Repeated Games and Limited Information Processing[*]
Preliminary and Incomplete

Olivier compte   Andrew Postlewaite

PSE, Paris   University of Pennsylvania

July, 2007; this version February, 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

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.

3

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 an interpretation 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. 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 is as below.

$$
\begin{aligned}
T_i^{-1}(N) &= \{(N, X, 1, 0), (s, X, y_i, 1)\} \\
T_i^{-1}(U) &= \{(N, X, 0, 0), (U, X, y_i, 0)\}.
\end{aligned}
$$

This table shows which combinations of state, actions and signals result in the two states. Player i will find himself in state N if he was previously in state N, received a good signal about the gift he received and the public signal was $z = 0$. He will also find himself in state N if the public signal was $z = 1$, regardless of his previous state, choice of action or private signal. When the public signal was $z = 0$, he will find himself in the other state, U, if he was in state N and received a bad private signal, or if he was in state U in the previous period.[5]

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.

---

[5] 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.

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:

$$
\begin{aligned}
T_i^{-1}(N) &= \{(N, X, 1, z)\} \\
T_i^{-1}(U) &= \{(N, X, 0, z), (U, X, y_i, z)\},
\end{aligned}
$$

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).

2.2. 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 s as a pair of signals (s_1, s_2) the players receive. Each strategy profile σ induces transition probabilities over states. By assumption each state s induces an action profile $\sigma(s)$, which in turn generates a probability distribution over signals, hence 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, $\phi_\sigma(s)$ denotes the (long run) probability that players are in state s.

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.[6] 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.

---

[6]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.

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 a mappings from S_i to A_i.[7] Also note that σ_i as defined should not be viewed as a strategy of the repeated game.[8]

We consider next the ergodic distribution induced by the strategy described above for p close to 1 and examine in turn possible deviations from that strategy.

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.[9]

When player i plays D at both N and U, player j continues to alternate between phases of cooperation and defection. 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

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

When player i plays C at both N and U, he avoids triggering some punishment phases. 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.[10] 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

---

[7]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.

[8]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.

[9]This is because $1/q = 1 + \sum_T Tq(1-q)^T$.

[10]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.
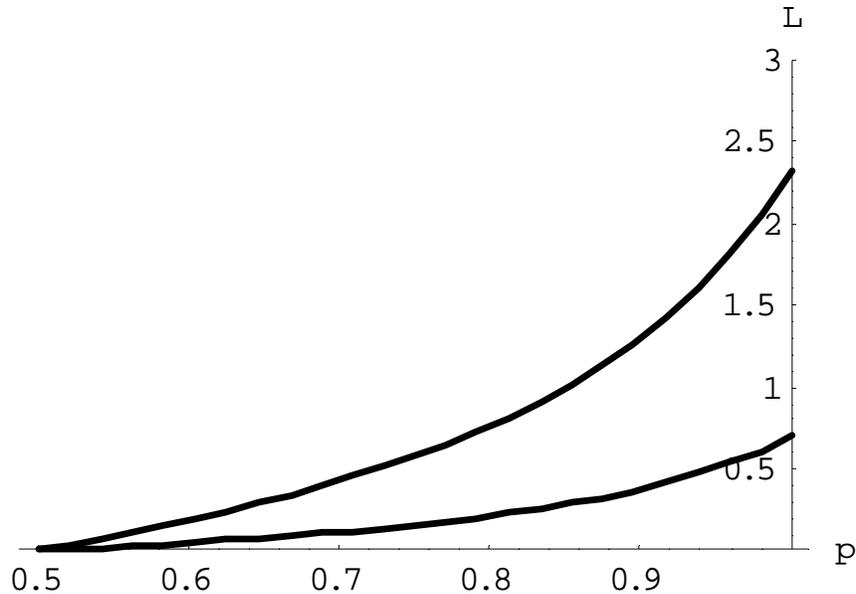
Figure 2.1:

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,[11] 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

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

$$L \;\; > \;\; \frac{\Delta}{1 + \Delta}.$$

To summarize, for p close to 1, there is a range of parameters (q, L) for which the proposed strategy is an equilibrium strategy.[12] This range of parameters is such that $\frac{\Delta}{1+\Delta} < L < \Delta$, 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:

---

[11] 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.

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

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 simply modifying 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. We might thus allow them to 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. Starting from our equilibrium strategy profile, it is actually difficult to modify it so that the modified strategies would constitute equilibrium when strategies are unconstrained. Equilibrium play in this first period following a return to state N would 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 counted differently than a signal received in later periods, in a way that takes into account the fact that the opponent, as well as the player, reacts differently to signals received shortly after a return to a cooperative phase than to signals received later.

