Repeated Games and Limited Information Processing*
Preliminary and Incomplete

Olivier compte
PSE, Paris

Andrew Postlewaite
University of Pennsylvania

May, 2008

Abstract

Many important strategic problems are characterized by repeated interactions among agents. There is a large literature in game theory and economics illustrating how considerations of future interactions can provide incentives for cooperation that would not be possible in one-shot interactions. Much of the work in repeated games assumes public monitoring: players observe precisely the same thing at each stage of the game. It is well-understood that even slight deviations from public monitoring increase dramatically the difficulty the problems players face in coordinating their actions. Repeated games with private monitoring incorporate differences in what players observe at each stage. Equilibria in repeated games with private monitoring, however, often seem unrealistic; the equilibrium strategies may be highly complex and very sensitive to the fine details of the stochastic relationship between players' actions and observations. Furthermore, there is no realistic story about how players might arrive at their equilibrium strategies.

We propose an alternative approach to understanding how people cooperate. Each player is endowed with a mental system that processes information: a mental system consists of a number of psychological states and a transition function between states that depends on observations made. In this world, a strategy is just a function from states to actions. Our framework has the following desirable properties: (i) players restrict attention to a relatively small set of simple strategies. (ii) the number of strategies that players compare is small enough that players might ultimately learn which perform well.

We find that some mental systems allow agents to cooperate under a broad set of parameters, while others are not conducive to cooperation.

1. Introduction

Cooperation is ubiquitous in long-term interactions: we share driving responsibilities with our friends, we offer help to relatives when they are moving and we write joint papers with our

*Compte: Paris School of Economics, 48 Bd Jourdan, 75014 Paris (e-mail: compte@enpc.fr); Postlewaite: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6207 (e-mail: apostlew@econ.upenn.edu). We thank the participants at many seminars at which this paper was presented for helpful comments. A large part of this work was done while Postlewaite visited the Paris School of Economics. The authors thank the Gould Foundation for financial support, and Postlewaite thanks the National Science Foundation for financial support.

colleagues. The particular circumstances of an agent's interactions vary widely across the variety of our long-term relationships but the mechanics of cooperation are usually quite simple. When called upon, we do what the relationship requires, typically at some cost. We tend to be upset if our partner seems not to be doing his part and our willingness to cooperate diminishes. We may be forgiving for a time but stop cooperating if we become convinced the relationship is one-sided. We sometimes make overtures to renew the relationship when opportunities arise, hoping to rejuvenate cooperation. Incentives to cooperate stem from a concern that the relationship would temporarily break down, while incentives to be less cooperative when the relationship feels one-sided stem from the fear of being taken advantage of by a non-cooperative partner.

These rules of thumb seem to be conducive to cooperation under a broad range of circumstances, including those in which we get only a noisy private signal about our partner's efforts in the relationship. We are interested in the following questions and in providing a framework to address them. Do there exist rules of thumb that perform well in the sense that despite agents' search for individual gains players are able to cooperate when there are gains from doing so *under a large variety of circumstances* (payoffs, signal structures)? What rules of thumb perform better than others?

The literature on repeated games that has addressed the issue of cooperation aims at answering the dual question: For given parameters of a particular game (payoffs, signal structures), what are the set of stable rules of behavior (i.e. equilibrium rules from which neither player would want to deviate). Analysis that begins with a specific game will likely not provide an answer to the question of what rules perform well in a large variety of circumstances.

The theory of repeated games has provided important insights about repeated interactions but is unsatisfactory in several regards. When signals are private, the quest for "stable" rules of behavior (or equilibria) has sometimes produced complex strategies that are finely tuned to the parameters of the game (payoffs, signal structure).[1] If the parameters of the game are changed slightly the rule fails to remain stable.[2]

Additionally, there is no realistic story of how players would arrive at the proposed equilibrium strategies. It seems extremely implausible that players could compute appropriate strategies through introspection. Furthermore, equilibrium strategies in repeated games with private signals typically rely on my knowing not only the distribution of signals I receive conditional on the other player's actions, but also the distribution of his signals given my actions, something I never observe. Even if one entertains the possibility that players compute equilibrium strategies through introspection there is the question of how the players might know these signal distributions. One might posit that players could "learn" the equilibrium strategies, but there are (at least) two difficulties. First, if players are to learn how to behave in repeated situations this should be incorporated into the modelling of the problem. Second, the set of strategies is huge and it is difficult to see how a player might learn which strategies work well.

We propose an alternative approach to understanding how people cooperate. Players restrict attention to a relatively small set of simple strategies. We are ultimately interested in the actual set of strategies players restrict attention to, but as a first step our goal is to find sets of strategies that have the following desirable properties: (i) the number of strategies in the set should be small enough that players might ultimately learn which perform well; (ii) the strategies should be simple in an intuitive sense; (iii) the sets should allow agents to cooperate under broad

[1] See, e.g., Piccione (2002) and Ely and Valimaki (2002).

[2] In other words, equilibria are not strict.

circumstances.

One can think of the choice of actions as arising in a hierarchical behavior system in which the top level (the set of strategies the player considers) describes the ways in which a person might play in a broad set of games. The second level of the hierarchical system is activated when an individual is faced with a specific game. At this point, the player chooses from the set of strategies determined at the top level. While the set of strategies considered remains unchanged as the "fine details" of the games vary, the choice from that set may.

The goal is a realistic description of cooperation when people are strategic. The games we play vary and our knowledge of the structure of the game may be limited. We need to adjust play through experimentation to learn which strategies perform well. Restrictions on the set of strategies allow players to identify best responses easily, and eventually adjust play as a function of the particular parameters of the game they are currently playing.

1.1. Strategy restrictions

The restrictions on the strategies available to players are a crucial element of our approach. We are not interested in arbitrary restrictions, but rather, on restrictions that might arise naturally. An individual's action in any period is generally assumed to be a function of the history to that point. Here, we restrict strategies to be functions of the *informational state* a player is in and limit the number of informational states available to a player. This restricts a player to behaving the same way for all histories of the game that lead to the same informational state.[3]

We think of the set of informational states not to be a choice variable, but rather a natural limitation of mental processing. We might feel cheated if we have put effort into a relationship and get signals that the other is not reciprocating. We can think of those histories in which one feels cheated as leading to a mental state (U)pset, and those histories in which one doesn't feel cheated as leading to a mental state (N)ormal. A *mental system* is the *set of mental states* one can be in along with a *transition function* that describes what combinations of initial mental state, actions and signals in a period lead to specific updated mental states. We will assume in most of what we do that the transition function does not depend on the fine details of the game – the payoffs and the monitoring structure – but in principle it might. For example, in circumstances in which it is extremely costly for my partner to put in effort, I may not become upset if he does not seem to be doing so. However, a fundamental aspect of the transition function is that the individual does not have control over it.

While the mental system may be the same across a variety of games, how one *responds* to being upset may be situational, that is, may depend on the particular game one is involved in, as well as on the behavior of the other player. If the cost of cooperation is very small, one might be hesitant to defect in state U and risk breaking a relationship that is generally cooperative, but not hesitate when the cost is large; whether in state U or state N, the individual may either cooperate or defect. Thus, while a player's *available* strategies depend only on that player's mental process - hence not necessarily on the fine details of the payoffs and informational structure of the game – the strategy he chooses will typically be sensitive to the specifics of the game at hand.

Our view is that there are limits to peoples' cognitive abilities, and evolution and cultural

[3] For many problems restricting a player to a finite number of informational states is natural. If there is a finite number of signals a player can receive following the play of a game in any period, and if players have bounded recall, the assumption that an individual has a finite number of informational states is without loss of generality.

indoctrination determine an individual's mental system consisting of the states he can be in and the transition function that moves him from one state to another. Children experience a large number of diverse interactions, and how they interpret those experiences are affected by their parents and others they are (or have been) in contact with. A parent may tell his child that the failure of a partner to have reciprocated in an exchange is not a big deal and should be ignored, or the parent can tell the child that such selfish behavior is reprehensible and inexcusable. Repeated similar instances shape how the child interprets events of a particular type. Even in the absence of direct parental intervention, observing parental reactions to such problems shape the child's interpretations.[4]

2. Model

Gift exchange.

There are two players who exchange gifts each period. Each has two possible actions available, $\{D, C\}$. Action D is not costly and can be thought of as no effort having been made in choosing a gift. In this case the gift will not necessarily be well perceived. Action C is costly, and should be interpreted as making substantial effort in choosing a gift; the gift is very likely to be well-received in this case. The expected payoffs to the players are as follows:

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

L corresponds to the cost of effort in choosing the "thoughtful" gift: you save L when no effort is made in choosing the gift.

Signal structure.

We assume that there are two possible private signals that player i might receive, $y_i \in Y_i = \{0, 1\}$, where a signal corresponds to how well player i perceives the gift he received. We assume that if one doesn't put in effort in choosing a gift, then most likely, the person receiving the gift will not think highly of the gift. We will refer to $y = 0$ as a "bad" signal and $y = 1$ as "good".

Formally,

$$p = \Pr\{y_i = 0 \mid a_j = D\} = \Pr\{y_i = 1 \mid a_j = C\}$$

We will assume that $p > 1/2$ and for most of the main text analysis we consider the case where p is close to 1.

In addition to this private signal, we assume that at the start of each period, players receive a public signal $z \in Z = \{0, 1\}$, and we let

$$q = \Pr\{z = 1\}.$$

We discuss the role of the signal z below.

[4] It is quite possible that there is a conflict between the two; parents may want to indoctrinate their children to be different from themselves.

2.1. Strategies

As discussed above, players' behavior in any period will depend on the previous play of the game, but in a more restricted way than in general. There is a finite set of possible *informational states*, S_i, that player i can be in, where a given informational state is a set of histories. Informational states capture the bounds on players' memories of the precise details of past play. For example, a player who has gotten one bad signal in the past twenty periods should not be assumed to remember this occurred in period sixteen or period seventeen. For simplicity, we assume that in the current example the players can be in one of two states U(pset) or N(ormal). The names are chosen to convey that at any time player i is called upon to play an action, he knows the mood he is in, which is a function of the history of (own) play and signals, but does not condition his action on finer details of the history.[5] One can interpret the restriction to strategies that are constant across the histories that lead to a particular informational state as being a limit on the player's memory or simply as a rule of thumb the player uses. S_i is exogenously given, not a choice. Player i's set of pure strategies is

$$\Sigma_i = \{\sigma_i, \sigma_i : S_i \longrightarrow A_i\}.$$

The particular state in S_i that player i is in at a given time depends on the previous play of the game. The *transition function* for player i is a function that determines the state player i will be in at the beginning of period t as a function of his state in period $t-1$, his choice of action in period $t-1$, and the outcome of that period – the signals y_i and z. As is the set of states for player i, the transition function is exogenous. A player who has made an effort in his choice of gift but receives a bad signal may find it impossible not to be upset, that is, be in state U.

