Games and Economic Behavior 91 (2015) 45–59

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Plausible cooperation [☆]

Olivier Compte [a], Andrew Postlewaite [b,*]

[a] Paris School of Economics, 48 Bd Jourdan, 75014 Paris, France
[b] Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6207, United States

ARTICLE INFO

Article history:
Received 4 April 2013
Available online 30 March 2015

JEL classification:
D01
D70

Keywords:
Repeated games
Private monitoring
Bounded rationality
Cooperation

ABSTRACT

There is a large repeated games literature illustrating how future interactions provide incentives for cooperation. Much of the earlier literature assumes public monitoring. Departures from public monitoring to private monitoring that incorporate differences in players' observations may dramatically complicate coordination and the provision of incentives, with the consequence that equilibria with private monitoring often seem unrealistically complex or fragile. We set out a model in which players accomplish cooperation in an intuitively plausible fashion. Players process information via a mental system – a set of psychological states and a transition function between states depending on observations. Players restrict attention to a relatively small set of simple strategies, and consequently, might learn which perform well.

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 restart 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 noncooperative partner. Such simple behavior seems 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, that is, when our partner does not always know if we are less than satisfied with his effort.

Despite the fundamental importance of cooperation in understanding human interaction in small or large groups, the theory of repeated games, while providing important insights about repeated interactions, does not capture the simple intuition in the paragraph above. When signals are private, the quest for "stable" rules of behavior (or equilibria) typically

☆ This paper is a substantially revised version of earlier working papers: "Repeated Relationships with Limits on Information Processing", "Effecting Cooperation" and a previous version of the current paper. We thank Ted Bergstrom, Larry Samuelson, Alvaro Sandroni 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 (grant SES-1260753).

* Corresponding author.
E-mail addresses: compte@enpc.fr (O. Compte), apostlew@econ.upenn.edu (A. Postlewaite).

http://dx.doi.org/10.1016/j.geb.2015.03.010
0899-8256/

produces strategies that are finely tuned to the parameters of the game (payoffs, signal structure),[1] or to the assumed sequencing/timing of actions and signals.[2] These "rules of behavior" fail to remain stable when the parameters of the game are changed slightly.[3] Their robustness to changes in timing is typically not addressed, nor is their plausibility. We propose an alternative theory/description of how strategic players accomplish cooperation via realistic and intuitively plausible behavior.

A descriptively plausible theory of cooperation should have realistic strategies. A strategy is a complex object that specifies behavior after all possible histories, and the number of possible histories increases exponentially with the number of interactions. If I and my spouse alternate cooking dinner and whoever cooks can either shirk or put in effort each time he cooks, there will be approximately a billion possible histories after one month. For each of these billion histories, both I and my spouse will have received imperfect signals about the effort put in by the other on the nights they cooked, and for each of the histories, I must decide whether or not to put in effort the next time I cook. It is inconceivable that I recall the precise history after even a month let alone after several years.

A more realistic description is that I rely on some summary statistic in deciding whether or not to put in effort – the number of times it seemed effort was put in over the past several times my spouse cooked, for example. In this way, histories are catalogued in a relatively small number of equivalence classes or states.[4] The pooling of histories into classes is intuitively plausible – not just a mathematically convenient pooling, and my action today depends only on the equivalence class containing the history. The strategies we consider will conform to these desiderata.

A second property of a plausible theory of cooperation is that the equilibrium behavior should be consistent with agents coming to that behavior. In the standard approach to repeated games there is typically 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 in repeated games with private signals.[5] Equilibrium strategies in such a setting typically rely on my knowing not only the distribution of signals I receive conditional on the other player's actions, but also on 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. Alternatively, one might posit that players could "learn" the equilibrium strategies. However, the set of strategies is huge and it is difficult to see how a player might learn which strategies work well. Even if one restricted attention to strategies that are deterministic functions of histories, finding an optimal strategy amounts to finding an optimal partition of the histories among all possible partitions.

