooperates on red. Notice that there are two tit-for-tat and two tat-for-tit strategies, as it is also possible to use the information system in which the role of the red and the green flags are reversed.

A useful example of a responsive strategy that is not always responsive is the *weak green team strategy*. The weak green team strategy is similar to the green team strategy, except that its information system gives a $1-\omega$ chance of a green flag whenever the opponent has a red flag. Thus the strategy is not responsive when a player's opponent has a red flag. If the opponent has a green flag then the player receives a green flag with probability $1-\omega$ only if he cooperates. Consider a population of consisting largely of agents who play this strategy. If a player playing weak green meets an opponent with a

red flag, it is still optimal to defect since defecting yields a gain of 1 and there is no impact on the flag awarded by this information system. Thus the behavior of the weak green team and the green team strategies seem identical. One obvious question we will have to deal with is why, when the green team strategy has probability 1/2 in the stochastically stable distribution, the weak green team strategy has no weight at all.

We turn now to the issue of how a player should choose his intentional behavior. He has no control over his opponent's current play, and utility is additively separable, so his first concern is with the current benefit from defecting versus that of cooperating. Our restrictions on the discount factor are designed so that the only other important concern players have is with the consequences of their current play on how they will be treated by next period's opponent.

At the end of a match, a player's prediction of his own flags is represented by $p(\phi^i_{t-1})[b^i_t]$, the (scalar) probability of a flag vector generated by the vector of information systems given that the observation of the previous match is ϕ^i_{t-1}. Also let $\{s = a, \hat{s} = \hat{a}\}$ be the set of flags b^i, b^j such that when s is played the action a is chosen $\hat{s}$ the action $\hat{a}$. Let ϑ^i be an arbitrarily given belief. We can define an *approximate gain function* G^i from using s instead of $\hat{s}$:

$$
\begin{aligned}
&G(s,\hat{s},\vartheta^i,\phi^i_{t-1}) \equiv \vartheta^i_{t-1}[\{s = D,\hat{s} = C\}] - \vartheta^i_{t-1}[\{s = C,\hat{s} = D\}] \\
&+ \sum_{b^i_t, s^j, b^j_t} \sum_{s^k \in S^R} p(\phi^i_{t-1})[b^i_t]\vartheta^i_{t-1}[s^j, b^j_t]\vartheta^i_{t-1}[s^k]\bar{s}^k(s(b^i_t, b^j_t), s^j(b^j_t, b^i_t), b^i_t, b^j_t; s^k)[C]\delta x \\
&- \sum_{b^i_t, s^j, b^j_t} \sum_{s^k \in S^R} p(\phi^i_{t-1})[b^i_t]\vartheta^i_{t-1}[s^j, b^j_t]\vartheta^i_{t-1}[s^k]\bar{s}^k(\hat{s}(b^i_t, b^j_t), s^j(b^j_t, b^i_t), b^i_t, b^j_t; s^k)[C]\delta x
\end{aligned}
$$

Notice that beliefs about next period flags are irrelevant, while beliefs about next period strategies are by assumption the same as this period; this is why $\vartheta^i_{t-1}[s^k]$ in this expression is used to represent beliefs about one period in the future. This approximate gain function captures the idea that only this period's action and the action of the next period's opponent have a significant impact on the player's payoff. It consists of four

terms. The first term is the utility from defection received when s defects and $\hat{s}$ does not. From this is subtracted the utility from defection when $\hat{s}$ defects and s does not. The third term is next period utility received from a cooperative opponent under s but not $\hat{s}$. From this is subtracted the final term, which is the next period utility received from an opponent who is cooperative under $\hat{s}$ but not s.

Our first lemma makes precise the extent to which this is the case. In the following we let $\vartheta(\theta_t)$ denote the beliefs of a player who only observes the public pool of observations.

Lemma 1: If

$$G(s,\hat{s},\vartheta(\theta_{t-1}),\phi^i_{t-1}) > \delta + \delta^2(1+x)/(1-\delta) + (1+\delta x)\xi$$

then for all θ^i_t $\rho^i_t(\theta_{t-1},\theta^i_{t-1},\phi^i_{t-1}) \neq \hat{s}$, that is, the intentional behavior is not $\hat{s}$.