We assume the transition function for the example, which we will refer to as the *leading example* below, is as in the figure below.

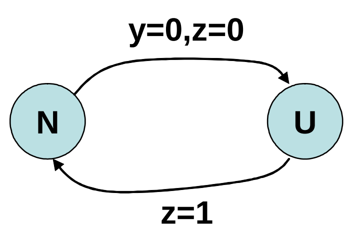


Figure 1: Transition function

This figure shows which combinations of actions and signals will cause the player to move from one state to the other. If player i is in state N, he remains in that state unless he receives signals

[5] For expository ease we assume that an individual's payoffs depend on outcomes, but not on the state he is in. The names that we use for the states suggests that the state itself could well be payoff relevant: whatever outcome arises, I will be less happy with that outcome if I'm upset. Our framework can easily accommodate state-dependence, and the qualitative nature of our conceptual points would be unchanged if we did so. We discuss this further in the discussion section below.

$y = 0$ and $z = 0$, in which case he transits to state U. If i is in state U, he remains in that state until he receives signal $z = 1$, at which point he transits to state N regardless of the signal y.[6]

To summarize, a player is endowed with a *mental system* that consists of a set of informational states the player can be in and a transition function that describes what triggers moves from one state to another. Our interest is in finding stable patterns of behavior. Our structure requires that players' strategies are stationary: they do not depend on the period. This rules out strategies of the sort "Play D in prime number periods and play C otherwise", consistent with our focus on rules of thumb that prescribe behavior as a function of the information close to hand at the time choices are made.

Our candidate behavior for the players will be as follows. For player i,

$$\begin{aligned} \sigma_i(N) &= C \\ \sigma_i(U) &= D. \end{aligned}$$

That is, player i plays C as long as he receives a gift that seems thoughtful, that is $y_i = 1$, or when $z = 1$. He plays D otherwise. Intuitively, player 1 triggers a "punishment phase" when he saw $y_1 = 0$, that is, when he didn't find the gift given to him appropriate. This punishment phase ends only when signal $z = 1$ is received.

The public signal z gives the possibility of "resetting" to relationship to a cooperative mode. If the signal z is ignored and the mental process is defined by

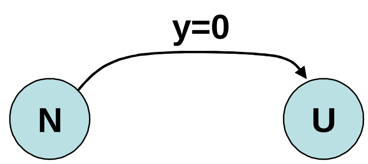


Figure 2: No "resetting" signal

then eventually, because signals are noisy, with probability 1 the players will get to state U under the proposed strategy and this will be absorbing: there would be nothing to change their behavior. The signal z allows for possible recoordination back to state N (and possibly cooperation).

In our leading example, players stop being upset for exogenous reasons. Alternatively, in a two-state mental system the players could move from state U back to state N after seeing a good signal: you stop being upset as soon as you receive a nice gift.

Formally, players may be either in state N or in state U, but are endowed with the following transition function.

[6] For this particular example, transitions depend only on the signals observed, and not on the individual's action. But in general it might also depend on the individual's action.

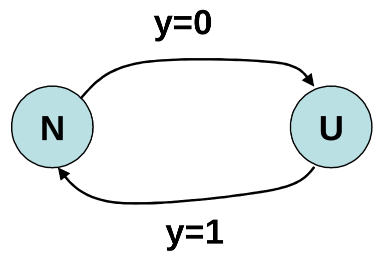


Figure 3: Forgiving transition function

A player endowed with this alternative mental process, who would cooperate in N and defect in U, would be following a TIT for TAT strategy.

2.2. Illustrative experiment

Before continuing with the formal description of our model, it is useful to give a real-world example to illustrate our idea of a mental system. Cohen *et al.* ran several experiments in which participants (students at the University of Michigan) were insulted by a confederate who would bump into the participant and call him an "asshole". The experiment was designed to test the hypothesis that participants raised in the north reacted differently to the insult than did participants raised in the south. >From the point of view of our model, what is most interesting is that the insult triggered a physical response in participants from the south. Southerners were upset by the insult, as shown by cortisol levels, and more physiologically primed for aggression, as shown by a rise in testosterone. We would interpret this as a transition from one mental state to another, evidenced by the physiological changes. This transition is plausibly not a choice on the participant's part, but involuntary. The change in mental state that is a consequence of the insult was followed by a change in behavior: southerners were more likely to respond in an aggressive manner following the insult than were northerners.

The physiological reaction to an insult – what we would think of as a transition from one state to another – seems not to be "hard-wired": physiological reactions to insults were substantially lower for northern students than for southern students. Indeed, the point of the Cohen *et al.* paper is to argue that there is a southern "culture of honor" that is inculcated in small boys from an early age. This culture emphasizes the importance of honor and the defense of it in the face of insults. This illustrates the view of our model expressed above that the transition function in our model can be thought of as culturally determined, but fixed from the point of view of an individual at the time decisions are taken.

2.3. Ergodic distributions and strategy valuation

For any pair of players' strategies there will be an ergodic distribution over the pairs of actions played. While in general the ergodic distribution may depend on the initial conditions, we restrict attention to transition functions for which the distribution is unique. The ergodic distribution gives the probability distribution over payoffs in the stage game, and we take the payoff to the players to be the expected value of their payoffs given this distribution.

Formally, define a state profile s as a pair of states (s_1, s_2). Each strategy profile σ induces transition probabilities over state profiles: by assumption each state profile s induces an action profile $\sigma(s)$, which in turn generates a probability distribution over signals, hence, given the transition functions T_i, over next period states. We denote by Q_σ the transition matrix associated with σ, and by ϕ_σ the ergodic distribution over states induced by σ. That is, $\phi_\sigma(s)$ corresponds to the (long run) probability that players are in state s.[7]

We associate with each strategy profile σ the value induced by the ergodic distribution. This corresponds to computing discounted expected payoffs, and taking the discount to 1.[8] We denote by $v(\sigma)$ this value (vector). Thus,

$$v(\sigma) = \sum_s g(\sigma(s))\phi_\sigma(s)$$

where $g(\sigma(s))$ is the payoff vector induced by the strategy profile σ for state profile s.

Equilibrium.

Definition: We say that a profile $\sigma \in \Sigma$ is an equilibrium if for any player i and any strategy $\sigma_i' \in \Sigma_i$,

$$v_i(\sigma_i', \sigma_{-i}) \leq v_i(\sigma).$$

This is a weaker notion of equilibrium than traditionally used in repeated games because of the restriction on the set of strategies to be mappings from S_i to A_i.[9] Also note that σ_i as defined should not be viewed as a strategy of the repeated game.[10]

We consider next the ergodic distribution induced by our candidate equilibrium strategy σ for p close to 1 and examine in turn possible deviations from that strategy. The transition over state profiles induced by σ is illustrated in figure 4.

[7] Formally, $Q_\sigma(s', s)$ is the probability that next state profile is s' when the current state is s, and the vector ϕ_σ solves $\phi_\sigma(s') = \sum_s Q_\sigma(s', s)\ \phi_\sigma(s)$.

[8] When discounting is not close to one, then a more complex valuation function must be defined: when σ is being played, and player i evaluates strategy σ_i' as compared to σ_i, the transitory phase from ϕ_σ to $\phi_{\sigma_i', \sigma_{-i}}$ matters. Note however that the equilibria we will derive are strict equilibria, to they would remain equilibria under this alternative definition for discount factors sufficiently close to 1.

[9] We restrict attention to pure strategies. However, our definitions can be easily generalized to accomodate mixed actions, by re-defining the set A_i appropriately, and having it include mixed actions. However, the spirit of our approach is that players should adjust play from experience, by checking from time to time the performance of alternative strategies. So if mixed actions are to be allowed, only few of them, rather than the whole set of mixed actions, should in our view be considered.

[10] A strategy of the repeated game is a mapping from histories to actions. The strategy σ_i, along with the mental system (S_i, T_i) would induce a repeated game strategy, once the initial state is specified.

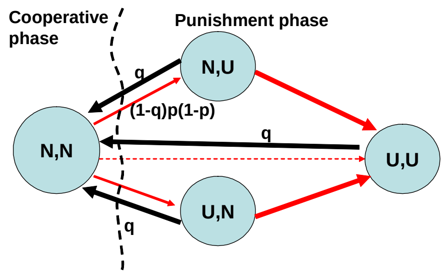


Figure 4: Transition over state profiles under σ.

When players follow our candidate equilibrium strategy, they alternate between cooperation and punishment phases. The probability of switching from cooperation to a punishment phase is $\pi = (1-q)(1-p^2)$ (since switching occurs when either player receives a bad signal and $z = 0$). The probability of switching from punishment to cooperation is q. Hence cooperative phases last on average $1/\pi$ periods, while punishment phases last on average $1/q$ periods.[11]

When player i plays D at both N and U, player j continues to alternate between phases of cooperation and defection. This is illustrated in figure 5.

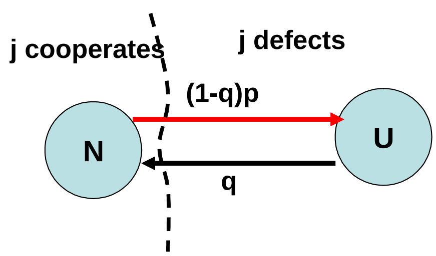


Figure 5: i defects always

Player i gets a higher payoff in cooperation phases; however, those phases are now much shorter, as his opponent switches to state U with probability $(1-q)p$. For p close to 1, a defection at N almost certainly generates a bad signal, which triggers a punishment phase of length $1/q$ with probability $1-q$, hence an expected cost

$$\Delta = \frac{1-q}{q}$$

corresponding to a per-period decrease in payoff of 1 for an expected number of period equal to Δ.