Our view is that the second property – that agents might come to behave in accordance with equilibrium predictions – calls for a restriction on the strategies that agents choose among. The nature of the restriction one imposes is a critical issue. In this paper, we propose a priori constraints on how agents process signals, motivated by plausible psychological considerations or cognitive limitations, and ask when the restricted family of strategies so generated is conducive to cooperation.[6]

To summarize, 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 based on a cataloging of histories that is intuitively plausible; (iii) the sets of strategies allow agents to cooperate under a broad set of circumstances; and (iv) equilibrium cooperation obtains in a way that is robust to the parameters of

---

[1] See in particular the belief free literature in repeated games (Piccione, 2002; Ely and Välimäki, 2002). See Compte and Postlewaite (2013) for a critique.

[2] Repeated relationships are typically modeled as a stage game played repeatedly, with the players choosing actions simultaneously in the stage game. In reality, the players may be moving sequentially and the signals they get about others' actions may not arrive simultaneously. The choice to model a repeated relationship as simultaneous play is not based on a concern for realism, but for analytic convenience. A plausible theory of cooperation should not hinge on the fine details of the timing of actions: we should expect that behavior that is optimal when play is simultaneous to be optimal if players were to move sequentially.

[3] Fundamental to the standard approach to repeated games with private signals is the analysis of incentives of one party to convey to the other party information about the private signals he received, either directly (through actual communication), or indirectly (through the action played). Conveying such information is necessary to build punishments that generate incentives to cooperate in the first place.

Incentives to convey information, however, are typically provided by making each player indifferent between the various messages he may send (see (Compte, 1998; Kandori and Matsushima, 1998)), or the various actions he may play (belief free literature). There are exceptions, and some work such as Sekiguchi (1997) or Compte (2002) *does* have players provided with *strict* incentives to use their observation. But, these constructions rely on fine tuning some initial uncertainty about the opponent's play (as in the work of Bagwell, 1995).

Finally, when public communication is allowed and signals are not conditionally independant, strict incentives to communicate past signals truthfully may be provided (Kandori and Matsushima, 1998), but the equilibrium construction relies on simultaneous communication protocols.

[4] Aumann (1981) suggested modeling agents as having a finite number of states of mind. This has led to the study of repeated game equilibria in which players are *constrained* to using finite automata (Rubinstein, 1986; Neyman, 1985, 1998; Ben-Porath, 1993), and to the study of repeated game equilibria in which strategies can be *implemented* by simple automata (Abreu, 1986), or approximated by finite automata (Kalai and Stanford, 1988).

[5] In a different context (that of repeated games with perfect monitoring), Gilboa (1988) and Ben-Porath (1990) have expressed a related concern, distinguishing between the complexity associated with *implementing* a repeated game strategy, and the complexity associated with *computing* best response automata. Our concern is not computational complexity per se, but rather the ability to perform relevant computations without precise knowledge of distributions.

[6] Although Aumann (1981) is not motivated by learning considerations, he mentions that assuming a bound on the number of states of mind would "put a bound on the complexity a strategy can have, and enable an analysis in the framework of finite games." In particular, in the context of a repeated prisoners' dilemma with perfect observations, he reports an example in which only few strategies are compared. Although Kalai et al. (1988) argue that the example lacks robustness, the path we follow is in the spirit of Aumann's example.

the game and the timing of players' actions. This goal motivates the model that we set out below. We do not claim that this model is unique in achieving our goal, only that it is a plausible model that satisfies our stated desiderata.

Before going on we should emphasize two things. First, cooperation is not always possible in our framework, and second, even when equilibrium cooperation is possible, there is also an equilibrium in which cooperation doesn't obtain. We will have nothing to say about what determines whether the cooperative equilibrium or the uncooperative equilibrium arises when both exist. Although this is an interesting question, and although we believe that our framework makes it more plausible that players would learn to cooperate, this is beyond the scope of this paper.

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, in restrictions that might arise naturally.