Proof: Let $V^i_t(s)$ be the expected present value according to i's beliefs based on information from time $t-1$ of the plan of playing s in the first period and optimally forever after. We show that $G(s,\hat{s},\vartheta(\theta_{t-1}),\phi^i_{t-1})$ is close to $V^i_t(s)-V^i_t(\hat{s})$; if $V^i_t(s)-V^i_t(\hat{s})$ is positive then $\rho^i_t(\theta_{t-1},\theta^i_{t-1},\phi^i_{t-1}) \neq \hat{s}$, so the result will follow.

First consider $G(s,\hat{s},\vartheta(\theta_{t-1},\theta^i_{t-1}),\phi^i_{t-1})$ in comparison to $V^i_t(s)-V^i_t(\hat{s})$. Notice that the utility $v(a^j_t)$ due to the opponent's action in the first period is independent of whether s or $\hat{s}$ is used, and so drops out of $V^i_t(s)-V^i_t(\hat{s})$. In comparison to $V^i_t(s)-V^i_t(\hat{s})$, $G(s,\hat{s},\vartheta(\theta_{t-1},\theta^i_{t-1}),\phi^i_{t-1})$ omits two other terms: the utility $\delta u(a^i_{t+1})$ due to the players own play in the second period, and the present value of all utility received in period 3 and later. The first term is at most δ in absolute value; the second at most $\delta^2(1+x)/(1-\delta)$. We conclude that

$$\left|G(s,\hat{s},\vartheta(\theta_{t-1},\theta^i_{t-1}),\phi^i_{t-1}) - \left[V^i_t(s)-V^i_t(\hat{s})\right]\right| \leq \delta + \delta^2(1+x)/(1-\delta).$$

Finally, we compare $G(s,\hat{s},\vartheta(\theta_{t-1},\theta^i_{t-1}),\phi^i_{t-1})$ with $G(s,\hat{s},\vartheta(\theta_{t-1}),\phi^i_{t-1})$. Observe that $\vartheta(\theta_{t-1},\theta^i_{t-1}) = (1-\xi)\vartheta(\theta_{t-1}) + \xi\vartheta(\theta^i_{t-1})$, so that

$$\left|\vartheta(\theta_{t-1},\theta^i_{t-1})-\vartheta(\theta_{t-1})\right| = \xi(\vartheta(\theta^i_{t-1})-\vartheta(\theta_{t-1})) \le \xi .$$

Since G is linear in ϑ with coefficient bounded by $1+\delta x$, the result follows.

☑

We call a stage game strategy is *strong* if it is always responsive, and if the information system it uses depends on only one of the two player flags and the action of the player in question. We denote the set of strong strategies by S^S. The two team strategies, tit-for-tat, and tat-for-tit are all examples of strong strategies. The key feature of a strong strategy is that if you expect to meet a player next period playing a particular strong strategy s^j, by a unique choice of current strategy, you can obtain a probability $1-\omega$ of getting x from that player next period. This is achieved by observing the flag (yours or your opponents) that will be used by the information system, and, since s^j is always responsive, playing the unique action that leads to a probability $1-\omega$ that s^j will cooperate next period. We refer to this strategy as $B(s^j)$, and somewhat loosely, refer to it as a *best response* to s^j.

Our second lemma characterizes the approximate gain from using the best response to a strong strategy as a function of the fraction of the population thought to be using that strategy. The basic idea is that if you play a best response to a strong strategy, then next period you get $(1-\omega)x$ from opponents using that strategy.

Lemma 2: Suppose that s is a strong strategy. If $\hat{s} \neq B(s)$ then

$$G(B(s),\hat{s},\vartheta(\theta_{t-1}),\phi^i_{t-1}) \ge \omega\left(-1+(1-2\omega)\left(2\vartheta(\theta_{t-1})[s]-1\right)\delta x\right),$$

provided the RHS is non-negative.

