The Extent of Cooperation*

Olivier Compte
PSE, Paris

Andrew Postlewaite
University of Pennsylvania

February 2009

Abstract

Compte and Postlewaite (2008, 2012) set out a model that assumed that peoples' behavior was organized via *mental systems*, and employed that framework to analyze how people might cooperate in a repeated relationship with private monitoring with plausible strategies. That paper focussed on the case in which two people played a repeated prisoners dilemma game. In this paper we show that the basic insights of that model can be extended to (i) repeated games in which play is sequential in each period rather than simultaneous; (ii) the games that are played in each period differ; (iii) repeated play in which players in a group are repeatedly randomly matched to play a repeated prisoners dilemma game; and (iv) how the framework can be used to understand incentive compatible social norms. We further analyze the evolutionary stability of the particular mental system in our earlier paper.

1. Introduction

Cooperation is difficult to sustain in repeated games when players get private signals of other players' actions. Cooperation in repeated games in which signals are public is typically supported by threats of punishment in the face of evidence of deviation from cooperative behavior. When players receive private signals about other's actions, punishment will be necessary if selfish behavior is to be suppressed. If the punishing actions are to be temporary, players must be able to coordinate both when to begin punishment phases and when to stop punishing and return to cooperative behavior. The difficulty in achieving this coordination in the face of private signals makes cooperation complicated in the standard model of repeated games.

Compte and Postlewaite (2009) (hereafter CP)[1] respectively) set out a model of behavior in a repeated game with private monitoring that restricted the set of strategies available to players

*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). Much, but not all, of the results in this paper appeared in an earlier paper, "Repeated Relationships with Limits on Information Processing". We thank Ted Bergstrom, Larry Samuelson and 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; their hospitality is gratefully acknowledged. The authors thank the Gould Foundation for financial support, and Postlewaite thanks the National Science Foundation for financial support.

[1]Compte and Postlewaite (2009) is a very much revised and shortened version of Compte and Postlewaite (2008).

in a natural way, and showed that in this restricted framework cooperation could be sustained with realistic strategies. CP defined a *mental system* as a finite set of mental states and a transition function describing a given player's change of mental state from one period to the next as a function of his action and the signals he receives in the period. Players are restricted to strategies that prescribe a pure action to every mental state they can be in. With this restriction, CP showed that cooperation in a repeated prisoners dilemma with private monitoring can be supported with intuitively plausible strategies.

That paper restricted attention to the classic two person prisoners dilemma repeated game. Players repeatedly played an identical game with an identical signal-information structure, simultaneously choosing their actions in each period. Our aim in that paper (and this one) is to set out a framework in which intuitively plausible strategies lead to cooperation in realistic repeated relationships. The restrictions embodied in the standard repeated game model are not, however, realistic. Two people in a repeated relationship tyically do not act simultaneously, but rather face a sequence of interactions in which one or the other is to act, but not both. In addition, the interactions they face are not identical: my partner may have prepared dinner when the I was not feeling well yesterday, I may do the laundry today while the my partner watches a favorite television program, and one of us may grocery shop tomorrow while the other 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 structures for all games in the sequence are identical.

We show in this paper that the restrictions that allow intuitively plausible cooperation in the repeated prisoners dilemma game also allow for cooperation in more general settings that include these kinds of repeated relationships. We also apply the framework to analyze repeated games with randomly matched pairs, and derive conditions under which cooperation in groups can be maintained when there is private monitoring. Lastly, we show how the framework can be used to understand incentive compatible social norms.

2. Model

Our analysis uses the basic model in Compte and Postlewaite (2009), which we describe next.

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 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.

2.1. Strategies

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.[2] Informational states capture the bounds on players' memories of the precise details of past play. For simplicity, we assume that in the current example the players can be in one of two states U(pset) or N(ormal). 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.

The analysis in CP centered on a leading example in which the transition function for both players was as in the figure below.