As suggested above, we shall first endow each player with a finite set of *mental states*, restricting a player to behaving the same way for all histories of the game that lead to the same mental state. In addition, we shall endow agents with *transition functions* that describe what combinations of initial mental state, actions and signals lead to specific updated mental states. Mental states and transitions jointly define what we call a *mental system*. Contrasting with traditional approaches to bounded rationality in repeated games (i.e., the automaton literature), we do not attempt to endogenize the mental system that would be optimal given the parameters of the problem and behavior of other players. We do not think of a mental system as a choice variable, but rather as a natural limitation of mental processing. For example, 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. In principle, transitions could be a complex function of past histories, with evidence accumulating in the background, up to the point where, suddenly, one feels cheated. A fundamental aspect of the transition function in our model is that it is exogenous: the individual does not have control over it.

A mental system characterizes how signals are interpreted and processed. Modeling mental systems as exogenous reflects our view that there are limits to peoples' cognitive abilities, and that evolution and cultural indoctrination should have more influence on one's mental system than the particular circumstances in which a specific game is played. 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 shapes the child's interpretations.

Taking the mental system as exogenous has one obvious consequence: this restricts the strategies *available* to agents. The restriction may be drastic when there are only few mental states. But there is still scope for strategic choices: even if upset, we assume that the agent still has a choice of what action to choose (either defect or cooperate in the case of a prisoner's dilemma). In other words, while the mental system is assumed to be the same across a variety of games, how one *responds* to being upset is assumed to be situational, and to 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 when upset, and risk breaking a relationship that is generally cooperative, but not hesitate when the cost is large.

From a positive perspective, our model provides a simple and plausible theory of how players manage to cooperate despite the fact that signals are private. Strategic choices are limited yet they are sufficient to highlight (and isolate) the two key differences between games with public and private monitoring:

(a) the difficulty in providing incentives to trigger punishments: a player may decide to ignore a bad signal, betting that he is the only one that got a bad signal.
(b) the difficulty in recoordinating to cooperation once a punishment phase has been triggered, as there is no longer a public signal that players can use to simultaneously go back to cooperation.

On issue (a), we find that providing incentives to trigger a punishment requires that the cost incurred from remaining cooperative while the other defects be sufficiently large. On issue (b), we find that mental systems that generate forgiveness (i.e. responsive to good signals) and some leniency (i.e. not too responsive to bad signals) are conducive to cooperation.

Technically, the approach taken here is simpler because the repeated interaction is reduced to a static game with a small number of strategies and checking incentives is easier. We check that one has incentives to defect when Upset (i.e. issue (a)) without checking that incentives hold for each possible history leading to the Upset mental state: histories are pooled.

Plan Section 2 provides a formal model which follows the steps described above. That section focuses on issue (a), assuming that there is occasionally a public signal that facilitates periodic synchronization. We analyze the circumstances in which cooperation is possible and discuss the robustness of our result. Next, in Section 3, we drop the assumption of a public signal and show how synchronization can be accomplished without such a signal (issue (b)). In Section 4 we discuss the results and possible extensions.

2. Model

Gift exchange There are two players who exchange gifts each period. Each has two possible actions available, one that corresponds to not making an effort in choosing a gift and a second corresponding to making a costly effort. Gifts may or may not be perceived as “thoughtful”, and a gift is more likely perceived as thoughtful when the other makes costly effort.

Payoff structure Actions are $\{C, D\}$ with C representing costly effort. The expected payoffs to the players are as follows:

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

where L corresponds to the cost of effort.

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 is correlated with the other player's action. Formally,

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

We assume that $p > 1/2$ so that one can refer to $y_i = 0$ as a “bad” signal and $y_i = 1$ as “good”.

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

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

The existence of a public signal z facilitates our exposition but can be dispensed with, as we demonstrate in Section 3.

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 traditional models. A player is endowed with a *mental system* that consists of a finite set S_i of *mental states* the player can be in, and a *transition function* T_i that describes what triggers moves from one state to another: the function T_i determines the mental 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 outcomes of that period and possibly previous periods. The restriction we impose is that a player may only condition his behavior on his mental state, and not on finer details of the history.[7] Given this restriction, all mappings from states to actions are assumed admissible. Player i's set of pure strategies is[8]:

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

We will illustrate the basic ideas with an example in which the players can be in one of two states U(pset) or N(ormal).[9] The names of the two states 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.[10] Both S_i and T_i are exogenously given, not a choice: 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 transition function for the example, which we will refer to as the leading example below, is as in Fig. 1.[11]

Fig. 1 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 $y_i = 0$ 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_i.[12] The mental system thus determines how observations are aggregated over time, hence how histories are pooled: some histories lead to state N, others lead to state U.