Proof: First observe that since each flag occurs with probability at least $\omega > 0$, $\hat{s}$ takes a different action than $B(s)$ with probability at least $\omega > 0$. Consider the event that the action taken in period t by $\hat{s}$ is different from the action taken by $B(s)$. In period $t+1$, if i meets an agent who uses s then this agent cooperates with probability ω if $\hat{s}$ was chosen

in period t and with probability $1-\omega$ if $B(s)$ was chosen in period t (s is strong and the two strategies call for different actions). If i meets an agent who does not use s then we may assume that this agent's choice of action depends on i's flag (otherwise there is no difference in i's period $t+1$ payoff stemming from his choice in t between the two stage game strategies.) Thus i's opponent cooperates with probability at most $1-\omega$ if $\hat{s}$ was chosen in period t and with probability at least ω if $B(s)$ was chosen in period t. Summing up these components, we get as a lower bound for the period $t+1$ component of G (in the event that the actions are different):

$$\begin{aligned}&\left[(1-\omega)\vartheta(\theta_{t-1})[s]-\omega\vartheta(\theta_{t-1})[s]+\omega(1-\vartheta(\theta_{t-1})[s])-(1-\omega)(1-\vartheta(\theta_{t-1})[s])\right]\delta x\\&=\left[(1-2\omega)(2\vartheta(\theta_{t-1})[s]-1)\right]\delta x\end{aligned}$$

The lower bound for G in period t is -1. Since the probability $\hat{s}$ and $B(s)$ play differently is at least ω when $-1+(1-2\omega)\left(2\vartheta(\theta_{t-1})[s]-1\right)\delta x>0$ the bound in the lemma follows from Lemma 1.

☑

A key property of the team strategies is that they are the only strong strategies that are best responses to themselves.

Lemma 3: If s is a strong strategy and $s=B(s)$ then s is either the red or green team strategy.

Proof: Suppose without loss of generality that s responds to green by cooperating. Suppose i is playing s and meets an opponent j also playing s with a green flag. Then since s is a best response to itself and cooperates, it must be in the expectation of receiving a green flag (so an opponent playing s will cooperate next period), and in the expectation that defecting will result in a red flag. Similarly, if i meets a red flag, he must expect to get a green flag for defecting and a red flag for cooperating. This uniquely defines the information system used by the green team strategy, and this is the only strong strategy that uses that information system.

☑

From the first three lemmas, if s is one of the team strategies, and

$$\vartheta(\theta_{t-1})[s] > \frac{1}{2} + \frac{\delta + \delta^2(1+x)/(1-\delta) + (1+\delta x)\xi + \omega}{2(1-2\omega)\omega\delta x} \equiv 1 - \frac{R[s]}{(1-\xi)K}$$

then the intentional behavior of all players is to play s. This equation defines $R[s]$ which, as we discuss below, is Ellison [1995]'s radius of the team strategy *s*. Notice that if δx is large and ξ and $\delta^2 x$ are small, then the right hand side of this expression is only slightly larger than 1/2. This says that the team strategies are "almost" 1/2-dominant.[15]

When the public sample satisfies the inequality $\vartheta(\theta_{t-1})[s] > 1 - R[s]/(1-\xi)K$ and the intentional behavior of all players is to play *s*, then in the absence of mutations ($\varepsilon = 0$) the fraction of public observations in which the particular team strategy is being used cannot decrease, and with positive probability must increase.[16] Consequently the same inequality is satisfied in the next period and the process converges to all players playing the team strategy, and all observations agreeing with this. The results of Ellison [1995] enable us to draw conclusions about the dynamics with mutation from the dynamics without mutation. If s is the green team strategy, then the number $R[s]$ is referred to by Ellison [1995] as the *radius* of the state Θ^G. This means that if the state is in Θ^G then a shock of fewer than $R[s]$ mutations followed by a sufficiently long period with no mutations will return the system to Θ^G.

If $R[s] > 1/2$, so that the team strategies were actually 1/2 dominant, it would follow that it would require fewer mutations to get to Θ^G than to depart, and as the mutation rate went to zero, this would mean that vastly more time would be spent at Θ^G than anywhere else, which is the conclusion of Proposition 2. Unfortunately, the bound

[15] Recall from the proof of Proposition 1 that 1/2 dominance means that if in the public sample slightly more than half of the observations are of one of the team strategies, then the intentional behavior of all players is to do the same.