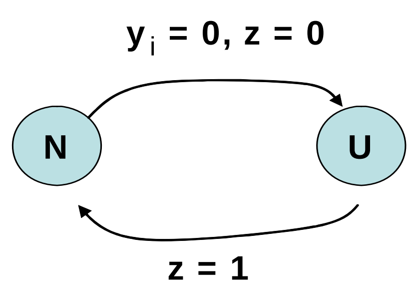

Figure 1: Transition

[2]See CP for a discussion and motivation for this structure.

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 $y = 0$ 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.[3]

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. CP analyzed when the following "cooperative" strategies:

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

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

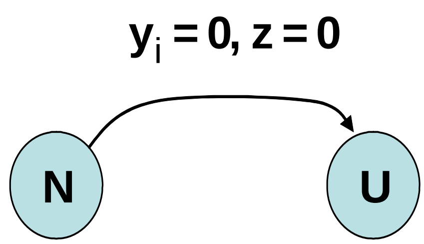

Figure 2: No "resetting"

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).[4]

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 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

[3] 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.

[4] The assumption of a public signal z that allows recoordination simplifies exposition. It is shown in CP how recoordination can be accomplished in the absence of public signals.

profile $\sigma(s)$, which in turn generates a probability distribution over signals, and 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.[5]

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.

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

2.3. Cooperation

CP showed that there was a range of the parameters q (the probability of resetting the relationship). p (the accuracy of the signals the players receive) and L (the gain from deviating) for which the cooperative strategies described above were an equilibrium. The argument can be summarized as follows.

When players follow the proposed 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]

[5] 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 $\phi_\sigma(s') = \sum_s Q_\sigma(s', s)\ \phi_\sigma(s)$.

[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.

[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 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 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, but 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 expected 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

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

Simply stated, if the gains from playing D while the other is playing C is sufficiently large, a player will gain by playing D always. There is also a lower bound on L that is consistent with the strategies being an equilibrium; if the loss when you cooperate and the other defects, L, is small, a player does better by always playing C and avoiding the costly punishment phase that follows the player's receiving an (incorrect) bad signal. This gain to player i when he plays C at both N and U is, of course, offset by the losses he incurs by remaining cooperative while player j is in a punishment phase. For the proposed strategies to be an equilibrium, 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 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 continues,[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

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

3. Extensions

The example in CP was kept simple in a number of ways to make clear how cooperation could be achieved when strategies are restricted. Some of the simplifications are not particularly realistic,

[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.

[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.

but 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 such extensions and modifications of the basic model.

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.

3.1. Sequential gift exchange

In CP 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, and then send a paper to you for your comments. Our effort levels – and consequently the signals received – are chosen sequentially rather than simultaneously.

We show next that allowing play to be sequential rather than simultaneous does not substantially alter the analysis in CP. To fix ideas, assume that in each stage player 1 moves first, then player 2, with signal z occurring at the end of the stage as before.[12] We examine the case where p is close to 1, and assume that players are endowed with the same mental process as in CP (described above), and characterize the conditions under which that process continues to enable cooperation.

As in the analysis of the game with simultaneous play, 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).[13]

Indeed since player 2 reacts within the same period to player 1's defection,[14] 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. Under 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 is:

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

[12] Other assumptions about when signal z occurs would not substiancially affect the analysis.

[13] Incentives are not altered for player 2. Note that players 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.

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

hence:

$$L > 1.$$

3.2. Heterogeneous stage games

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. A satisfactory analysis of any strategic behavior should 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. That 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 across periods as well. This is a caricature 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. We will outline how our approach can be extended to settings in which there is a sequence of games with possibly many actions, that are varying over time, and with a possibly richer signal structure than in CP. One of our central points is that our basic mental system, properly extended, may allow for cooperative behavior under broad circumstances.