---

[7] Note that our structure requires that players' strategies be stationary: they do not depend on *calendar time*. This rules out strategies of the sort “Play D in prime number periods and play C otherwise”, consistent with our focus on behavior that does not depend on fine details of the history.

[8] We restrict attention to pure strategies in this paper; see Compte and Postlewaite (2012) for a discussion of mixed strategies.

[9] The restriction to two mental states is for expository purposes. The basic insights that cooperation can be sustained via intuitively plausible strategies continue to hold when agents have more finely delineated mental states; we discuss this at the end of this section.

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

[11] See Compte and Postlewaite (2008) for a discussion of more complicated transition functions and for a discussion of the robustness of the analysis of this model to changes in the timing of decisions and heterogeneity in monitoring technologies.

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

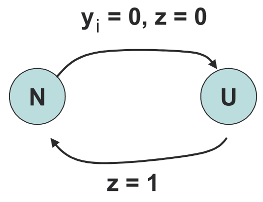

Fig. 1. Transition.

The simple form of the transition function – that given the previous state and the action taken, it depends only on the most recent signal – is for simplicity. In principle, the transition function could depend on more than this most recent signal, for example, whether two of the past three signals was "bad", or it could also be stochastic. We consider stochastic transitions in Section 3.

Given the mental system above, our candidate behavior for each player i will be as follows,

$$\sigma_i(N) = C$$
$$\sigma_i(U) = D.$$

That is, player i plays C as long as $y_i = 1$, or when $z = 1$. He plays D otherwise. Player 1 triggers a "punishment phase" when he sees the bad signal, $y_i = 0$.[13] This punishment phase ends only when signal $z = 1$ is received.

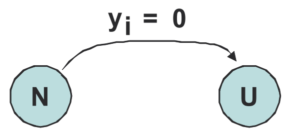

Fig. 2. No "resetting".

The public signal z gives the possibility of "resetting" the relationship to a cooperative mode. If the signal z is ignored or absent and the mental process is defined by Fig. 2, then eventually, because signals are noisy, with probability 1 the players will get to state U under the proposed strategy and this will be absorbing: there would be nothing to change their behavior. The signal z allows for possible recoordination back to state N (and possibly cooperation).

In our leading example, players transit simultaneously from U to N for exogenous reasons. Alternatively, in a two-state mental system the players could move from state U back to state N after seeing a good signal.

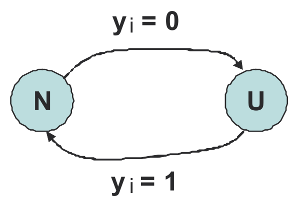

Fig. 3. Forgiving transition.

Fig. 3 describes such a transition function. A player endowed with this alternative mental process, who would cooperate in N and defect in U, would be following a TIT for TAT strategy.[14]

2.2. An illustrative experiment

Before continuing with the formal description of our model, it is useful to give a real-world example to illustrate our idea of a mental system. Cohen et al. (1996) ran several experiments in which participants (students at the University of Michigan) were insulted by a confederate who would bump into the participant and call him an "asshole". The experiment was designed to test the hypothesis that participants raised in the north reacted differently to the insult than did participants raised in the south. From the point of view of our model, what is most interesting is that the insult triggered a physical response in participants from the south. Southerners were upset by the insult, as shown by cortisol levels, and more physiologically primed for aggression, as shown by a rise in testosterone. We would interpret this as a transition

[13] The play induced by this strategy is similar to the equilibrium play in Sekiguchi (1997).
[14] We show below that cooperation is essentially impossible if players have this mental process.

from one mental state to another, evidenced by the physiological changes. This transition is plausibly not a choice on the participant's part, but involuntary. The change in mental state that is a consequence of the insult was followed by a change in behavior: Southerners were more likely to respond in an aggressive manner following the insult than were northerners. Moreover, Southerners who had been insulted were more than three times as likely to respond in an aggressive manner in a word completion test than were Southerners in a control group who were not insulted. There was no significant difference in the aggressiveness of Northerners who were insulted and those who were not.