[16] The only reason the number of team strategy observations will fail to increase is if the new observations of both players playing the team strategy displace only existing observations of the team strategy. Unless all observations are already of this type, there is a positive probability this does not happen.

in Lemma 2 is fairly tight: if the slightly less than half of the population were playing tat-for-tit, then the intentional behavior is to always defect. However, tat-for-tit is not terribly interesting as a stage-game strategy, as it is not a best response to anything. Ellison [1995] shows that Θ^G is the stochastically stable distribution if the radius is larger than the co-radius, where the co-radius is the number of mutations needed to get back to Θ^G from initial conditions that have positive asymptotic probability in the intentional dynamic, which tat-for-tit clearly does not. Our strategy for proving Proposition 2 is to show that the co-radius is smaller than the radius for all initial conditions that have positive asymptotic probability in the intentional dynamic.

Our immediate goal is to calculate the co-radius for initial conditions in which some of the players not playing the team strategy are playing a strategy which is not strong. An example of this situation is the case where half of the population is playing the green team strategy and the other half is playing the weak green team strategy. The weak green team strategy, recall, is not strong because it is not always responsive: if a player's opponent has a red flag, he gets a green flag regardless. Why should the intentional behavior in this situation be to choose the green team strategy rather than the weak green team strategy? The two strategies are very similar, however if the green team strategy is used, consider the occasion when an opponent with a weak red flag and strong green flag meet; in this case cooperation will occur against a weak red flag. The following period, whether the new opponent is using the green or weak green strategy, there is a $1-\omega$ chance of getting x. On the other hand if the situation is reversed, so that the weak green team strategy is used, when an opponent with a strong red flag and weak green flag meet. Then the following period there is only a $(1-\omega)/2$ chance of getting x as the player will likely (with probability $1-\omega$) get a strong red flag for failing to defect. The next lemma shows that this is more generally a problem with strategies that are not strong: unlike strong strategies, they cannot guarantee a $1-\omega$ chance of x if and only if the correct current choice is made.

Lemma 4: Suppose that s is one of the team strategies. Then if $\hat{s} \neq s$ and $\sim S^S$ is the set of strategies that are not strong then

$$G(s,\hat{s},\vartheta(\theta_t),\phi_t^i) \geq \omega\left(-1+\left((1-2\omega)\left(2\vartheta(\theta_t)[s]-1\right)+\vartheta(\theta_t)[\sim S^S]\left((1-2\omega)\omega^3\right)\right)\delta x\right)$$

provided the RHS is non-negative.

Proof: First we consider the case of strategies that are not strong, but are always responsive. Information systems used by such strategies have the property that for some period t flag combination agent i's $t+1$ flag depends on both his and his opponent's period t flags in a non-trivial way. For this information system, each pair of flags, one for i and one for his opponent, occurs with probability at least ω^2. Since the player can only observe one of the two flags there must exist a flag combination for which the agent is unsure about the "right" action, that is, the action that guarantees a $1-\omega$ probability of cooperation from the always responsive but not strong strategy in the following period. Since the agent sees only one of the two flags and since each flag occurs with probability at least ω there is at least a ω chance that the agent takes the "wrong" action for this particular flag combination. Thus there is at least a ω^3 chance that next periods probability of cooperation is ω. Therefore, when agent i is in a population containing opponents who use an always responsive but not strong stage game strategy the probability of cooperation by such an agent in the following period is bounded below by

$$(1-\omega^3)(1-\omega)+\omega^3\omega = (1-\omega)-\omega^3(1-2\omega)$$

Thus the maximum difference in the probability of cooperation in period $t+1$ between $\hat{s}$ and s when facing such an agent is given by

$$(1-\omega)-\omega^3(1-2\omega)-\omega = (1-2\omega)-\omega^3(1-2\omega).$$

For a next period opponent's strategy that is not always responsive, then there is a flag pair for the corresponding information system such that the probability of cooperation in the next period is independent of the action of player i. The greatest

possible difference between the probability of cooperation in the next period as a consequence of a current action is $(1-\omega)-\omega = 1-2\omega$. To find a bound on the maximum difference in the probability of cooperation in period $t+1$ between $\hat{s}$ and s we multiply this by the probability that such a flag pair does not occur, that is $1-\omega^2$ and find

$$(1-2\omega)(1-\omega^2) = (1-2\omega)-\omega^2(1-2\omega)$$
$$\leq (1-2\omega)-\omega^3(1-2\omega)$$