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: each distinct belief can be interpreted as a distinct state of mind, but the number of distinct beliefs typically grows over time.[13]

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.

---

[13] See however Phelan and Skrzypacz (2006) and Kandori and Obara (2007).

Of course, for a given profile σ, one may still define the belief that a player holds at a given informational state s_i: it would correspond 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). However, the belief of player i does not determine which state he is in. In general, player i's belief at state s_i may depend on the strategy profile σ.

For example, at state $s_i = U$, the belief that player i would hold under the strategy σ that plays C in N and D in U is:

$$\Pr_\sigma(s_j = U \mid s_i = U) = \frac{1}{2}1 + \frac{1}{2}(1 - q) = 1 - \frac{q}{2}.$$

Indeed, players have equal chances of triggering a punishment phase. When player j triggers, player j is in state U whenever player i is, so:

$$\Pr_\sigma(s_j = U \mid s_i = U, \, j \text{ triggers}) = 1$$

When player i triggers, player j is still in N in the first period, and then switches to U with probability close to 1. So

$$\Pr_\sigma(s_j = U \mid s_i = U, \, i \text{ triggers}) = (0 + (1 - q)(1/q))/(1/q) = 1 - q.$$

Now, when player 1 plays the strategy σ^C that cooperates at both N and U rather than σ, then:

$$\Pr_{\sigma^C, \sigma_2}(s_2 = U \mid s_1 = U) = 1/2.$$

so the belief that player 1 would hold in U is not the same when he plays σ^C than that he would hold in U when he plays σ.

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 above. First, the signals that players get are not helpful to predict the signal received by the other player, and second, as our computation of beliefs at state U illustrates, however accurate signals are there are times when a player cannot predict with any accuracy what his opponent will do.

3.2. Another two-state mental process

Players in our example have two possible mental states, and transition functions according to which only signal $z = 1$ can move them back to the normal state. It is instructive to consider an alternative mental process in which players would move from state U back to cooperation (state N) after seeing a good signal: you stop being upset as soon as you receive a nice gift.

Formally, we still assume that players may be either in state N or in state U, but we now consider the following transition functions:

$$T_i^{-1}(N) = (s_i, X_i, 1, z)$$
$$T_i^{-1}(U) = (s_i, X_i, 0, z)$$

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. With only two states, I either choose D in state N as in the proposed strategies in the example, or I choose C. Choosing C in both states N and U, however, cannot be an equilibrium since the best response to this is to always play D. 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:

$$
\begin{aligned}
T_i^{-1}(N) &= \{(N, X, 1, 0), (M, C, 1, 0), (s, X, y_i, 1)\} \\
T_i^{-1}(M) &= \{(N, X, 0, 0)\} \\
T_i^{-1}(U) &= \{(M, C, 0, 0)), (M, D, y, 0))(U, X, y, 0)\}.
\end{aligned}
$$

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 example. If a player in state M gets a bad signal he transits to state U, and if he gets a positive signal 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., in case 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.
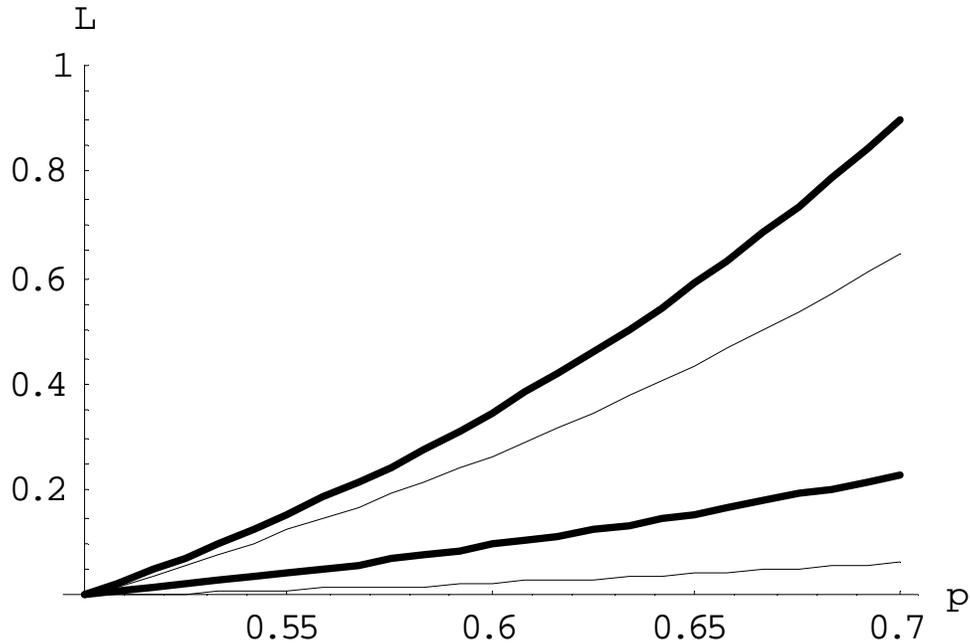
Figure 3.1:

The following figure illustrates this. We have set the value q to 0.1. For the set of parameters (p, L) between the thick lines, (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.

We make several comments about this example.

- Intuitively, the additional state allows for cooperation by reducing the probability that the relationship transits to a punishment phase following bad signals even when players play C, which is inevitable. Allowing for a stochastic transition function is an alternative modification of the mental system in our leading example that could accomplish this as well. 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 increase the expected length of the cooperative phase, making it more valuable. 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

13

state allows not only for more complex strategies to effect 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, and yet $(\bar{\sigma}, \bar{\sigma})$ is not an equilibrium. To see why, consider the case where p is close to 1. Under $(\bar{\sigma}, \bar{\sigma})$, punishment phases are triggered under two events: either when players are in state (M, N) (in which case player 1 triggers), or when players are in state (N, M), in which event player 2 triggers. When player 1 plays according to η instead (while player 2 continues to play according to $\bar{\sigma}$), player 1 plays C at M, hence he avoids triggering punishments. However player 2 continues to trigger, and player 1 loses $(1 + L)$ in the first period of the punishment (after that, he gets a second bad signal and moves to U). Both events have equal probability. Under the first event, player 1 gains:

$$\frac{1}{q} * 1 - (1 + L + (\frac{1}{q} - 1) * 0) = \frac{1}{q} - 1 - L$$

Under the second event however, he loses L. So player 1 does not have an incentive to play C at M when:

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

or equivalently when:

$$L \geq 1/2 \cdot (\frac{1}{q} - 1)$$

When $q < 1/2$, this condition is stronger than $L \geq 1 - q$.

3.4. 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.

Nevertheless, despite these restrictions, one might wonder whether an equivalent of the one-shot deviation principle holds in our framework. For repeated games with no restrictions on strategies the one-shot deviation principle says the following: to determine whether a pair of strategies is an equilibrium it is enough to check whether for any history a player wants to deviate from the proposed strategy in that period only, returning to the equilibrium strategy thereafter.

It might seem 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 should it not be the case that it is not profitable for me to deviate each time I am in state U? Actually, no; returning to our main example, it may be that deviating once at U is not profitable, while deviating each time I am in state U is profitable. The reason is that by deviating one time, you just postpone one period your triggering the punishment phase, while by deviating each time, you avoid to trigger punishment phases in events where only you have received 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.

4. The scope of a mental process

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 losing 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 vary. 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. Compared to the simultaneous play case, propects are thus very similar.

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, incentives conditions are altered (for player 1).[14]

Indeed since player 2 reacts within the same period to player 1's defection,[15] 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

---

[14]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.

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

than the timing issue. The standard model assumes that there is 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.

One of our central points is that within our mental system, the strategies in the main example are equilibrium strategies for a (relatively) broad set of parameters. The suggestion that the two players play C when in state N and play D in state U should remain optimal even as the stage game itself varies from period to period.

We explain more formally below how valuation functions can be extended to encompass those cases. Let ξ denote the payoff and signal characteristics of the game being played at a particular date: $\xi = (L, p, q)$. We have in mind that ξ would be drawn in each period from some finite set Ξ, according to some distribution h. As before, each player's mental process is characterized by mental states $S_i = \{N, U\}$ and a transition T_i over mental states:

$$T_i(s_i, X_i, y_i, z, \xi) \in S_i$$

>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 as:

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

Equilibrium conditions can then be defined as before.

Our central point is that for T as defined in the example, and properly extended, that is

$$
\begin{aligned}
T_i^{-1}(N) &= \{(N, X_i, 1, 0, \xi), (s, X_i, y_i, 1, \xi)\} \\
T_i^{-1}(U) &= \{(N, X_i, 0, 0, \xi), (U, X_i, y_i, 0, \xi)\}.
\end{aligned}
$$

the strategy σ that plays C in N and D in U is an equilibrium for many values of (p, q, L), as well as for many distributions h over Ξ.