The physiological reaction to an insult – what we would think of as a transition from one state to another – seems culturally driven: physiological reactions to insults were substantially lower for northern students than for southern students.

Indeed, the point of Cohen et al. (1996) is to argue that there is a southern "culture of honor" that is inculcated in small boys from an early age. This culture emphasizes the importance of honor and the defense of it in the face of insults. This illustrates the view expressed above that the transition function in our model can be thought of as culturally determined.

2.3. Ergodic distributions and strategy valuation

For any pair of players' strategies there will be an ergodic distribution over the pairs of actions played.[15] The ergodic distribution gives the probability distribution over payoffs in the stage game, and we take the payoff to the players to be the expected value of their payoffs given this distribution.

Formally, define a state profile s as a pair of states (s_1, s_2). Each strategy profile σ induces transition probabilities over state profiles: by assumption each state profile s induces an action profile $\sigma(s)$, which in turn generates a probability distribution over signals, and hence, given the transition functions T_i, over next period states. We denote 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.[16]

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.[17] 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.[18]

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.[19,20]

2.4. Successful cooperation

We are interested in equilibria in which the players cooperate at least some of the time asymptotically. This requires players playing the strategy "play C in N and D in U".[21] We denote this strategy by σ_i^* for player i. We look for the set of parameters p, q and L for which $\sigma^* \equiv (\sigma_1^*, \sigma_2^*)$.

---

[15] While in general the ergodic distribution may depend on the initial conditions, we restrict attention to transition functions for which the distribution is unique.

[16] Formally, define $Q_\sigma(s', s)$ as the probability that next state profile is s' when the current state is s. That is, Q_σ is the transition matrix over state profiles induced by σ. The vector ϕ_σ solves $\phi_\sigma(s') = \sum_s Q_\sigma(s', s)\ \phi_\sigma(s)$.

[17] 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, so they would remain equilibria under this alternative definition for discount factors sufficiently close to 1.

[18] The distinction between ex ante and interim incentives is irrelevant in our framework. When a strategy profile $\sigma = (\sigma_1, \sigma_2)$ is played, the value that player i obtains is $v_i(\sigma)$ and it is computed by considering the ergodic distribution over state profiles induced by σ. Neither the date at which this computation takes place, nor the beliefs that player i might have about the other player's current mental state are specified.

[19] We restrict attention to pure strategies. However, our definitions can be easily generalized to accommodate 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.

[20] Note that σ_i as defined should not be viewed as a strategy of the repeated game. 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.

[21] This is because there cannot be an equilibrium where a player cooperates always, or when he plays D in N and C in U. Indeed, when player 2 plays D in N and C in U, then defecting always is the best response for player 1 because it maximizes the chance that player 2 cooperates.

For given p and q, L cannot be too large. If L is sufficiently large it will pay a player to deviate to "always defect": the occasional reward of L to the deviating player will be more than enough to compensate for causing the opponent to very likely switch from N to U.

In addition, L cannot be too small. When a player gets a bad signal he is not sure if his opponent is now in U and is playing D, or if the signal is a "mistake". If it's the latter case, playing D will likely lead to a spell of noncooperation. If L is small, there is little cost to playing C to avoid this; thus there is a lower bound on L that is consistent with the strategies above being an equilibrium.

There is thus an upper bound and a lower bound on L. The upper bound reflects a standard consideration, that gains from deviating from cooperation should not be too large; otherwise they cannot be offset by punishment phases. The lower bound is specific to private monitoring games: private signals can always be ignored, so incentives to trigger punishment have to be provided; and such incentives obtain when the cost of remaining cooperative while the other defects is too large.