We now use these bounds to calculate *G*. Notice that the non-responsive strategies do not appear in *G* since they treat all players the same way no matter how they play. From the definition *G* can be divided therefore into two components: one from meeting *s* at $t+1$, and a second component from meeting all other responsive strategies

$$G(s,\hat{s},\vartheta(\theta_t),\phi_t^i) \geq \vartheta(\theta_t)[\{\hat{s}\neq s\}]\Big(-1+\big(1-2\omega\big)\vartheta(\theta_t)[s_{t+1}^j = s]\delta x +$$
$$(1-\vartheta(\theta_t)[s_{t+1}^j = s])\Big(\Pr(\{s^j = C\}|s,b_t^i,\theta_t)-\Pr(\{s^j = C\}|\hat{s},b_t^i,\theta_t)\Big)\delta x\Big)\geq$$
$$\vartheta(\theta_t)[\{\hat{s}\neq s\}]\big(-1+\big(1-2\omega\big)\vartheta(\theta_t)[s_{t+1}^j = s]\delta x$$
$$+(1-\vartheta(\theta_t)[s_{t+1}^j = s])\min_{b_t^i}\Big(\Pr(\{s^j = C\}|s,b_t^i,\theta_t)-\Pr(\{s^j = C\}|\hat{s},b_t^i,\theta_t)\Big)\delta x\Big)$$

The previous argument demonstrated that

$$\min_{b_t^i}\Big(\Pr(\{s^j = C\}|s,b_t^i,\theta_t)-\Pr(\{s^j = C\}|\hat{s},b_t^i,\theta_t)\Big)$$
$$\geq\Big(-(1-2\omega)+\vartheta(\theta_{t-1})[\sim S^S]\omega^3(1-2\omega)\Big)$$

Substituting in the bound, since $\vartheta(\theta_t)[\hat{s}(b_t^j)\neq s(b_t^j)]\geq\omega$ the lemma follows.

☑

Next we calculate the co-radius for initial conditions in which players not playing the team strategy are playing several different strong strategies. Our basic intuition is that these strategies tend to interfere with one another, and consequently, a team strategy can take over despite constituting a portion of the population less than the radius.

Lemma 5: Suppose that s is one of the team strategies, and that $\hat{s},\breve{s}$ are strong strategies with $s\neq\hat{s}\neq\breve{s}$. Then if $\hat{s}\neq s$

$$G(s,\hat{s},\vartheta(\theta_t),\phi_t^i) \geq$$
$$\omega\Big(-1+\big((1-2\omega)\big(2\vartheta(\theta_t)[s]-1\big)+\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}(\omega^2-2\omega^3)\big)\delta x\Big)$$

provided that the RHS is non-negative.

Proof: First observe that as in the proof of Lemma 4 we can divide G into components corresponding to meeting s and not meeting s:

$$G(s,\hat{s},\vartheta(\theta_t),\phi_t^i) \geq$$
$$\vartheta(\theta_t)[\{\hat{s}\neq s\}][-1+(1-2\omega)\vartheta(\theta_t)[s_{t+1}^j=s]\delta x+$$
$$(1-\vartheta(\theta_t)[s_{t+1}^j=s])\min_{b_t^i}\big(\Pr(\{s^j=C\}|s,b_t^i,\theta_t)-\Pr(\{s^j=C\}|\hat{s},b_t^i,\theta_t)\big)]\delta x$$