Of course, we do not mean to suggest that T is always good. There are distributions for which T is a poor mental process. Consider for example the case of two possible realizations $\xi_0 = (L_0, p_0, q)$ and $\bar{\xi} = (\bar{L}, \bar{p}, q)$ with p_0 close to $1/2$ and L_0 close to 0. We choose $\bar{\xi}$ so that under T, if $\bar{\xi}$ were occuring with probability close to 1, cooperation would be sustainable. Assume now that the realization ξ_0 occurs frequently. Then transition to the upset state would become too frequent to sustain cooperation. This could happen despite the fact that there is little to gain from defecting under ξ_0. The problem is that players become Upset even though the signal

received in very poorly informative under ξ_0. The following transition function, which ignores signal y_i under the event ξ_0, would permit cooperation:

$$
\begin{aligned}
T_i^{-1}(N) &= \{(N, X_i, y, z, \xi_0), (N, X_i, 1, 0, \bar{\xi}), (s, X_i, y_i, 1, \xi)\} \\
T_i^{-1}(U) &= \{(N, X_i, 0, 0, \bar{\xi}), (U, X_i, y_i, 0, \xi)\}.
\end{aligned}
$$

Intuitively, because signal y_j is ignored by j under ξ_0, there is an additional incentive to play D in N, as this will be of no consequence. However, if L_0 is sufficiently small this will not be worthwhile.

4.3. 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,[16] 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

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

---

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

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 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, $Q_t \leq \min(t, \bar{Q})$ 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 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 (until a signal $z = 1$ occurs).[17] 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[18]

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

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.

---

[17]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 $\log K$.

[18]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.

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

4.4.1. "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.[19]

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 $\Pr(z_i) = q$ and $\Pr(z_1 = z_2)$ close enough to 1.

While signals can be private, correlation across signals cannot be too weak. We explore below the case of independent signals, and show that cooperation can not be sustained in equilibrium.

Formally, we consider the mental process as before:

$$\begin{aligned} T_i^{-1}(N) &= \{(N, X, 1, 0), (s, X, y_i, 1)\} \\ T_i^{-1}(U) &= \{(N, X, 0, 0), (U, X, y_i, 0)\}, \end{aligned}$$

with the qualification that T_i is now defined over Z_i rather than Z. We investigate whether and when the strategy that plays C in N and D in U is an equilibrium.

To fix ideas, we consider a special case here, and leave the general case to the appendix. We assume that p is close to 1, and that $q^2 << 1 - p << q$; thus q is small, but large compared to $1 - p$. Under σ, cooperation phases last $\frac{1}{2(1-p)}$ on average (because each player has a chance $(1-p)$ of switching to U in each period), but 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). Because punishment phases last much longer than cooperation phases, the value to following σ will be very close to 0. In contrast, by playing C in both states, player i ensures that "punishment" phases last only $\frac{1}{q}$ periods on average, while cooperation phases last $\frac{1}{1-p}$ periods; player i can thus secure a payoff close to 1 by playing C in both states. This is a profitable deviation, and the strategy σ cannot be an equilibrium.

4.4.2. 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 "special" gift 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.

---

[19] 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.

We consider a modification of the main example. Player 2's set of possible actions is as before, $\{C, D\}$. Player 1 has, independently across time, one of two sets of actions available:

$$A_1 = \{C, D\} \text{ with probability } 1 - q$$
$$A_1' = \{C, D, G\} \text{ with probability } q.$$

While we interpret C as a standard or routine gift, we interpret G as a particularly special gift (one that extraordinary effort went into). Action G requires a cost a, and it gives an expected benefit $b + L$ (instead of $(1 + L)$) to the one who receives it. When player 1 has the three actions available, expected payoffs become as follows:

<div align="center">

2

		C	D
1	C	$1, 1$	$-L, 1+L$
	D	$1 + L, -L$	$0, 0$
	G	$1 + L - \mathbf{a}, -L + \mathbf{b}$	$-\mathbf{a}, \mathbf{b}$

</div>

Signal structure.

We assume that for each player $i = 1, 2$, there are three possible signals $y_i \in \{0, 1, 2\}$ that player i gets. A signal corresponds to how player i perceived the gift he received. We assume that if a player did not put effort into choosing the gift, then the person receiving the gift will most likely not think highly of the gift.