Deviation is deterred if

$$L < \Delta. \tag{2.1}$$

[11] This is because punishment lasts $1+T$ periods with probability $q(1-q)^T$, and because

$$1 + \sum_T Tq(1-q)^T = 1/q.$$

When player i plays C at both N and U, he avoids triggering some punishment phases, illustrated in figure 6.

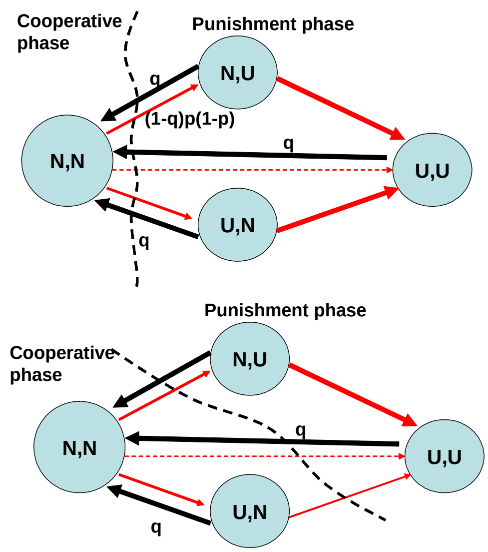


Figure 6: i cooperates always

However, he remains cooperative while player j is in a punishment phase. So L must be high enough for this option to be unattractive. More precisely, conditional on both players being in state N, there are events where only player i receives a bad signal, and events where only player j receives a bad signal.[12] Under the first event, player i initially gets 1 instead of $1+L$, however he avoids the punishment phase, hence he makes a net gain of $\Delta - L$. Under the second event, nothing changes in the first period (because player i is still in state N), but he then gets $-L$ instead of 0 as long as the punishment phase lasts,[13] hence an expected cost equal to $L(\frac{1}{q}-1) = L\Delta$. Since these two events have equal probability, playing C at N and U is not a profitable deviation if

$$\begin{aligned}\frac{1}{2}(\Delta - L) + \frac{1}{2}(-L\Delta) &< 0\text{, that is} \\ L &> \frac{\Delta}{1+\Delta}.\end{aligned}$$

To summarize, for p close to 1, there is a range of parameters (q, L) for which the proposed strategy is an equilibrium strategy.[14] This range of parameters is such that $\frac{\Delta}{1+\Delta} < L < \Delta$,

[12] There are also events where both receive bad signal, but when p is close to 1, these are very unlikely events, and we can ignore them here. However, they would affect the computation in the general case where p is not close to 1.

[13] This is because when p is close to 1, player i switches to U with probabilty close to 1, hence he would have started playing D under the candidate equilibrium profile, while here, he does not.

[14] It is easy to check that playing D at N and C at U gives a lower value than always playing D.

where $\Delta = 1/q - 1$. It is easy to check that outside this range, the only equilibrium entails both players defecting at both states.

Similar analysis can be performed for other values of p. The following figure delineates, for q set to 0.3, the set of parameters (p, L) for which σ is an equilibrium:

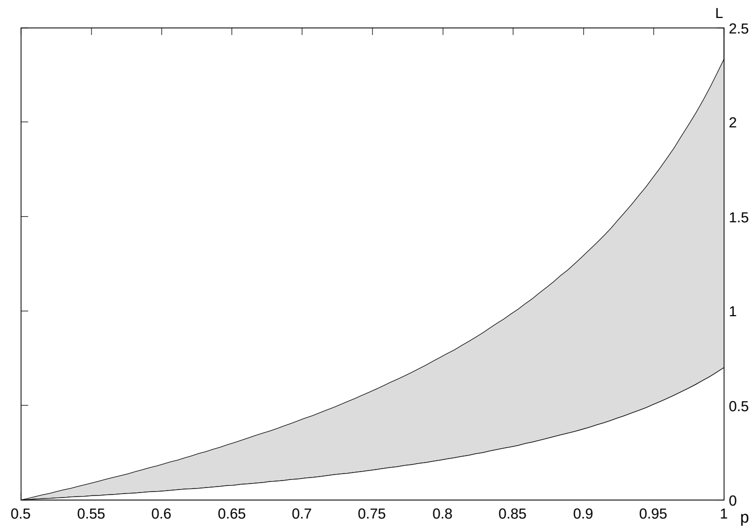


Figure 7: $p - L$ combinations that allow cooperation

3. Discussion of example

This example illustrates how cooperation can be achieved when strategies are constrained. Before going on, it is useful to compare this approach with the standard approach and discuss why cooperation is difficult when strategies are not constrained. We will then show that there are other two-state mental processes that seem reasonable and yet fail to support cooperation. Following that, we discuss richer mental processes.

3.1. Comparison with standard approach

In the example, the strategies "play C when in N and play D when in U" are in equilibrium: they are best responses to each other under the restriction that players use strategies in Σ_i that condition actions on states. These strategies however would not be best responses to each other without restrictions. Consider for example the first period following a punishment phase (players have just seen signal $z = 1$ and have consequently just returned to state N). Suppose player 1 receives a bad signal. Player 1 then transits to state U, and the equilibrium strategy in the example calls for him to play D, sending the play into the punishment phase again. However, at this first period after $z = 1$, player 2 was in state N as well, and most likely will remain in state N in the next period. So player 1 would be better off not triggering a punishment.

One might hope to get a standard "unconstrained" equilibrium by a simple modification of our equilibrium strategy profile, allowing players to condition play on longer histories, and, for example, on whether they are just returning to N or not. In this way, a player could ignore signals in the first period following a return to state N. However, the best response to someone who behaves in that way is to play D in the first period following a return to state N. Equilibrium play in this first period following $z = 1$ would thus have to involve mixed strategies, and later play would have to be finely tuned to the history of signals later received. For example, a signal received in the first period following a return to N would have to be interpreted differently than a signal received in later periods, in a way that takes into account the fact that the opponent reacts differently to signals received shortly after a return to a cooperative phase than to signals received later. Thus, one cannot modify the strategies in our example in a simple way to effect a standard unconstrained equilibrium.

One difficulty in the search for strategies that support cooperation when strategies are not constrained (i.e. the standard approach) is precisely the possibility that beliefs are sensitive to the histories of signals received, and possibly the entire history of signals received. In general, there is then no natural way for players to be in one of a small number of states of mind in an equilibrium of the repeated game: distinct beliefs over the opponent's continuation play are likely to generate distinct best responses, hence, according to our interpretation, distinct states of mind. But since the number of distinct beliefs typically grows over time, so does the number of states of mind.[15]

In contrast, our approach takes information states and the transition across states as exogenous. In other words, we take as given the way players aggregate past information about play, that is, how they pool past histories. The strategic issue thus reduces to determining how one should play at each informational state, that is, choosing a mapping σ_i from states to actions, a much simpler task. In particular, our analysis is not based on a player forming beliefs about his opponent's future play, and maximizing given those beliefs. Rather, we just compute the long-run value associated with any possible deviation, and this long-run value does not depend on current beliefs.[16]

3.2. Another two-state mental process

We compare our leading example to the forgiving transition described earlier, whereby a good signal makes a player switch back to the normal state.

[15]See however the discussion of Mailath and Morris (2002), Phelan and Skrzypacz (2006) and Kandori and Obara (2007) and the belief free literature in the related literature Section below.

[16]One can define rational beliefs for a player at a given informational state s_i, for a given profile σ, by considering the ergodic distribution over state profile $\phi_\sigma(s_1, s_2)$ and by computing the conditional probabilities $\Pr_{\phi_\sigma}\{s_j \mid s_i\}$. Such a belief corresponds to the average of the beliefs given the possible histories leading to state s_i (the average taken with respect to the ergodic induced by σ).

Note however that while beliefs can be defined, we don't have in mind that players would know the mental process and strategy of the other player and use that knowledge to compute a belief. Beliefs should jusr be viewed as conditional probabilities, that is, as a way to understand correlations across each player's state, under a given strategy profile.

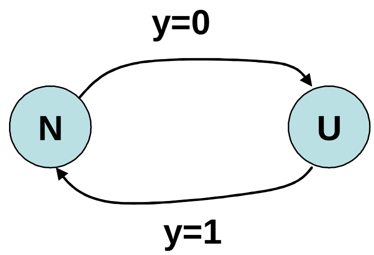


Figure 3: Forgiving transition function

We show below that with such a mental process, (for almost all values of p) the only equilibrium entails both players defecting in both states.

Consider first the case where player 2 follows the strategy σ that plays C in N and D in U. If player 1 adopts the same strategy, then by symmetry, the induced ergodic distribution puts identical weight on (NN) and (UU) on one hand, and on (NU) and (UN) on the other hand. (Intuitively, the dynamic system has equal chances of exiting from (NN) as it has of exiting from (UU).)

$$\phi_{\sigma,\sigma}(NN) = \phi_{\sigma,\sigma}(UU) \text{ and } \phi_{\sigma,\sigma}(NU) = \phi_{\sigma,\sigma}(UN) \tag{3.1}$$

The value to player 1 from following that strategy is thus

$$\begin{aligned} v(\sigma,\sigma) &= \phi_{\sigma,\sigma}(NN) + (1+L)\phi_{\sigma,\sigma}(UN) - L\phi_{\sigma,\sigma}(NU) \\ &= \phi_{\sigma,\sigma}(NN) + \phi_{\sigma,\sigma}(UN) = \frac{1}{2}(\phi_{\sigma,\sigma}(NN) + \phi_{\sigma,\sigma}(UN) + \phi_{\sigma,\sigma}(NU) + \phi_{\sigma,\sigma}(UU)) \\ &= \frac{1}{2}. \end{aligned}$$

Now if player 1 cooperates in both states (σ^C), player 2 will switch back and forth between states N and U, spending a fraction p of the time in state N. The value to player 1 from following that strategy is thus:

$$v(\sigma^C, \sigma) = p + (1-p)(-L)$$

and it exceeds 1/2 if

$$p > 1 - \frac{1}{2(1+L)}.$$

If player 1 defects in both states (σ^D), player 2 will again switch back and forth between states N and U, but now spending a fraction $1-p$ of the time in state N. The value to player 1 from following that strategy is thus:

$$v(\sigma^D, \sigma) = (1-p)(1+L)$$

which exceeds 1/2 as soon as $p < 1 - \frac{1}{2(1+L)}$.