We now turn to a formal derivation of these two bounds, showing that they can be compatible. We refer to ϕ_{ij} as the long-run probability that player 1 is in state $i \in \{U, N\}$ while player 2 is in state j when both players follow σ^*.[22] By definition we have

$$v_1(\sigma_1^*, \sigma_2^*) = \phi_{NN} - L\phi_{NU} + (1+L)\phi_{UN}.$$

By symmetry, $\phi_{NU} = \phi_{UN}$, so this expression reduces to

$$v_1(\sigma_1^*, \sigma_2^*) = \phi_{NN} + \phi_{UN} = \Pr_{\sigma^*}(s_2 = N) \equiv \phi_N.$$

Consider now the alternative strategy σ^D (respectively σ^C) where player 1 plays D (respectively C) in both states U and N. Also call ϕ_j^D (respectively ϕ_j^C) the long-run probability that player 2 is in state $j \in \{U, N\}$ when player 1 plays the strategy σ^D (σ^C) and player 2 plays σ_2^*. We have:

$$v_1(\sigma^D, \sigma_2^*) = (1+L)\phi_N^D.$$

This expression reflects the fact that playing σ^D induces additional gains when the other is in the normal state; but this has a cost because of an adverse effect on the chance that player 2 is in the normal state ($\phi_N^D < \phi_N$ when $p > 1/2$). The expressions above imply that the deviation to σ^D is not profitable when

$$L \le \bar{L} = \frac{\phi_N}{\phi_N^D} - 1. \tag{2.1}$$

When player 1 plays σ^C, he obtains:

$$v_1(\sigma^C, \sigma_2^*) = \phi_N^C - L\phi_U^C.$$

This expression reflects the fact that playing σ^C changes the probability that player 2 is in N ($\phi_N^C > \phi_N$ when $p > 1/2$) in a way that benefits player 1: player 2 is more likely to be in N when player 1 always cooperates than when he follows σ^*, because he avoids triggering a punishment/Upset phase when he receives bad signals by mistake. But this has a cost: he loses L whenever player 2 is in U. The deviation to σ^C is thus not profitable when

$$L \ge \underline{L} \equiv \frac{\phi_N^C - \phi_N}{\phi_U^C} = \frac{\phi_U - \phi_U^C}{\phi_U^C} = \frac{\phi_U}{\phi_U^C} - 1 \tag{2.2}$$

The following Proposition, which is obtained by using exact expressions for the long run distributions, shows that the inequalities are compatible for any $p > 1/2$ and any $q \in (0, 1)$, and that it is an equilibrium for both agents to play the prescribed strategies.

Proposition 1. *Consider the transition rule for the leading example, the strategies $\sigma_1 = \sigma_2 = \sigma^*$ and the bounds on $\underline{L}$ and $\bar{L}$ described above. Then*

i. *for any $p > 1/2$ and any $q \in (0, 1)$, we have $0 < \underline{L} < \bar{L}$. Additionally, for any q, both $\bar{L}$ and $\underline{L}$ are increasing with p;*
ii. *for any $L \in (\underline{L}, \bar{L})$, it is an equilibrium for both agents to play the strategy C in N and D in U.*

The shaded region in the graph (Fig. 4) shows how the range of L for which cooperation is possible varies as a function of p for the particular value of q equal .3.

[22] Long run probabilities are easy to compute. For example, the long run probability ϕ_{NN} satisfies

$$\phi_{NN} = q + (1-q)p^2\phi_{NN}$$

(both players are in state N either because resetting to cooperation occurred, or because players were already in state N and no mistake occurred). Solving, this gives $\phi_{NN} = \frac{q}{1-(1-q)p^2}$.