Consider two people facing a general finite action game (with possibly many actions), that they play repeatedly and who discount the payoffs they receive. The interesting case is that in which there is a pair of actions, a_1^*, a_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 a_1^*, a_2^* to play a stage game Nash equilibrium. Suppose, however, that there is a monitoring problem as 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 a_1^* and a_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:

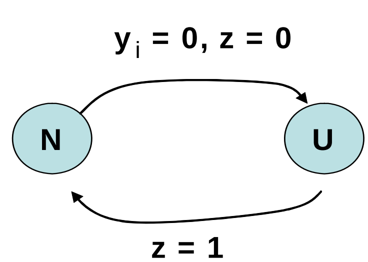

Figure 3: Transition

Player i's strategy is to play a_i^* in state N and $\bar{s}_i$ in state U, where $(\bar{a}_1, \bar{a}_2)$ 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 a_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 (a_1^*, a_2^*) Pareto dominate playing the Nash equilibrium $(\bar{a}_1, \bar{a}_2)$, 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 a_1^* in state N and playing something other than $\bar{a}_i$ 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 $(\bar{a}_1, \bar{a}_2)$ 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).[15] 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 modeler'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.

Finally, in CP, we suggested an intuitively plausible cataloguing of histories (defined by means of a plausible transition function). We suggest here a natural extension of CP to contexts in which the signal structure is richer than that explored in CP. A natural extension would be to define transitions so that a transition to the upset state occurs only if the likelihood ratio of defection versus cooperation is sufficiently large (given the signal received). Formally, one could set a threshold β that determines which signals are bad:

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

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.

[15] For ease of exposition, we assume below that each game ξ has a unique Nash equilibrium.

3.3. Heterogeneous agents

CP 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 may 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 altered 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 signal, so by choosing C in state U, player 1 mostly avoids triggering 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 triggering 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.

3.4. 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 the 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.

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.

As one would expect, there is a bound on how large the gain from deviating and playing D in N, and that bound depends on Q[17]

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

and a constraint on how small the gain can be to ensure that it is optimal to play D in U

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

The main difference from the two person leading example is that there, when player i deviated by playing D in N, 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 $\log K$.

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). (When K gets very large, a player has little chance to meet infected players before the end of the punishment phase, hence Q gets close to $\sum_t t(1-q)^t q = \Delta$).

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.

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

[17] The details of the calculations are left to the appendix.

4. Evolutionary pressures

We argued for the plausibility of our mental system framework on two grounds. First, that that individuals have a finite capacity for recalling histories of play, as captured by our assumption of a finite number of mental states. Second, the transition function that describes the movements among an individual's mental states is exogenously given. Both in CP and in the examples analyzed above, we illustrated how cooperation could be supported by realistic behavior in repeated relationships with private signals. The mental system in the examples had two states, U and N, and a transition function that takes a player from N to U when a bad signal is received, and takes him from U to N when he receives the signal $z = 1$. One can accept our explanation that there are biological constraints that rule out an infinite number of mental states, but be concerned that there are evolutionary pressures that would lead to an increase in the number of mental states: individuals with a larger number of mental states would seem to have an advantage over others with fewer states.

This is not necessarily the case. More mental states permit more complex strategies, but this is a two-edged sword. The possibility of more complex strategies can inhibity cooperation, as we show in the next section.

4.1. Pressures to increase to three states

Assume player 2 has a mental process (S, T) as described in figure 1 below, with $S = \{N, U\}$ and T as in our example. If player 1 has the same mental process as player 2, then the strategy $\sigma(N) = C$, $\sigma(U) = D$ supports cooperation under the usual conditions.

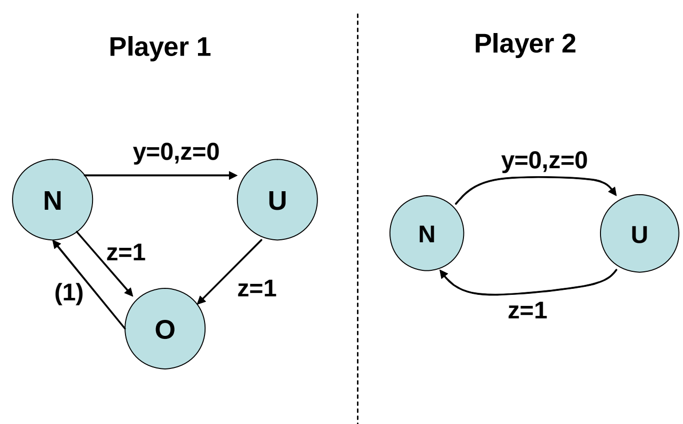

Figure 4

Now assume that player 1 has an additional state O that he uses right after signal $z = 1$. Also assume that player 1's mental system is (S', T') as illustrated in the figure (under this mental process, the first date we exit from the upset set, I ignore my signal), and that his strategy is:

$$\sigma(N) = \sigma(0) = C \text{ and } \sigma(U) = D.$$

If player 2 continues to use $\sigma(N) = C$, $\sigma(U) = D$, then the value to player 1 of following that strategy is larger than before (because player 1 avoids triggering unnecessary punishment - at the first date, a bad signal may arrive, yet for sure, the other player has played N).

So there would seem to be an evolutionary pressure towards the three state mental process. Now can cooperation be sustained? We have to see how incentives of player 2 are changed.

Incentives to play D at N: In that case, player 1 remains cooperative for two consecutive periods. So, for p close to 1, the condition becomes:

$$2L < \Delta$$

Incentives to play C at U: for p close to 1, the condition does not change because the probability that 1 is in U when 2 is in N is very small. For other values of p the condition would change, and become more stringent because it becomes more likely that player 1 does not trigger first. The condition would become:

$$\alpha(\Delta - L) + (1 - \alpha)(-L\Delta) < 0$$

with $\alpha > 1/2$, where α depends on p and q. So the condition would become more stringent as well.

To summarize, if player 1 mutates to the alternative mental system (S', T'), then the set of parameters (p, L) for which cooperation can be sustained is reduced.

When cooperation is still possible, what is the gain that player 1 makes? He avoids trigerring unnecessary punishments in the first period of the cooperation phase, so when p is close to 1, the gain is close to

$$(1 - p)(L - \Delta)$$

So if we have in mind that p is close to one and that L varies, there is no evolutionary pressure: the gain is small and the loss is comparable to the measure of events for which cooperation becomes impossible.

If p is not close to 1, or if L does not vary much and allows for cooperation even when 1 uses (S', T'), then there may be an evolutionary pressure towards (S', T'). However, note that (S', T') against himself is very bad: only defections may arise when two such people meet, because they learn that playing D is S is best. So if these are the only two mental processes available, we should end up, when p is not too small, with mostly (S, T) and a fraction of (S', T') comparable to $(1 - p)$.

We emphasize that the example is *not* to argue that more mental states is necessarily bad. Rather, our aim is to illustrate that even before biological costs of adding mental states is taken into account, increasing the number of mental states *may* be harmful.

4.2. Pressure towards stochastic transitions

We can imagine evolutionary pressure not only on the number of mental states, but also on the transition functions. In the example in CP, players transit from state N to U when they receive a bad signal. What if they transit with probability less than 1 in this case? The signals that the players receive are noisy, and as a result they periodically enter a costly punishment phase. Reducing the probability that they transit from N to U will decrease the frequency of such transitions. We examine next the evolutionary pressure on the transition function.

Suppose the transition function for the agents is as in the figure below.