Assume that $\hat{s},\breve{s}$ cooperate when the opponent has a green flag and defect if the opponent has a red flag. This is without loss of generality since $\hat{s},\breve{s}$ use different information systems, and different flags anyway. Suppose player i uses strategy $\hat{s}$. Since $\hat{s},\breve{s}$ are strong it follows that for every pair of flags corresponding to the information systems of $\hat{s},\breve{s}$ there is a unique pair of actions for player i that ensures a $1-\omega$ probability of a green flag in period $t+1$. Also observe that every possible pair of flags for the two information systems corresponding to $\hat{s},\breve{s}$ occurs with probability at least ω^2 in period t. Thus with probability at least ω^2 a player who uses $\hat{s}$ will be defected against by one of the two strategies $\hat{s},\breve{s}$ with probability $1-\omega$. Or in other words he will be cooperated with at most ω of the time. All other strategies cooperate with probability no more than $1-\omega$. Thus a player who uses $\hat{s}$ can expect cooperation from his opponent (in period $t+1$) with a probability no greater than

$$\begin{aligned}&\omega[\omega^2\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}]\\&+(1-\omega)[1-\omega^2\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}]\\&=(1-\omega)-(1-2\omega)\omega^2\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}\end{aligned}$$

A player who uses the team strategy will be defected against with probability no larger than $1-\omega$ by all players who do not use the team strategy. Thus we get that

$$
\begin{aligned}
&(1-\vartheta(\theta_t)[s^j_{t+1}=s])\min_{b^i_t}\left(\Pr(\{s^j=C\}|s,b^i_t,\theta_t)-\Pr(\{s^j=C\}|\hat{s},b^i_t,\theta_t)\right)]x\\
&\geq(1-\vartheta(\theta_t)[s^j_{t+1}=s])[\omega-\\
&(1-\omega)+(1-2\omega)\omega^2\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}\\
&=(1-\vartheta(\theta_t)[s^j_{t+1}=s])(1-2\omega)\\
&+(1-2\omega)\omega^2\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}]
\end{aligned}
$$

Now it follows that

$$
\begin{aligned}
&[-1+(1-2\omega)\vartheta(\theta_t)[s^j_{t+1}=s]\delta x+\\
&(1-\vartheta(\theta_t)[s^j_{t+1}=s])\min_{b^i_t}\left(\Pr(\{s^j=C\}|s,b^i_t,\theta_t)-\Pr(\{s^j=C\}|\hat{s},b^i_t,\theta_t)\right)]\delta x\\
&\geq([-1+(1-2\omega)(2\vartheta(\theta_t)[s^j_{t+1}=s]-1]+(1-2\omega)\omega^2\min\{\vartheta(\theta_{t-1})[\hat{s}],\vartheta(\theta_{t-1})[\breve{s}]\})\delta x
\end{aligned}
$$

since $\vartheta(\theta_t)[\hat{s}(b^j_t)\neq s(b^j_t)]\geq\omega$ the lemma follows.

☑

Proof of Proposition 2: Recall that the radius is the number of mutations required to escape the basin of one of the team strategies *s*, and in Lemma's 2 and 3 we showed that it is given by

$$R[s]\equiv(1-\xi)K\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right].$$

From Ellison [1995], to prove Proposition 2, we need to show that under the hypothesis that for any $\omega>0$ there exists κ, $\bar{u},\underline{v}>0,\bar{\xi}>0$ such that if $\xi<\bar{\xi}$, $K/n\geq\kappa$, $\delta x\geq\bar{u}$, and $\delta^2 x\leq\underline{v}$ the coradius is less than the radius, where the coradius is the minimum number of mutations required to return to the basin of one of the team strategies from any state that has positive recurrence under the limit dynamic ($\varepsilon=0$).

Observe first that if *s* is one of the team strategies, $G(s,\hat{s},\vartheta(\theta_t),\phi_t^i)$ will not decrease if observations in θ_t are replaced by observations of *s* being used. It follows that the basin of *s* is reached as soon as $G(s,\hat{s},\vartheta(\theta_t),\phi_t^i)$ is positive for all $\hat{s}$.

From Lemma 4 a sufficient condition for $G(s,\hat{s},\vartheta(\theta_t),\phi_t^i)$ to be positive for all $\hat{s}$ is

$$\omega\left(-1+\left((1-2\omega)\left(2\vartheta(\theta_t)[s]-1\right)+\vartheta(\theta_t)[\sim S^S]\left((1-2\omega)\omega^3\right)\right)\delta x\right)>0,$$

which may be rewritten as

$$K\vartheta(\theta_t)[s]>K\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-\frac{\omega^3}{2}\vartheta(\theta_t)[\sim S^S]\right].$$