Finally, if player 1 follows the strategy $\widehat{\sigma}$ that plays D in N and C in U, then, as above, the dynamic system has equal chances of exiting from (NN) as it has of exiting from (UU). Therefore, equalities (3.1) hold for the profile $(\widehat{\sigma}, \sigma)$, and the value to player 1 from following $\widehat{\sigma}$ thus remains equal to 1/2. It follows that unless $p = 1 - \frac{1}{2(1+L)}$, the strategy profiles (σ, σ) and

$(\widehat{\sigma}, \sigma)$ cannot be equilibria. Similar considerations show that the strategy profile $(\widehat{\sigma}, \widehat{\sigma})$ cannot be an equilibrium. As a result, only strategy profiles that are constant across state may be in equilibrium, hence the only equilibrium entails defecting in both states. To summarize:

> **Proposition:** *If $p \neq 1 - \frac{1}{2(1+L)}$, and if each players' mental process is as defined above, then the only equilibrium entails defecting in both states.*

3.3. Adding more mental states

Players in our example had two possible mental states and two possible (pure) actions, which limited them to four pure strategies. This clearly limits them both in the existence of strategies that might lead to cooperation and in the possible profitable deviations. Adding a state can allow for more flexible reaction to signals that might permit cooperation which would have been impossible with only two states. For example, when signals are not very informative and q is small, the prescribed strategies in the example may not be an equilibrium. When there is substantial probability that I received a bad signal when the other player chose C, I might prefer not to trigger a punishment phase that could last for a long time. Hence the combination of not very accurate signals of the opponent's action and low probability of escaping the punishment regime may preclude an equilibrium in which the two cooperate.

Adding a state can allow cooperation that would be impossible with only two states for some parameters. Suppose there is a state M in addition to the states N and U, and define transitions as follows:

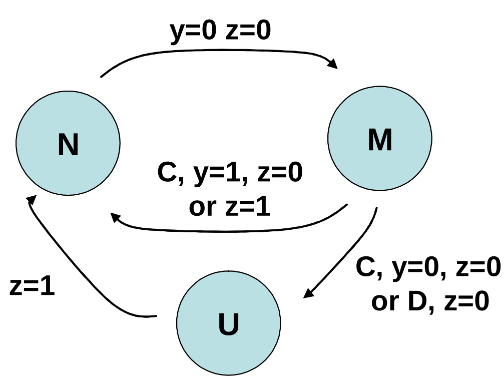


Figure 8: "Slow" transition to state U

The interpretation of this is that if a player in state N gets a bad signal, he transits to state M, rather than state U as in the leading example. If the first signal a player in state M gets is bad, he transits to state U, and if it is positive, he returns back to state N as though the bad signal that sent him to M never occurred. Players remain in state U whatever action is played or private signal received, until signal $z = 1$ occurs. Signal $z = 1$ always sends players back to state N. The additional state, along with the amended transition function, allows for strategies that punish an opponent if there is some evidence he is not cooperating, but the evidence that triggers punishment can be either a single bad signal (e.g., the strategy plays C in N only, which we denote by $\bar{\sigma}$), or two consecutive bad signals (e.g., the strategy plays C in N and M, and D in U , which we denote by η). We find that the strategy profile (η, η) that entails both players cooperating in states N and M and punishing in state U is an equilibrium for some parameters for which cooperation is impossible with two states.

The following figure illustrates this. We have set the value q to 0.3. For the set of parameters (p, L) in the shaded area, (η, η) is an equilibrium profile. For the parameters (p, L) between the thinner lines, (σ, σ) was an equilibrium profile in our simple two state case.

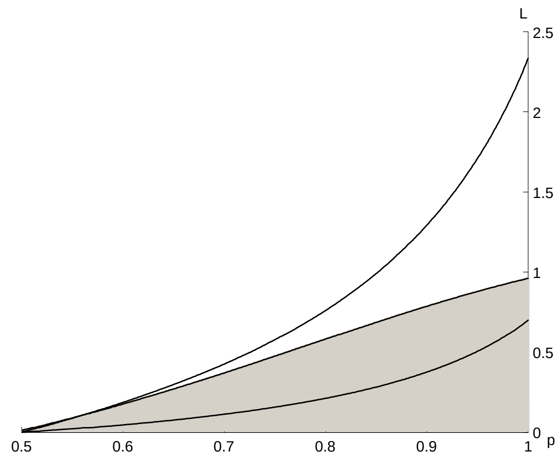


Figure 9: values of (p, L) for which (η, η) is an equilibrium when $q = 0.3$.

We make several comments about this example.

- Intuitively, the additional state has two effects. Under η, it allows for longer cooperation phases by reducing the probability that the relationship transits to a punishment phase (such transitions are inevitable even when players cooperate), and it allows players to be more cautious about deciding that their opponent has switched to U by waiting for two consecutive negative signals. This has two effects on incentives. It makes defecting always more attractive because it takes two consecutive bad signals to make the other player switch to U. It also makes cooperating less attractive, because when a player eventually switches to U, it is very likely that the other player is already in U, hence incentives to play D in U are stronger. This is illustrated in Figure 9: compared to the two-state case, the $p - L$ region where cooperation is possible (under (η, η)) shifts downward.

- Allowing for a stochastic transition function is an alternative modification of the mental system in our leading example that could accomplish the two effects above. In the example, a player transits from state N to state U if he observes a bad signal (and $z = 0$). Suppose the transition function is changed so that a player transits from state N to state U with probability $1 - \mu$ after seeing a bad signal. This will make transition to U less likely, hence incentives to play D at N stronger. However, when a player finds himself in state U, it is more likely that his opponent is in state U as well, hence players will have stronger incentives to play D at U. So for some parameters, this can make cooperation possible when it would not have been possible with the deterministic transition function in the example.

- In addition to making cooperation possible for parameters under which it was impossible with only two states, a third state can increase the *value* of cooperation for parameters even if cooperation was possible with two states (i.e. $v(\eta, \eta) > v(\sigma, \sigma)$, as the punishment state is triggered less often.

- Adding a state is not unambiguously good, however. As mentioned above, an additional state allows not only for more complex strategies to achieve cooperation, but more complex strategies for deviating. Despite the fact that both (σ, σ) and $(\bar{\sigma}, \bar{\sigma})$ generate the same behavior (under both profiles, once a player observes a bad signal, he continues to defect until a signal $z = 1$ arises), there are parameters for which (σ, σ) is an equilibrium, but $(\bar{\sigma}, \bar{\sigma})$ is not. For $(\bar{\sigma}, \bar{\sigma})$ to be an equilibrium, players must have an incentive to play D in M. This entails a strictly tighter constraint because playing C at M is less costly than playing C at U: when the opponent is already in U, the player is likely to get a negative signal and switch to U next period. For p close to 1 for example, the cost will be for only one period rather than for the duration of the punishment phase. Formally, the incentive constraint becomes:

$$\frac{1}{2}(\Delta - L) + \frac{1}{2}(1 - q)(-L) \geq 0.$$

Rearranging terms, we have

$$L \geq \frac{\Delta}{2 - q} = \frac{1 - q}{q(2 - q)},$$

a condition that is always more restrictive than $L \geq \frac{\Delta}{1+\Delta} = 1 - q$. Figure 10 shows how the incentive condition changes for other values of p.

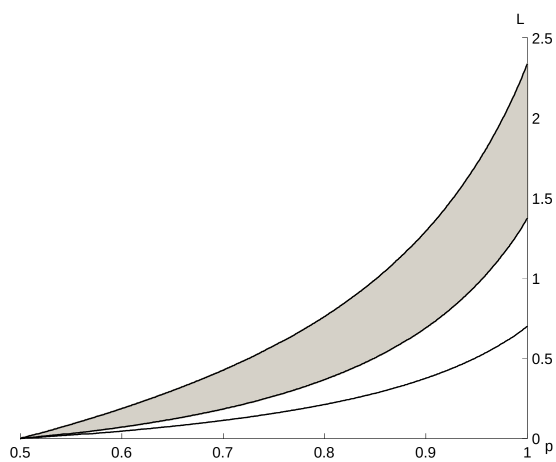


Figure 10: values of (p, L) for which $(\bar{\sigma}, \bar{\sigma})$ is an equilibrium when $q = 0.3$.

3.4. Available strategies and One-shot deviations

We discussed above that if there were no restrictions on strategies a player would like to deviate from the equilibrium strategy in the example by not playing D if he received a bad signal the first period after he returned to state N. Players in our framework are restricted to choosing which action they will play when in state N and what action they will play in state U. This restricts the set of strategies available, including the deviations available to them.

Our approach puts limits on the strategies available to players in various ways. First, as in standard Bayesian models, a player's state pools many histories of the game, and he is restricted to behave in the same way every time he finds himself in a particular state. For example, a player in our main example can choose to play C rather than D when he is in state U, but he cannot play C for some subset of the histories that lead to state U and play D for other histories that lead to U. If players are to behave differently for different subsets of histories, they must have different mental states that allow them to distinguish the sets. In the previous subsection, state M plays precisely that role. With this additional state, the player has available a strategy that allows him to play C after a bad signal (instead of going to state U and playing D), and condition future play on the realization of next period's signal, i.e. reverting to state N after a good signal, or switching to state U after a second bad signal. For any expanded mental system, however, there will always be strategies for which deviations from that strategy cannot be accomplished within that system.

Second, some deviations are not available because current actions have consequences on the player's continuation state of mind. In the previous subsection, if a player play D in state M, he moves to state U. He cannot decide, upon seeing the signal y_i, whether he wants to continue to behave as though he were still in state N. He will be in state U regardless because in this example, a defection triggers a switch to state U. In other words, as in models with imperfect recall, after playing D in M, the player will be in state U and does not distinguish whether he is in U because he received two consecutive bad signals or because he played D in M.

Because of these restrictions, we may find equilibria for which there would be a profitable one-shot deviation; the player may simply not have available a strategy that would mimic the standard one-shot deviation.