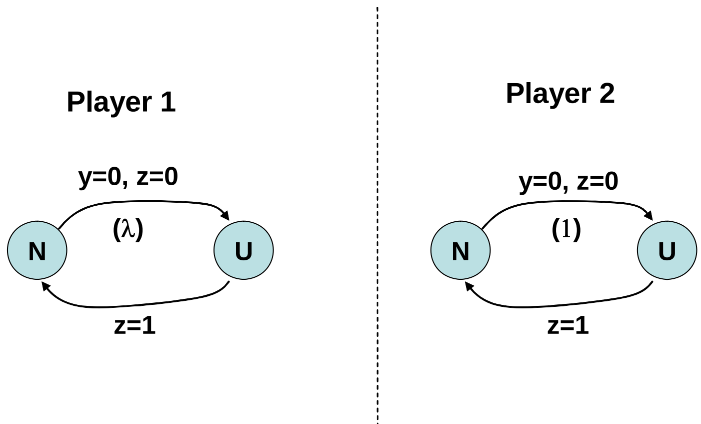

Figure 5

The scalar $\lambda \in [0,1]$ denotes the probability that player 1 switches to the upset state from the normal state when signals $y = 0$ and $z = 0$ have been observed by player 1. We denote by $T^{(\lambda)}$ that transition. Player 2's transition remains as before. The evolutionary pressure towards smaller values of λ for player 1 depends on the range of parameters for which σ^* ($\sigma^*(N) = C$ and $\sigma^*(U) = D$) is an equilibrium when $T_1 = T^{(\lambda)}$ and $T_2 = T^{(1)}$. We examine the case where p is close to 1.

4.2.1. Incentives of player 1

Incentives to play C at N are as before: if player 1 plays D all the time, player 2 will be playing C a fraction $1/(1+\Delta)$ of the time. So as before, $(1+L)/(1+\Delta) < 1$.

Incentives to play D at U are now stronger, because now it is more likely that we exit from (NN) to go to (NU) than to go to (UN), so there is now a higher chance that player 2 has transited. The condition is now:

$$\lambda p(1-p)[\Delta - L] + p(1-p)(-\Delta L) \leq 0,$$

hence,

$$L \geq \frac{\lambda\Delta}{\lambda + \Delta}$$

which is less stringent than before. Hence, the set of parameters for which it is optimal for player 1 to play σ^* increases when λ is less than 1.

4.2.2. Incentives of player 2

Incentives to play C at N are weaker, because player 1's reaction is delayed: when player 2 plays D all the time, player 1 exits from the cooperative phase with probability $\lambda(1-q)$, and he exits from U with probability q, hence expected durations are respectively $\frac{1}{\lambda(1-q)}$ and $1/q$.

Player 2 gains L per period in the cooperative phase, and looses 1 per period in the defecting phase, hence the condition:

$$\frac{1}{\lambda(1-q)}L < \frac{1}{q}$$

or equivalently

$$L < \lambda\Delta.$$

The condition is thus more stringent than before.

For 2's incentive to play D at U, there are two effects. Playing D is less costly than before, because player 1 reacts more slowly on average. However, it is also more likely that player 2 becomes upset before player 1 does, which creates an incentive for player 2 not to trigger the punishment. Set Δ^* to solve $\Delta^2 = 1 + \Delta$. We show below that if $\Delta > \Delta^*$, the latter effect dominates and creating incentives for player 2 to trigger the punishment is more difficult.

Recall that as in the case where $\lambda = 1$ examined earlier, whether player 2 plays D or C at U, entry to and exit from NN is unchanged, and so does entry and exit from UN. The important difference is that NU is more likely when player 2 remains cooperative in U. The cost of course is that player 2 loses L in UU.

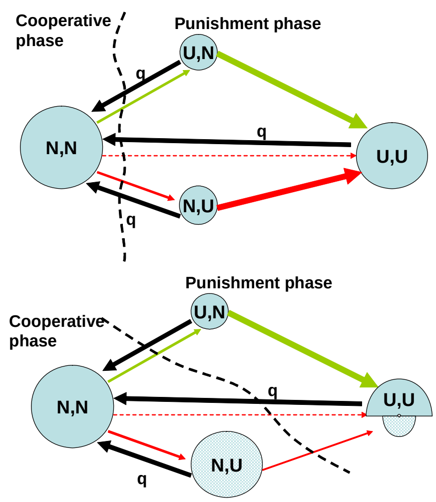

Figure 6

In the event exit from NN leads players to UN (which happens with probability $(1-q)p(1-p)$), the strategy σ^C (which plays C at both states) generates a loss equal to ΔL (because conditional on UN, the expected length of stay in UU is Δ (as before).