In calculating the co-radius, we can look for the most favorable mutations for returning to the basin of the team strategies. So we suppose that mutations take place by replacing non-strong strategies with the team strategy *s*. Then

$$K\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-\frac{\omega^3}{2}\vartheta(\theta_t)[\sim S^S]\right]+1$$

mutations lead to the basin of *s*. So the condition for the co-radius less than the radius may be written

$$(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]$$
$$>\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-\frac{\omega^3}{2}\vartheta(\theta_t)[\sim S^S]\right]+1/K$$

In other words, if the radius is smaller than the co-radius, it must be that

$$\vartheta(\theta_t)[\sim S^S]$$
$$\leq\frac{2}{\omega^3}\left[\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]\right]+1/K\right]$$

$$\vartheta(\theta_t)[S^S] \geq$$
$$1-\frac{2}{\omega^3}\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}+1/K-(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]\right]$$

Applying Lemma 5 in the same way for any two strong strategies $\hat{s} \neq \breve{s}$ we find the condition for the co-radius to be less than the radius is

$$(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]$$
$$>\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-\frac{\omega^2}{2}\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}+1/K$$

and so if the radius is smaller than the co-radius, it must be that

$$\min\{\vartheta(\theta_t)[\hat{s}],\vartheta(\theta_t)[\breve{s}]\}$$
$$\leq\frac{2}{\omega^2}\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]+1/K\right]$$

We may also check that there are 8 strong strategies. It follows that if the radius is smaller than the co-radius then for some strong strategy $\hat{s}$

$$\vartheta(\theta_t)[\hat{s}] \geq$$
$$1-\frac{2}{\omega^3}\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}+1/K-(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]\right]$$
$$-\frac{14}{\omega^2}\left[\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}-(1-\xi)\left[\frac{1}{2}-\frac{\delta+\delta^2(1+x)/(1-\delta)+(1+\delta x)\xi+\omega}{2(1-2\omega)\omega\delta x}\right]+1/K\right]$$

On the other hand, by Lemma 3, only the team strategies are best response to themselves, and by Lemma 2, as soon as

$$\vartheta(\theta_t)[\hat{s}]>\frac{1}{2}+\frac{1}{2\delta x(1-2\omega)}$$

all players must play a best-response to $\hat{s}$. Moreover, no more than $2n$ observations can be added to the common pool in a single period, so $\vartheta(\theta_t)$ can increase at most by $2n/K$ each period. This implies that if

$$\vartheta(\theta_t)[\hat{s}] > \frac{1}{2} + \frac{1}{2\delta x(1-2\omega)} + \frac{2n}{K}$$

on a best-response cycle, then $\hat{s}$ is a team strategy.

Combining these two facts, we find that the radius must be greater than the co-radius along a best-response cycle if

$$\frac{1}{2} - \left[\frac{2}{\omega^3} + \frac{14}{\omega^2}\right] \times$$
$$\left[\frac{1}{2} + \frac{1}{2\delta x(1-2\omega)} + 1/K - (1-\xi)\left[\frac{1}{2} - \frac{\delta + \delta^2(1+x)/(1-\delta) + (1+\delta x)\xi + \omega}{2(1-2\omega)\omega\delta x}\right]\right]$$
$$> \frac{1}{2\delta x(1-2\omega)} + \frac{2n}{K}$$

or

$$\frac{1}{2} > \frac{1}{\delta x(1-2\omega)}\left[\frac{1}{2} + \frac{2}{\omega^2} + \frac{\xi(1+\delta x)}{\omega^3} + \frac{\delta}{\omega^3(1-\delta)}\right] + \frac{1}{K}\left[2n + \frac{2}{\omega^2}\right] + \frac{\xi}{\omega^2} + \frac{\delta}{1-\delta}\frac{1}{\omega^3(1-2\omega)}$$

The result now follows from algebraic manipulation.

☑

5. Conclusion

We examine particular Prisoners' Dilemma games and a particular inferential process with noise in both the selection of strategies and the transmission of information about past play. In this setting, we show that cooperative play emerges as the unique long-run stochastically stable distribution. We conclude by examining the robustness of these results to variations in both the parameter values and the details of the inferential process.

Our analysis crucially depends on the restriction of the parameter space to those parameter values where next period cooperation is important and later cooperation is not.