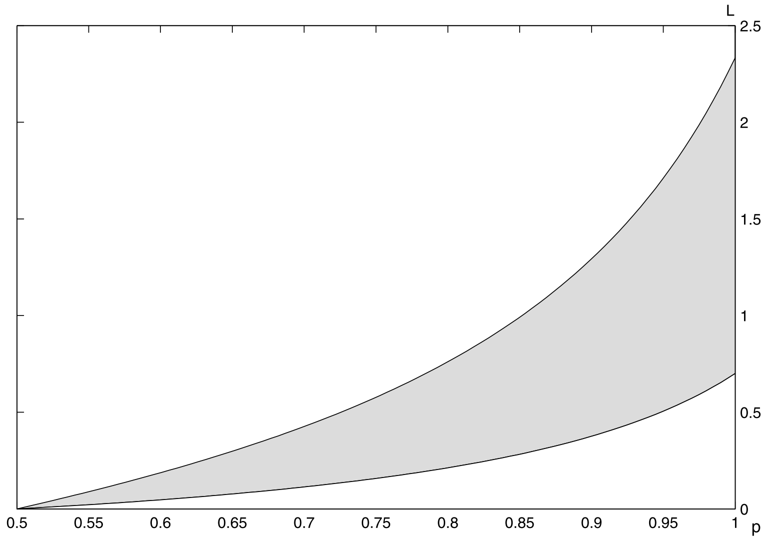

Fig. 4. $p - L$ combinations that allow cooperation for $q = .3$.

Note that as long as the signal about the opponent's effort is informative, there are some values of L for which cooperation is possible. Both the lower bound and upper bound on such L's are increasing as p increases, as is the size of the interval of such L's.

The detailed proof is in Appendix A.[23] We check here that the bounds $\bar{L}$ and $\underline{L}$ are compatible for p close to 1.

When p is close to 1, mistakes are rare, so $\phi_N \simeq 1$. When player 1 always defects, player 2 essentially receives only bad signals. If in state N, this signal triggers a change to U with probability $(1-q)$. Since it takes on average $1/q$ periods to return to N, the fraction of the time player 2 spends in N when player 1 always plays D is $\frac{1}{1+(1-q)/q} = q$, hence $\phi_N^D \simeq q$. This gives us the upper bound

$$\bar{L} = \frac{1}{q} - 1.$$

Now assume player 1 follows σ^C. Mistakes by player 2 occur with probability $(1-p)$, leading with probability $(1-q)$ to an Upset phase of length $1/q$, hence $\phi_U^C \simeq (1-p)(1-q)/q$. When player 1 is following σ^*, he now reacts to bad signals, and these mistakes also induce player 2 to switch to U the following period (unless signal $z=1$ arises). So in expectation, the length of time during which player 2 is upset following a mistake by player 1 is $(1-q)\phi_U^C$, and thus $\phi_U \simeq \phi_U^C + (1-q)\phi_U^C$. This implies:

$$\underline{L} = \frac{\phi_U}{\phi_U^C} - 1 \simeq 2 - q - 1 = 1 - q < \bar{L}.$$

2.5. Tit-for-Tat

A mental system generates restrictions on the set of strategies available to players and these restrictions eliminate potentially profitable deviations. It is not the case, however, that seemingly reasonable mental systems necessarily make cooperation possible. The Forgiving mental system (see Fig. 3) causes a player to be responsive to good signals: good signals make him switch back to the normal state.

Along with the strategy of playing C in N and D in U, this mental system induces a Tit-for-Tat strategy. With such a mental system however, for almost all values of p the only equilibrium entails both players defecting in both states.

Proposition 2. *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.*

We leave the proof of this proposition to the appendix, but give the basic idea here. With the Forgiving mental system, player 2's action only depends on the most recent signal. Thus, at any date, player 1's expected payoff *at that date* is only determined by his current action, and the action he took in the previous period. It follows that if all strategies were

[23] We thank an anonymous referee for improving the second part of the proposition.

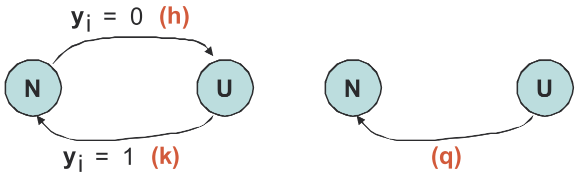

Fig. 5. Modified Tit-for-Tat and independant resetting.

feasible, then a particular *deterministic* sequence of play would be optimal: either C at all periods, or D at all periods, or the infinite alternation CDCDC.... The only way a stochastic sequence using signals could be optimal is if all these deterministic sequences are equally optimal, hence yield the same payoff.