Conversely, one might expect that a version of the one-shot deviation principle would hold in our framework: if it does not pay for me to deviate one time when I am in state U (and play C once), then it should not be the case that it is profitable for me to deviate each time I am in state U. This is not the case however. Returning to our main example, suppose that I am in state U and I consider deviating from my strategy one time only and play C. Returning to my equilibrium strategy after this involves my playing D while in U. At best, this will lead to a delay in entering the punishment phase by one period. On the other hand, if I play C each time I am in U I avoid completely triggering the punishment phase any time only I receive a bad signal. Formally, deviating one time at U (and playing C) is not profitable when

$$\Pr_\sigma(s_j = U \mid s_i = U)(-L) + \Pr_\sigma(s_j = N \mid s_i = U)(q(-L) + (1-q)(1)) < 0$$

or equivalently, when

$$L > \frac{q/2}{q(1-q/2)}(1-q).$$

This condition is less stringent than the equilibrium condition we found earlier.

Comments

1. If a player has a profitable one-shot deviation, he will have a profitable mixed strategy deviation. Choosing with small probability the action in a profitable one-shot deviation will give

a gain to the player without the cost entailed in playing that action each time the player is in that state. If we did not restrict to pure strategies, or allowed players to play differently with small probability, then there would be no profitable one-shot deviations.

2. In our framework, one could define an alternative notion of a one-shot deviation whereby a player could decide to change state once (from U to N, say, and continue behaving as if he were now in state N). Such deviations could be profitable, depending on whether it would be profitable for a player to add a small amount of noise to a particular transition.

4. Robustness

The example was kept simple in a number of ways to make clear how cooperation could be achieved when strategies were restricted. Some of the simplifications are not particularly realistic, but they can be relaxed without affecting the basic point that cooperation is possible even when agents get private signals if strategies are restricted. We discuss next several of the extensions and modifications of the basic model.

These extensions are not meant only as a robustness check though. As mentioned in the introduction, our goal is a realistic description of cooperation when people are strategic and the structure of the games they play varies. In the face of the variety of the games we play, players' mental processes should be viewed as the linchpin of cooperation. The extensions below are meant to capture the scope of a given mental process.

4.1. Sequential gift exchange

In our main example players moved simultaneously, choosing the effort levels in gifts for the other. Simultaneous choice is often an unrealistic assumption. If you and I are to cooperate in our research efforts, we send drafts of our papers to each other. You send me a paper that you have written, and I make comments on it. I then send a paper to you and you make comments. Our choices of effort – and consequently the signals received – are chosen sequentially rather than simultaneously.

Our assumption of simultaneous play is for pedagogical reasons, and allowing play to be sequential rather than simultaneous does not substantially alter our analysis. To fix ideas, assume that in each stage player 1 moves first, then player 2, with signal z occuring at the end of the stage as before. We again examine the case where p is close to 1, and assume that players are endowed with the same mental process as before. We check the conditions under which that same mental process continues to enable players to support cooperation in equilibrium.

As before, if I choose D in state N, you are likely to receive a bad signal, resulting in your being in state U, and triggering a punishment phase; if I choose C in state U, I avoid triggering a punishment phase in the event you are still in state N, however I incur a loss in the event you are in state U. Thus, this case is very similar to the simultaneous play case.

There are differences however: play is sequential, so when player 1 plays D, player 2 most likely receives a bad signal, and she may thus react immediately (i.e. within the same period) to player 1's defection. As a result, incentive conditions are altered (for player 1).[17]

[17] Incentives are not altered for player 2. Note that player 1 and 2 are not in a symmetric position because switching back to state N (after signal $z = 1$) may only occur after player 2 moves.

Indeed since player 2 reacts within the same period to player 1's defection,[18] player 1 gets 0 if he plays D at both N and U. So incentives to play C at N are trivially satisfied for player 1. In contrast, and precisely because player 1 does not gain as much as before from defecting, incentives to play D at U are more difficult to satisfy. Conditional on the state being (N, N), consider the events leading to player 1 being in state U. Either player 2 receives a bad signal (then switches to U and plays D, so that with probability p player 1 transits to U as well), or player 2 receives a good signal but subsequently player 1 receives a bad signal. Both these events have the same probability. Under the first event, player 1 loses $(-L)$ for the duration of the punishment phase if he plays C in U. In the second event, player 1 avoids triggering a punishment phase by playing C in U, and saves Δ (rather than $\Delta - L$). So player 1's incentive constraint becomes:

$$\frac{1}{2}(\Delta) + \frac{1}{2}(-L\Delta) < 0,$$

hence:

$$L > 1.$$

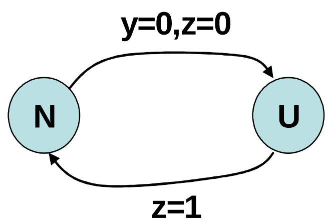
4.2. Heterogeneous stage games.

The previous subsection pointed out that our analysis is robust to some perturbations in how a repeated interaction is modeled. It is customary to model such relationships with a repetition of a simultaneous stage game, while in fact, the relationship may involve a sequence of transactions in which a single player has a choice. In general, we would like the insights of our analysis to be robust to changes in the fine details of how we model the phenomenon in question. The standard repeated game model abstracts from details in ways that can be important other than the timing issue. The standard model assumes that a given stage game is played repeatedly. Specifically, it is assumed that the payoffs in play are identical and, if there is imperfect monitoring, the monitoring structure is identical as well. This is a charicature of the typical relationship that we want to understand. Cooperation between two people living together is a prototypical relationship that we might wish to understand. One person may have prepared dinner when the other was not feeling well yesterday, the other may do the laundry today while the other watches a favorite television program, and the first may grocery shop tomorrow while the second sleeps late. The payoffs in each of the transactions can differ, as may the details of the monitoring structure. Rather than a repeated game, there is a sequence of transactions the two face, and it is implausible that the payoffs and the monitoring structure for all games in the sequence are identical. We will outline how our approach can be extended to such general settings. One of our central points is that our basic mental system, properly extended, may allow for cooperative behavior under broad circumstances.

Consider two people who are faced with a general finite action game that they play repeatedly and who discount the payoffs they receive. The interesting case is that in which there is a pair of actions, s_1^*, s_2^*, for the stage game that, if played, give each player a higher payoff than in any pure strategy Nash equilibrium in the stage game. The players would then prefer to cooperate and play action pair s_1^*, s_2^* to play a stage game Nash equilibrium. Suppose, however, that there

[18] This is because p is assumed to be close to 1.

is a monitoring problem as in our examples above, making cooperation difficult; in particular, suppose that, as before, players get a signal of their partner's action that is highly accurate, but not perfect. Suppose there is a *norm* that specifies how each player is to act in the stage game, that is, that specifies a pair of actions s_1^* and s_2^* that are to be played in the stage game. One can then map this problem into the example that we analyzed above as follows. The mental system is as before:



Transition function as before

Player i's strategy is to play s_i^* in state N and $\bar{s}_i$ in state U, where $(\bar{s}_1, \bar{s}_2)$ is a Nash equilibrium of the stage game. The signal y now corresponds to a player receiving a signal about whether or not his partner has played the action prescribed by the norm, and as before, z is a public signal through which the players may recoordinate to cooperation. Signal $y_i = 0$ is a signal to i that j has not played s_j^* as prescribed by the norm, while $y_i = 1$ is a signal that he has. For the case in which the payoffs from cooperating by playing (s_1^*, s_2^*) Pareto dominate playing the Nash equilibrium $(\bar{s}_1, \bar{s}_2)$, the structure of the problem is essentially the same as our leading example. Whether cooperation, in the form of following a specific norm, will be an equilibrium will depend on the specific payoffs in the game at hand. Player i will consider playing something other than s_1^* in state N and playing something other than $\bar{s}_i$ in state U, and whether the play prescribed by the norm is an equilibrium will depend on the magnitude of the gain from deviating when the opponent is cooperating and the magnitude of the loss when playing the Nash equilibrium $(\bar{s}_1, \bar{s}_2)$ relative to cooperation.

One can extend this way of modeling cooperation when there is a given game that is repeatedly played to a pair that repeatedly interacts in a variety of games. Suppose there is a finite set Ξ of games that may be played, and denote by h the stochastic process over games. For each game $\xi \in \Xi$ there is a set of actions A_i^ξ for each player i, a payoff function g^ξ, and a monitoring structure (Y^ξ, q^ξ).[19] When player i is in game ξ there is a set of "bad" signals $Y_{i,0}^\xi$ that will cause i to transit from state N to state U (unless $z = 1$).

A generalized *norm of behavior* prescribes a way to play in each game that arises. A strategy σ_i for player i then specifies a rule $a_i^\xi \in \{D_i^\xi, C_i^\xi\}$ for each possible state $s_i \in \{U, N\}$; C_i^ξ corresponds to playing the prescribed cooperative action for game ξ, and D_i^ξ corresponds to i's playing his part of the stage game Nash equilibrium of the game ξ. From the modeller's perspective, the dynamic system is at any date in some state (s, ξ), where $s = (s_1, s_2)$, and (σ, h) generates transitions over these states. We denote by $\phi_{\sigma,h}(s, \xi)$ the ergodic distribution over states induced by (σ, h). As before, we define the value associated to a strategy profile σ

[19]For the sake of illustration, we assume below that each game ξ has a unique Nash equilibrium.

as:

$$v(\sigma) = \sum_{s,\xi} g^{\xi}(\sigma(s))\phi_{\sigma,h}(s,\xi).$$

Equilibrium conditions can then be defined as before.

The above discussion takes transitions to be exogenous. A natural extension would be to define transitions in such a general setting so that a transition to the upset state occurs only if the likelihood ratio of cooperation versus defection is sufficiently small (given the signal). Formally, one could set a threshold β that determines which signals are bad:

$$Y_{i,0}^{\xi} = \{y_i, \frac{\Pr\{y_i \mid C_i^{\xi}, D_j^{\xi})}{\Pr\{y_i \mid C_i^{\xi}, C_j^{\xi})} \geq \beta\}$$

Thus, if a player is currently facing an interaction in which the signal about the partner's action is not very informative, an otherwise negative signal might not move him to the upset state.

4.3. Heterogeneous agents

We assumed that agents were identical, that is, that the cost and benefit of favors was the same and their "monitoring technologies" were the same. As is the standard assumption that the games in each period are the same, the assumption that the players are identical is unrealistic. It is clear that the basic qualitative analysis remains unchanged if players have different costs and benefits of favors, provided that for each player, the cost and benefit fall within the limits described in the main example. Of course, large differences in costs and benefits across players will be problematic, as the limits described would not be satisfied for both players.