Formally,

$$p = \Pr\{0 \mid D\} = \Pr\{1 \mid C\} = \Pr\{2 \mid G\}$$

[20] and, for all other pairs (y, X), i.e. $(y, X) \notin \{(0, D), (1, C), (2, G)\}$

$$\Pr\{y \mid X\} = \frac{1 - p}{2}.$$

We will consider again the case that p is close to 1.

States and transitions.

We assume as before that each player i may be in one of two states U or N. The transition function that we consider is below. For player 1,

$$T_1^{-1}(N) = \{(N, C, 1), (s, G, y_1)\}$$
$$T_1^{-1}(U) = \{(N, C, 0), (s, D, y_1)\},$$

and for player 2:

$$T_2^{-1}(N) = \{(N, C, 1), (s, X, 2)\}$$
$$T_2^{-1}(U) = \{(N, C, 0), (s, D, 0), (s, D, 1)\},$$

Compared to our original example, this transition function specifies state N follows when player 2 observes signal 2 or player 1 plays G. This assumes that in either case a switch back to the N state in the same way as signal z was resetting cooperation.

[20] Since, by assumption, player 2 does not have the option of choosing G, player 1 will never see signal $y = 2$.

Strategies:

As before, we consider strategies that condition only a player's state.[21]

$$\Sigma_1 = \{\sigma_1, \sigma_1 : S_1 \longrightarrow A'_1 \times A_1\} \text{ and } \Sigma_2 = \{\sigma_2, \sigma_2 : S_2 \longrightarrow A_2\}$$

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

$$
\begin{aligned}
\sigma_1(N) &= C \\
\sigma_1(U) &= G \text{ if available, otherwise } D
\end{aligned}
$$

and for player 2,

$$
\begin{aligned}
\sigma_2(N) &= C \\
\sigma_2(U) &= D.
\end{aligned}
$$

Intuitively, either player triggers a punishment phase when he chooses C and gets a bad signal. This punishment phase lasts as long as player 1 does not have the option to choose G. As soon as he has the option, he plays G to return to cooperation, which will happen if player 2 gets the signal $y_2 = 2$.

Analysis:

When players follow our candidate equilibrium strategy, as before, if p is close to 1 they alternate between long phases of cooperation and relatively short punishment phases. There are two main differences however.

First, player 2 only switches back to N if he sees $y_2 = 2$. So unless player 1 chooses G when available at U, players will not be able to exit from (U, U), (or only with the very small probability of an error in the signal). Second, switching from punishment to cooperation involves other payoffs than before: previously, at a date where players got $z = 1$, they would immediately switch to cooperation and get $(1, 1)$. Now, at the date where G becomes available to player 1 and player 2 is at state U, players get $(-a, b)$.

When player 1 plays D at N and (G, D) at U, he generates as before a very short cooperative phase. Additionally, he now has to incur the cost of resetting cooperation (he gets $-a$ instead of 1), so the constraint becomes:

$$L \le \Delta + 1 + a.$$

Player 2's incentive to play C at N is checked similarly. At the date when G becomes available, player 2 now gets b instead of 1, hence the constraint:

$$L \le \Delta + 1 - b.$$

When player 1 plays C at N and (G, C) at U, then as before, from state (N, N), in the event where player 1 receives a bad signal and action G is not immediately available, he avoids triggering a punishment phase and thus saves $\Delta + 1 + a - L$. However, in the event where player 2 receives a bad signal (hence switches to U and plays D) and when G is not available to player 1 at the following date, he plays C and loses L at each date of the punishment phase (until G

---

[21]Note that for player 1, the strategy specifies an action in A'_1 and an action in A_1, depending on which set of action is available.

becomes available), hence an expected loss of $L\Delta$. Thus for player 1 to have an incentive to choose D at U when G is not an option it must be that[22]

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

hence the conditions:

$$\frac{\Delta + 1 + a}{1 + \Delta} < L < \Delta + 1 - b.$$

4.5. 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 can not be satisfied for both players.

Differences in the players' signal technology 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.

---

[22]The two events described above have equal probability. There are also events where player 1 transits to U because recoordination did not work: player 2 did not get signal $y_2 = 2$ despite player 1 choosing G. When p is close to 1, however, such events are of second order.

5. Further discussion

5.1. 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 would obviously affect the possibilities of cooperation.

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.

5.2. 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.

5.3. Related literature

To be done.

6. Bibliography

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

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

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

Compte, O.,

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

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

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.

Okuno-Fujiwara, M. and A. Postlewaite, "Social Norms in Matching Games," *Games and Economic Behavior*, 1995, Vol. 9, pp79-109.

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.