In the event exit from NN leads to NU (which happens with probability $(1-q)\lambda p(1-p)$), the strategy σ^C avoids trigering a punishment phase. Consider a punishment phase started by player 2. Since the transition of player 1 is stochastic, player 1 remains in N a fraction

$$\pi = \frac{q}{q + (1-q)\lambda}$$

of the time (until $z = 1$ occurs) (during which player 2 would an extra per period payoff L), and in U a fraction $1 - \pi$ of the time (during which player 2 would incur a per period loss of 1). It follows that by triggering the punishment, player 2 gets

$$\frac{1}{q}[\pi L + (1-\pi)(-1)].$$

Given the respective probabilities of reaching NU rather than UN, the condition becomes:

$$\lambda[-\Delta L] + \frac{1}{q}[\pi L + (1-\pi)(-1)] > 0,$$

or equivalently

$$L \geq \frac{\Delta\lambda}{1 + \lambda\Delta[\frac{1+\lambda\Delta}{1+\Delta}]}. \quad (4.1)$$

Note that for $\lambda = 1$, we have the usual condition.

Define $f(\lambda)$ as the RHS of the above inequality. It is easy to verify that $f'(1)$ has the same sign as $1 + \Delta - \Delta^2$. So for $\Delta > \Delta^*$, the condition (4.1) is more stringent when λ is smaller than 1.

To summarize: if we fix p close to 1 and $\Delta > \Delta^*$, then the effect of reducing λ is to unambiguously reduce the range of values of L for which cooperation can be sustained.

Finally, we can examine whether in the range where players can sustain cooperation, there is an incentive for player 1 to decrease λ.

One can check that if

$$L < \Delta\frac{1+\Delta}{1+2\Delta}$$

then player 1 (actually each player) would be better off with a lower value of λ (so long as cooperation can still be sustained). The intuition is that by switching to U with probability λ smaller than one (under the event $y = 0, z = 0$), player 1 gives another chance to cooperation (after all, there is equal chance that either player triggers the punishment). By staying in the normal state, he checks that indeed it is the other player that has already triggered the punishment. This strategy is worthwhile so long as making a gift (continuing to cooperate) is not too costly.

The reason that a stochastic transition may be better for a player is related to the fact that the one-shot deviation principle (but not quite the usual version though) fails in our case. Going slightly away from the determistic transition would amount to allowing a player to decide whether he really wants to be upset whenever the signals call for being upset. When $L < \Delta\frac{1+\Delta}{1+2\Delta}$, he would indeed prefer to continue behaving as if he was not upset.

*To summarize: For a **fixed** $L \in (\Delta\frac{1+\Delta}{1+2\Delta}, \Delta)$, there is no incentive to move to a stochastic transition. For a **fixed** $L \in (\frac{\Delta}{1+\Delta}, \Delta\frac{1+\Delta}{1+2\Delta})$, there is an incentive to move to a stochastic transition.[18] One then might want to conclude that there would be an*

[18] This is similar to Wilson's (2004) analysis showing the benefit of stochastic transition functions.

evolutionary pressure away from our deterministic transition. The analysis above, however, suggests that if the mental process is applied to a variety of environmnets, one must determine how a change in the mental process affects cooperation accross all possible realizations of the environment. In our example, with $\Delta > \Delta^*$, a smaller λ would reduce the scope for cooperation, so even if there are realizations of L below $\Delta\frac{1+\Delta}{1+2\Delta}$, there may not necessarily be evolutionary pressure towards stochastic transitions.

5. Appendix

Cooperation with many players

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,[19] 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

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

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

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

punishment phase (i.e. until a signal $z = 1$ occurs).[20] 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[21]

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

6. Bibliography

Compte, O., and A. Postlewaite [2008], “Repeated Relationships with Limits on Information Processing,” mimeo, University of Pennsylvania.

Compte, O., and A. Postlewaite [2012], “Plausible Cooperation,” mimeo, University of Pennsylvania.

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

[20] 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. 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 $\log K$.

[21] 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.