Differences in the players' signal technologies can also be problematic. Suppose, for example that player 1 receives a moderately accurate signal about 2's action, while 2 receives an almost perfect signal about 1's action. The strategies in the example that supported cooperation will not be equilibrium strategies with this change in signal technology. In the example when the signal accuracies of the two players are the same, player 1 had an incentive to play D in U because when player 1 received a bad signal, the following two events are equally likely. Event 1: Player 2 was in state N, had played C and 1 received an incorrect signal; Event 2: Player 2 had previously received an incorrect signal and was in state U and played D. So when player 1 chooses C in state U, half of the time, he avoids triggering a punishment phase, but half of the time, he bears the cost of remaining cooperative against an upset player. But with the alterred signal technologies in which player 2 receives an almost perfect signal, player 1 switches to U most likely because he (player 1) has received an incorrect 1 signal, so by choosing C in state U, player 1 mostly avoids trigerring a punishment phase. Player 1's best response is thus to play C rather than D in state U, hence the strategies are not an equilibrium when player 2 gets an almost perfect signal.

It isn't necessary that player 2 receive an almost perfect signal for the strategies in the example to fail to be equilibria. If player 2's signal is significantly more accurate than player 1's signal, player 1 may still find that by playing C in U, he is much more likely to avoid trigerring punishment phases rather than bearing the cost of being cooperative against an Upset player 2. The accuracies of the players' signals do not have to be exactly the same for the strategies described in the example to be equilibrium, but there is a limit on how different they can be.

4.4. Many players

Many of the insights of the two-person gift exchange problem carry over to larger groups. Suppose that there are K players, where K is even, and in each period, half of the population is randomly matched with the other half. As in the two-person case analyzed above, we assume that players may either be in state U or N, switching to state U after a bad signal, and switching back to state N after the realization $z = 1$. We examine the conditions under which our candidate strategy profile (cooperate at N and defect at U) is an equilibrium.

Under our candidate strategy profile, players will alternate between cooperation phases (in which all players are in state N), and punishment phases (in which some players are in state U). The probability of switching from cooperation to a punishment phase is $\pi = (1-q)(1-p^K)$ (since switching occurs when one player receives a bad signal and $z = 0$). The probability of switching from punishment to cooperation is q as before.

Given K, if p is sufficiently close to 1, then as before, the cooperation phase will be much longer than punishment phase (in expectation). There are two main differences with the previous case, though. First, when a punishment phase starts, it takes some time before all players switch to state U. Hence a player who plays D may continue to meet many players in state N. This makes the incentives to play C at N weaker. Second, when a player in state N gets a bad signal, that player understands that there is only a $1/K$ chance that he was the first player to get a bad signal. It is only in the case that he was the first player to get a bad signal that playing C averts the punishment phase, hence the incentive constraint to play D at U will easier to satisfy.

Incentive to play C at N : If player i deviates to playing D at both states, this will propagate through future random matches to the whole population. The length of the punishment phase is random: until the public signal $z = 1$. If a punishment phase lasts t periods,[20] call Q_t the expected number of "uninfected" players (that is, those who have not yet seen a bad signal) that player i will meet during that punishment phase, and define

$$Q = \sum_t Q_t(1-q)^t q.$$

Q corresponds to the average number of uninfected players that player i meets in a punishment phase, taking into account the fact that the length of the punishment phase is random. The constraint for a player to have an incentive not to play D in state N is then

$$(1+L)(1+Q) < \sum_t t(1-q)^t q = \frac{1}{q},$$

or equivalently,

$$\Delta > L + Q(1+L),$$

hence

$$L < \frac{\Delta - Q}{(1+Q)}.$$

To compare with the previous 2 player case, note that there, for p close to 1, we had $Q_t \approx 0$ for all t (hence $Q \approx 0$) because when player i defected, the other player would immediately

[20] This event has probability $(1-q)^t q$.

switch to state U with probability close to 1. Here, it takes some time before players switch to U, hence $Q > 0$. Nevertheless, there is a bound on the time it takes for bad signals to propagate through the whole population. Because bad signals will propagate exponentially in the population, $Q_t \le \min(t, \bar{Q})$ where $\bar{Q}$ is of the order $\log K$.

There is also the constraint that player i should play D in state U. When player i plays C at both N and U, he avoids triggering some punishment phases. Offsetting this, however, he remains cooperative in punishment phases. Conditional on both players being in state N, consider the events where only one player receives a bad signal. In the event that player i receives the bad signal, player i avoids triggering a false alarm (and saves $\Delta - L - Q(1+L)$). However, in the event that some player $j \neq i$ received the bad signal, a punishment phase starts. Let π denote the probability that i becomes "infected" before a signal $z = 1$ arises (if Δ is large compared to Q, then π is close to 1). In that event, player i loses L in each period of the punishment phase (i.e. until a signal $z = 1$ occurs).[21] Thus it will be optimal for a player to play D after first seeing a bad signal if

$$\frac{1}{K}(\Delta - L - Q(1+L))) + (1 - \frac{1}{K})\pi(-\Delta L) < 0,$$

or equivalently[22]

$$L > \frac{\Delta - Q}{1 + Q + (K-1)\pi\Delta}$$

We see from these inequalities how the feasibility of cooperation in this society changes as the group gets large. The constraint that a player should play C when in state N becomes harder to satisfy (holding p and q fixed). (Recall that Q is on the order of $\log K$, and hence is also becoming large when K increases.) This is intuitive; the larger the group, the longer will be the expected time that I will continue to match with uninfected players who play C when matched with me. The other inequality however, that a player should play D when in state U, becomes easier to satisfy. This is also intuitive: it is less likely that a player in a large group who receives a bad signal is the first to do so.

5. When z is not public: Resetting the relationship to cooperation

In many situations, there may be no public signal that can facilitate the coordination back to cooperation. How can players then coordinate a move back to cooperation? We examine in turn the case where z is almost public, and next the case where each player receive signals z_i that are independently distributed. **We then show how recoordination can occur despite the absence of public signal.**

[21] We omit here the fact that by playing C, a player slows down infection and consequently may face "uninfected" players for a longer period of time. However this term is negligible when K is large: it is of the order of at most $(\log K)/K$. Intuitively, the effect is smaller than the effect of randomly switching one player each period from state U to state N. In that case, infection would still spread to the whole population, but it would take slightly longer, and be comparable to $\log(K + \log K)$ rather than $\log K$.

[22] Note that $\frac{1}{1+(K-1)\pi}$ corresponds to the probability that player i is the first to switch to U, given that he switches to U.

5.1. "Almost public" signal z

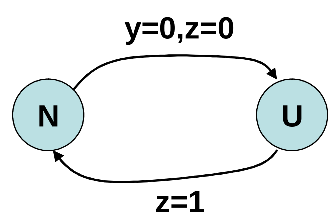
Because of the noisiness in the private signals the players get, the relationship will periodically fall into disrepair with neither exerting effort in choosing a gift. Without some way to "reset" the relationship this would be an absorbing state with no further cooperation. The public signal z allows the players to recoordinate and start cooperation afresh. The limits on the probability of the public (or nearly public) signal are intuitive: if q is too low, players may prefer not to punish when they get a signal that the other has not put in effort, and if q is too high the punishment phase will not be sufficiently painful to deter deviation.[23]

It isn't essential, however, that the signal be precisely public. If the parameters (p, q, L) are such that incentives are strict, then by continuity, incentives will continue to be satisfied if each player receives a private signal $z_i \in Z_i = \{0, 1\}$ with the property that $\Pr(z_i = 1) = q$ and $\Pr(z_1 = z_2)$ close enough to 1. As we show below however, while signals can be private, there are cases where correlation across signals. cannot be too weak.

5.2. Independent signals z_1, z_2.

We explore below the case of independent signals, and first show that when p is close to 1, cooperation can not be sustained in equilibrium.

Formally, we consider the mental process as before with the qualification that T_i is now defined over Z_i rather than Z. We investigate whether and when the strategy σ that plays C in N and D in U is an equilibrium.



Proposition: *Fix the mental system as above. For any fixed $q \in (0, 1)$, for p close enough to 1, the strategy profile where each player cooperates in N and defects in U cannot be an equilibrium.*

Assume p is close to 1. Under σ, cooperation phases last $\frac{1}{2(1-p)(1-q)}$ on average (because each player has a chance $(1 - p)(1 - q)$ of switching to U in each period), and punishment phases last $\frac{1}{q^2}$ (since only in events where both players get signal $z = 1$ at the same date that recoordination on cooperation is possible).[24] So for p close enough to 1, the value to following the proposed strategy profile is 1.

Compared to the case where z is public, the incentives to play C at N are unchanged: if player 1 plays D at both states, his opponent will be cooperative once every $1/q$ periods on average, hence the condition

$$L < 1/q - 1 \tag{5.1}$$

[23] A richer mental system can allow for cooperation even when the probability of the public signal is too large or too small for the two state mental system in the example. We discuss this below.

[24] Omitting terms comparable to $(1 - p)$, this is true whether the current state profile is (U, U), (U, N) or (N, U).

still applies.

Incentives to defect at U however are much harder to provide. As before, by cooperating at U, player 1 ensures that a punishment phase is not triggered in the event state profile is UN. But there is another beneficial effect. In the event state profile NU occurs, the punishment phase that follows will last only $1/q$ periods (as simultaneous transition to N is no longer required). So player 1 will only have incentives to defect at U when:

$$\frac{1}{2}L(1/q) > 1/q^2,$$

or equivalently

$$L > 2/q,$$

a condition which is incompatible with inequality (5.1).

Thus strategy σ cannot be an equilibrium: the length of punishment phases is subtantially reduced when playing C at U, which makes playing C at U an attractive option.

TO BE DONE: show that results holds for p not close to 1 as well.