An alternative approach would be to assume that a player's next period flags depend only on his current period action and his current opponent's flags. Like the assumption on the discount factor, this means that the consequences of current behavior do not extend beyond next period. However, we see little justification for limiting the strategy space in this way, so prefer to emphasize that our results are valid for a particular range of parameters.

A second essential element of the analysis is the combination of assumptions that enable us to conclude that strategies can be uniquely inferred from observations. To see the importance of these assumptions, consider an alternative scenario where we drop the assumption that players understand that opponents will not use dominated strategies. Then they can no longer infer the strategy from the observation of the action, flag and information service. Suppose that players believe that the action is the same for the unobserved flag as for the observed flag. Even in that case, players can infer the probability distribution of actions conditional on flags, which is the relevant information to solve the optimization problem. Because there is flag noise, typically there will be enough red and green flags in the sample to draw reliable inferences about the conditional distribution of actions. However, occasionally, flag noise will lead to a situation in which there are either only red or only green flags in the public and private samples. If that happens the strategy of always defect becomes optimal. By assumption, the probability that all flags are the same color is much greater than the probability of a mutation. Hence, the system will collapse to always defect much more rapidly than it can move to the team strategy (or any other strategy) through mutation. We conclude in this case that the unique long-run stochastically stable distribution places all weight on always defect, reversing our results.

Although this is an important limitation on our analysis, we think that it is less important than it seems for two reasons:

- It is possible to construct a mechanism by which strategies are directly observable; for example, players can write their strategy down and have an agent play on their behalf. At the end of the period, the paper is revealed. Indeed, we can allow the information system to depend on whether a written strategy of this type is played, or whether the player plays on his own behalf. A variation on the green team strategy which assigns a green flag only if the observable green team strategy is employed, together with its red counterpart will then be the unique long-run steady state. In other words, if some strategies are observable, and others not, the evolutionary process will itself choose the observable strategy, especially if punishment is given for failing to use an observable strategy.
- It is also possible to consider sampling procedures that include and discard observations based on the color of the flag. For example, a rule can be employed that if for a particular information system there are fewer than $\omega/2$ red flags, then observations with red flags are never discarded from the sample unless they are replaced with another red flag observation. This means that inferences about the distribution of actions conditional on flags are always dominated by sample information rather than priors. Moreover, the employment of these sampling procedures makes sense, as the goal of players is to draw inferences based on data rather than priors.

References

Bergin, J. and B. Lipman [1995]: "Evolution with State Dependent Mutations," Queens University.

Ellison, G. [1994]: "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, 61:567-588.

Ellison, G. [1995]: "Basins of Attraction and Long-Run Equilibria," MIT.

Fudenberg, D. and D. K. Levine [1995]: "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19 : 1065-1090.

Fudenberg, D. and D. K. Levine [1998]: *Learning in Games*, (Cambridge: MIT Press), forthcoming.

Fudenberg, D. and D. Kreps [1989]: "Repeated Games with Long-Run and Short-Run Players," MIT #474.

Fudenberg, D. and D. Kreps [1990]: "Lectures on Learning and Equilibrium in Strategic-Form Games," CORE Lecture Series.

Harsanyi, J. [1973]: "Games with Randomly Disturbed Payoffs," *International Journal of Game Theory*, 2: 1-23.

Kandori, M. [1992]: "Social Norms and Community Enforcement," *Review of Economic Studies*, 59: 61-80.

Kandori, M., G. Mailath and R. Rob [1993]: "Learning, Mutation and Long Run Equilibria in Games," *Econometrica*, 61: 27-56.

Monderer, D., D. Samet and A. Sela [1994]: "Belief Affirming in Learning Processes," Technion.

Morris, S., R. Rob and H. Shin [1993]: "p-dominance and Belief Potential," *Econometrica*, 63: 145-158.

Rubinstein, A. [1986]: "Finite automata play the repeated prisioners dilemma," *Journal of Economic Theory*, 39, 1, 83-96.

Young, P. [1993]: “The Evolution of Conventions,” *Econometrica*, 61: 57-83.

Young, P. and D. Foster [1991]: “Cooperation in the Short and in the Long Run,” *Games and Economic Behavior*, 3:145-56.