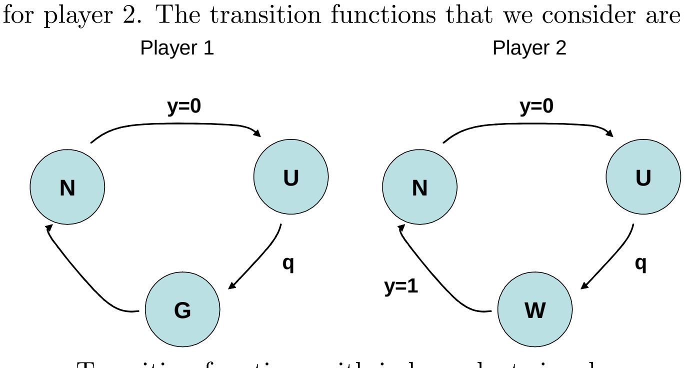
5.2.1. No public signal

In many situations, there may be no public signal that can facilitate the coordination back to cooperation. How can players then coordinate a move back to cooperation? One possibility is that rather than trying to have both players simultaneously switch back to cooperation, one player can be designated a leader in the relationship, and make a gift (i.e. cooperate) as a signal that he understands that the relationship has broken down and needs to be restarted. Such a possibility cannot be achieved under the simple mental process we have examined so far. However, a more sophisticated one can achieve that possibility, as we illustrate below.

We will consider again the case that p is close to 1. We assume that each player receives a private signal $z_i \in \{0,1\}$ in each period, with the property that the signals are independently distributed and $\Pr(z_i = 1) = q$.

States and transitions.

We assume that each player i may be in one of three states: U, N, and G for player 1, and U, N and W for player 2. The transition functions that we consider are shown below.



Transition functions with independent signals z_i

These transitions specify that when in state N, a bad signal always triggers a switch to state U. Once in state U, players 1 and 2 independently transit respectively to state G or W (with probability q in each period). Once in state G, player 1 always switches back to state N, while, once in state W, player 2 switches back to state N only if he receives a good signal $y_2 = 1$.

Strategies:

As before, we consider strategies that condition only a player's state.

$$\Sigma_1 = \{\sigma_1, \sigma_1 : S_1 \longrightarrow A_1\} \text{ and } \Sigma_2 = \{\sigma_2, \sigma_2 : S_2 \longrightarrow A_2\}$$

Our candidate equilibrium strategy pair will be as follows. For player 1,

$$\sigma_1(N) = C, \sigma_1(U) = D \text{ and } \sigma_1(G) = C$$

and for player 2,

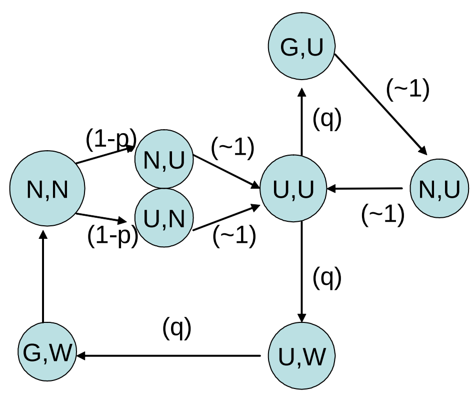
$$\sigma_2(N) = C, \sigma_2(U) = D, \sigma_2(W) = D$$

Intuitively, when the state profile is (N, N), both players cooperate and each player may receive a bad signal and trigger a punishment phase. Once a punishment phase starts, two events may occur:

Either player 2 moves to state W before player 1 moves to G (or at least no later than one period after player 2 does). In that case, the most likely event is that players will coodinate back to (N, N) (with probability close to 1).[25] Alternatively, player 1 moves to state G more than one period before player 2 moves to state W. In that case, the most likely event is that players switch to (N, W) or (N, U), hence coordination back to (N, N) will take longer.

We illustrate the main transitions for the case where p is close to 1 and q is small, but not too small:

$$0 < 1 - p \ll q \ll 1.$$



Transition of mental state pairs

Throughout the rest of this section, we will focus on that case. We analyse the more gereral case where q is larger in the appendix.

Analysis:

[25] This is because once in state profile (G, W), player 1 plays C and moves to N, while player 2 receives (with probability close to 1) signal $y_2 = 1$, hence also moves to N.

When players follow the proposed strategy profile, players alternate between long phases of cooperation (of lenth $1/\pi$ with $\pi = 2(1-p)$), and relatively short punishment phases (of approximate lenght $2/q$).

Incentives of player 1 at U. Under the proposed equilibrium strategy profile, the expected loss that player 1 incurs (compared to being in the cooperative phase) until coordination back to cooperation occurs is approximately $2/q$.[26]

When player 1 cooperates at U, he avoids triggering a punishment phase in the event (U, N), so the occurences of punishment phases are reduced by 1/2. Besides, punishment phases are shorter, as coordination back to cooperation occurs as soon as player 2 transits to W (hence punishment lenght is reduced to $1/q$), however they are more costly per period of punishment, as player 1 looses an additional L in each period. The condition is thus:

$$\frac{2}{q} < \frac{1}{2}(\frac{1}{q})(1+L)$$

or equivalently:

$$L > 3.$$

Incentives of player 2 at N: Defection generates short period of cooperation (cooperation lasts 2 periods), during which player 2 gains an additional payoff of L, and long periods of punishment (that last $1/q$ periods) during which player 2 looses 1. Hence the condition

$$2L < \frac{1}{q}$$

Other incentives are easier to check and automatically satisfied, so they are omitted.

So as before, we obtain two bounds on L. The following figure describes, for p close to 1, the set of values of L and q for which the proposed strategy profile is an equilibrium.

[26] The exact cost is larger because there are few periods at which player 1 cooperates while player 2 still defects. A better approximation is that that cost is $2/q + \frac{1}{2} + 3(1+L)$. To see why, compute, conditional on being in (U, U), the expected loss that player 1 incurs (compared to being in the cooperative phase) until coordination back to cooperation occurs. We denote by Δ that loss. We have:

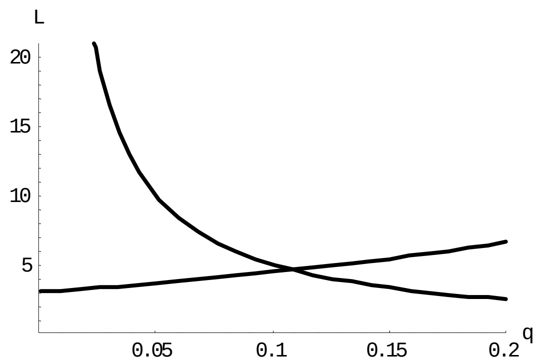
$$\Delta = 1 + q[2(1+L) + \Delta] + q[1/q + 1 + L] + (1-2q)\Delta$$

or equivalently:

$$\Delta = \frac{2}{q} + 3(1+L)$$

Because there is equal chance of going to UU through NU or UN, the expected loss from a punishment phase is:

$$\frac{1}{2}(1+L) + \frac{1}{2}(-L) + \Delta = \frac{1}{2} + \Delta$$



Values of p and L for which cooperation is possible

6. Further discussion

6.1. Extensions of the model

1. We take transitions to be exogenous but they need not be independent of the game at hand. If a player is in a repeated relationship in which the signal about the partner's action is poorly informative, a bad signal might not move him to the upset state. It may be that transition to the upset state occurs only if the likelihood ratio of cooperation versus defection is sufficiently small (given the signal). More generally, if there are many signals that one might receive, only sufficiently informative signals might trigger a transition to a different state.

2. The set of possible mental states, in general, are driven by emotions and memory. If signals are public, and player i can observe that player j did not like the gift, player i might feel Guilty, an additional mental state. If transitions to G never occur and transitions are as before (hence never take the other's signal into account), nothing changes to our analysis: our equilibrium would remain an equilibrium. However, if transitions to the state G occur, then other outcomes may be sustainable as equilibria (equilibria similar to the public equilibrium for example).

More generally, our approach suggests thinking first of mental states, then of transitions that seem reasonable given the mental states, and ultimately of a class of games for which the corresponding mental system is useful.

3. There is no utility attached to mental states in our model; the states U and N are no more than collections of histories. It is straightforward to extend our model to the case in which utility is attached to states, or to particular sequence of states (going from upset to normal). Strategies that constitute a strict equilibrium in our model would remain an equilibrium for small values attached to utilities in particular states or movements among states, but large values might affect the possibilities of cooperation. Suppose that in our main example, there is a negative utility attached to being in state U (relative to being in state N). When player 1 is then in state N, deviating will be more costly than in the case we analyzed. Consequently, there will be a larger set of parameters for which players will not choose the strategy of playing D in both states. What if player 1 plays C in both states? It is easy to see that doing so will not decrease the proportion of time that 1 is in state U. Player 1 is playing C when the pair of is (N, N) or (N, U), so we need only check the pairs (U, N) and (U, U). But in both cases,

1 is already in state U and will remain in that state until $z = 1$. Hence, playing C will neither decrease the chance of transiting from state N to U, nor will it quicken his departure from U. Summarizing, modifying the example so that there is a decrease in utility when a player is in state N (or suffers a decrease in utility when he transits from N to U) can only increase the set of parameters for which the cooperative strategies are an equilibrium.

4. We have taken the mental system – the states and transition function – to be exogenously given. We did, however, suggest that one might think of these as having been formed by environmental factors. In the long run, evolution might influence both the set of mental states that are possible and the transition function. While beyond the scope of this paper, it would be interesting to understand how evolution shapes mental systems. We have pointed out above that it is *not* the case that evolution should necessarily favor more complicated mental systems; adding more states to a mental system that allows cooperation might make cooperation then impossible.

While evolution might shape the mental systems one would expect to see in the long run, cultural transmission would logically be a way mental systems are shaped in the short run. We have illustrated the notion of mental systems with our leading example in which there is a unique equilibrium, but it is easy to see that one might easily have multiple equilibria in such a model. Models such as this might be useful in understanding why behaviors that are innocuous in some cultures trigger emotional, and sometime violent, reactions.

6.2. Experiments

Our aim has been to set out a model that allows cooperation in ongoing relationships via behavior that can be represented by realistic strategies. Our goal is to provide a broader conceptual framework than standard repeated games permit. While our goal was primarily conceptual, the model may suggest experiments. If people are allowed to play a repeated game with private monitoring similar to the one we analyze, in principle one could examine the actions people play and see whether they can be reasonably be structured as our model suggests. That is, one can ask what histories trigger a deviation from cooperation, whether those histories can be classified as a mental state, and whether the same histories continue to structure behavior as parameters of the game vary.

There have been experiments that investigate how players behave in repeated prisoner's dilemma games with uncertainty of different kinds. In general, cooperation is less likely when actions are "noisy", that is, there is randomness in what players observe about what their opponents have done[27], and when the participants receive random payoffs. However, in these experiments, while players may not see exactly what their opponent has done, they know precisely what signals the opponent has seen. There has been relatively little work done on what we have argued is the realistic setting in which players do not observe perfectly the actions taken by their opponents. Miller (1996) has run tournaments in which strategies for playing repeated prisoner's dilemma games with uncertainty are pitted against each other, with more successful strategies "multiplying" more rapidly than less successful strategies. The form of the uncertainty that Miller considers is essentially as in our model. The action played in any period is seen with some noise by the opponent. Miller compares the strategies that succeed in this

[27] See, e.g., Bendor et al. 1999 and Barbara Sainty, 1999.

tournament when there is no noise in the observation of the opponent's action and when the signals are 99% accurate and 95% accurate. While there are differences in this particular exercise and our model,[28] it is interesting to note two things. First, the successful strategies that emerge are simpler than they might have been in the sense that they use fewer states than they might. In other words, there seems to be some "optimal" number of states in the sense that strategies that used more states did strictly worse. The second interesting thing about Miller's results is that when the error rate in the signals of what an opponent has done increases from 1% to 5%, the number of states used by the most successful strategies decreases. While we don't take this as compelling evidence for our general point of view, it is consistent with our framework.

6.3. Related literature

Although we have emphasized the difficulty in supporting cooperation when signals are private, there are monitoring structures for which cooperation is relatively easy to sustain. These are case where each player can be sure – or almost sure – of his opponent's state of mind. Mailath and Morris (2002) analyze repeated games in which players get imperfect information about past play. Consider games in which players get a public signal, that is, a signal that both see. Cooperation may be possible here, because if players choose pure strategies, they know exactly their opponents' state of mind: the public signals and the actions prescribed by the equilibrium strategies will be common knowledge. Mailath and Morris then consider a perturbation of this information structure whereby each player gets the public signal with a small amount of noise. They focus on the case that players' signals are *almost public*: for any signal a player receives, the probability that other players have received the same signal is close to one. Mailath and Morris show that if players' strategies depend only on a finite number of past signals, the introduction of the small amount of noise into the players' signals about past play doesn't matter; strategies that comprise an equilibrium when there is no discrepancy in what players observe remain equilibrium strategies when players' signals are almost public.

This result provides sufficient conditions under which cooperation can be possible even when players receive private signals, but those conditions are quite stringent. In particular, when signals are almost public, I can predict very accurately what other players will next do. This is in sharp contrast to our example. First, the signals that players get are not helpful to predict the signal received by the other player, and second, however accurate signals are there are times (in state U for example) when a player cannot predict with any accuracy what his opponent will do.

It has long been understood that some Nash equilibria are sufficiently complicated that it is implausible that players will be able to identify the strategies and play them. One approach to taking into account the complexity of strategies is to assume players use finite automata to implement their strategies. A finite automaton consists of a finite number of states and a transition function, as in our model. The complexity of a player's strategy is defined to be the minimal size of a machine that can implement that strategy.[29] We differ from this literature in several respects. First, the literature using automata to implement strategies has players choosing both the transition function and the mapping from states to actions, taking fixed only the number of states available given the automaton's size. In contrast, we take players' transition

[28] Miller considers finitely repeated games and allows transition functions to vary, for example.

[29] See, e.g., Ben-Porath (1986) and Abreu and Rubinstein (1988).

functions as fixed with players' choices being only the mapping from states to actions. Second, to our knowledge, this literature does not consider games with private monitoring. Third, the earlier literature used automata primarily as a tool to capture complexity, our modeling strategy takes more seriously mental systems as being a plausible, if crude, model of the process by which players may interact. There has been work in single-person decision making problems that is analogous to the papers using automata to capture complexity costs.[30] While we take agents' transition functions as fixed, the focus of this literature is on characterizing the optimal transistion rule.

In our model a player must choose an action at a given date that depends only on which of the finite number of states that he is in at that time. The number of realized histories goes to infinity so a state is an information set that over time contains a large set of histories, and a player might prefer to choose different actions at different histories that led to a given state if he could distinguish the histories. This structure is analogous to the "absent-minded driver problem"[31] in which a driver who wishes to exit a limited access highway "at the second exit" must decide what to do when he arrives at an exit but cannot recall whether he has already passed an exit.

The model we study reduces to a stochastic game of a particular kind in which each player has his own state variable, and each player observes only his own state.[32] Most of the literature in stochastic games assumes that in each period, there is a state that is known to both players,[33] while our interest is in the case that players do *not* know their partner's state. In addition, our interest is primarily in understanding the conditions under which cooperation is possible for specific classes of games.

Kandori (1992) and Okuno-Fujiwara and Postlewaite (1995) analyze the possibility of cooperation when players are repeated randomly matched to play a prisoner's dilemma game with limited information about previous play. Those papers demonstrate that cooperation can be supported with limited information, but the nature of the limited information is different from that in this paper. In particular, there is no uncertainty about the action that an individual has taken in a given interaction unlike in our model.

Abdulkadiroglu and Bagwell (2007) analyze a two-person favor exchange model in which in each period one or the other, or neither, of the agents may be able to do a costly favor for the other agent.[34] When in a given period an agent has the opportunity to do a favor, this fact will be known to him but not to the potential recipient of the favor. If a favor is done, it is publicly observed by both agents. It is assumed that the benefit to the recipient of a favor is greater than the cost of doing the favor, it is efficient that all possible favors be done. It is straightforward to see that there is an incentive problem in inducing an agent to do a costly favor when his opponent will not know that such a favor was possible however. Abdulkadiroglu and Bagwell analyze the equilibria in this problem and demonstrate how relatively simple strategies can support significant cooperation. This problem differs from our problem in the nature of the asymmetry of information. In the problem Abdulkadiroglu and Bagwell analyze, agents do not

[30] See Wilson (2004) and Cover and Hellman (1970) for such models of single-person decision problems and Monte (2007) for a strategic treatment of such models.

[31] See Piccione and Rubinstein (1997).

[32] We thank Eilon Solan for this observation.

[33] However, see Altman et. al. (2005).

[34] Mobius (2001) and Hauser and Hopenhayn (2005) analyze similar models.

know precisely the game they face in any period. If an agent cannot do a favor for his opponent, he does not know whether or not his opponent can do a favor for him. But when an agent chooses to do a favor, both agents observe this with no uncertainty. In contrast, in our model agents know precisely the game in any period, but get an imperfect signal of the action taken by their opponent. It is this latter imperfection that makes coordination and cooperation in our framework so difficult.

7. Bibliography

Abdulkadiroglu, Atila and Kyle Bagwell [2007], "Trust, Reciprocity and Favors in Cooperative Relationships," mimeo, Columbia University.

Abreu, D. and A. Rubinstein [1988], "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, No. 6, 1259-1281.

Altman, E., K. Avrachenkov, R. Marquez and G. Miller [2005], "Zero-Sum Constrained Stochastic Games with Independent State Processes," *Mathematical Methods of Operations Research* 62, 375-386.

Bendor, Jonathan, Roderick M. Kramer, and Suzanne Stout [1991], "When in Doubt ...Cooperation in a Noisy Prisoner's Dilemma," *Journal of Conflict Resolution* 35, pp 691-719.

Ben-Porath, E. [1986], "Repeated Games with Finite Automata," *Journal of Economic Theory*, 59, 17-32.

Ben-Porath, E. and M. Kahneman [1996], "Communication in repeated games with private monitoring," *Journal of Economic Theory* 70, 281–297.

Bereby-Meyer, Yoella, and Alvin E. Roth [2006], "The Speed of Learning in Noisy Games: Partial Reinforcement

and the Sustainability of Cooperation," *American Economic Review 96,* 1029-1042.

Cohen, D., R. E. Nisbett, B. Bowdle and N. Schwartz [1996], "Insult, Aggression, and the Southern Culture of Honor: An 'Experimental Ethnography'," *Journal of Personality and Social Psychology* 70, 945-960.

Compte, O.,

Cover, T. and M. Hellman [1970], "Learning with Finite Memory," *Annals of Mathematical Statistics* 41, 765-782.

Ellison, G., "Learning, Local Interaction, and Coordination," *Econometrica*, Vol. 61, No. 5 (Sep., 1993), pp. 1047-1071.

Ely, J. C. and J. Välimäki [2002], A robust folk theorem for the prisoner's dilemma, *J. Econ. Theory* 102, 84–105.

Hauser, C. and H. Hopenhayn [2005], "Trading Favors: Optimal Exchange and Forgiveness," mimeo, UCLA.

Kandori, M., "Introduction to Repeated Games with Private Monitoring, *Journal of Economic Theory,* 2002, Vol 102, pp 1-15.

Kandori, M., "Social Norms and Community Enforcement," *Review of Economic Studies*, 1992, Vol 59, pp. 63-80.

Kandori, M., and I. Obara, "Finite State Equilibria in Dynamic Games," mimeo, 2007.

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

Miller, John H. [1996], “The Coevolution of Automata in the Repeated Prisoner’s Dilemma,” *Journal of Economic Behavior and Organization* 29, pp 87-112.

Mobius, M. [2001], "Trading Favors," mimeo, Harvard University.

Monte, D., “Reputation and Bounded Memory: A Theory of Inertia and Skepticism,” mimeo, 2007.

Okuno-Fujiwara, M. and A. Postlewaite, "Social Norms in Matching Games," *Games and Economic Behavior*, 1995, Vol. 9, pp 79-109.

Phelan, C. and A. Skrzypacz, "Private Monitoring with Infinite Histories," mimeo, 2006.

Piccione, M., The repeated prisoner’s dilemma with imperfect private monitoring, *Journal of Economic Theory* 102 (2002), 70–83.

Piccione, M. and A. Rubinstein [1997], "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior* 20, 3-24.

Rubinstein, A. [1986], "Finite Automata Play the Repeated Prisoner’s Dilemma," *Journal of Economic Theory* 39, 83-96.

Sainty, Barbara. “Achieving Greater Cooperation in a Noisy Prisoner’s Dilemma: An Experimental Investigation.” *Journal of Economic Behavior and Organization*, 1999, 39(4), pp. 421–35.

Wilson, A. [2004], "Bounded Memory and Biases in Information Processing," mimeo, University of Chicago.