The argument in the proof requires more work because the infinite alternation is not feasible: only 4 strategies are available to the agent. What we show however is that the tit for tat strategy yields a payoff that is a weighted average of the three others, so it can only be optimal if the agent is indifferent between all 4 strategies. This requires $p = 1 - 1/(2(1+L))$.[24]

In the next section, we shall introduce another class of mental system in which mental states have some permanence. They will have the property that a player's mental state may remain unchanged with positive probability, *whatever signal received last period*. We will show that under some conditions, cooperation can then be achieved with strict incentives.

3. Resetting the relationship to cooperation

A central issue in relationships where monitoring is private is ensuring that players have incentives to trigger punishments. When monitoring is public, all players see the same signal, so there is no problem in coordinating a punishment phase: it is common knowledge what other players know. The issue in private monitoring is as in our example: when a player gets a bad signal, it may be that he is the only one having received it and that the other still plans on cooperating, making it a nontrivial decision for that player decide whether to begin a punishment phase or ignore the bad signal.

The lack of common knowledge among players results in a second key issue – how do players get back to cooperation once a punishment has been triggered. We finessed the second issue in the analysis above by assuming the public signal z that facilitated recoordination after a punishment period. A public signal z is a relatively simple way for the players to recoordinate, but as we emphasized, not necessary. We demonstrate next how players can coordinate a move back to cooperation in the absence of a public signal.[25]

Modified Tit-for-Tat We consider a stochastic modification of the Forgiving mental system above. Suppose the transitions from N to U are modified so that a bad signal only causes a transition to U with probability h. Also suppose that the transitions from U to N are modified so that a good signal triggers a transition back to N with probability k. Finally suppose that before each period starts, each player independently "forgets" with probability q and moves back to state N. This "forgetting" is similar to the resetting probability q in previous section, but differs in that the players' resetting is *independent*.

Transitions are summarized in Fig. 5.

For some configurations of parameters h and k, the strategy σ^* which plays C in N and D in U will be an equilibrium for a broad set of values of p and L, demonstrating that robust cooperation can be achieved without public signals.

To analyze this transition, we follow Section 2.4. Each strategy profile induces long-run probabilities over state pairs. The valuations and the bounds $\underline{L}$ and $\bar{L}$ can be expressed as a function of these long-run probabilities in exactly the same way (see Eqs. (2.1) and (2.2)), and the same conclusion holds: deterring deviations to σ^C and σ^D requires $\underline{L} < L < \bar{L}$.

One difference with Section 2.4 however is that the long run probabilities take different values (they are functions of p, k, h and q), and they are more difficult to compute. A second difference is that these two bounds alone are no longer sufficient to characterize equilibrium.

We state here our main result, the proof of which is in Appendix A:

Proposition 3. *$\underline{L} < \bar{L}$ if and only if $k > h > 0$. Furthermore, there is a range of values of h, k, q, p and L for which σ^* is an equilibrium.*

A corollary of Proposition 3 is that if $k = 1$ and $h = 1$ as in the Forgiving mental system (Fig. 3), then no cooperation can be achieved, even if independent forgetting ($q > 0$) is introduced, thus strengthening Proposition 2.

Another corollary of Proposition 3 is that if $k = 0$, so that resetting is only achieved through forgetting, then no robust cooperation can be achieved. Compared to the initial mental system (Fig. 1), players are similarly unresponsive to good signals. The difference is that forgetting is independent across players. This lack of coordination in resetting cooperation prevents cooperation altogether.

[24] In particular, when the latter condition holds, the player is indifferent between cooperation at all dates and defecting at all dates. Note however that the existence of such an equilibrium uses the symmetry assumption in the payoff structure. Asymmetries would destroy it.

[25] Compte and Postlewaite (2008) also consider recoordination utilizing asymmetric mental systems.

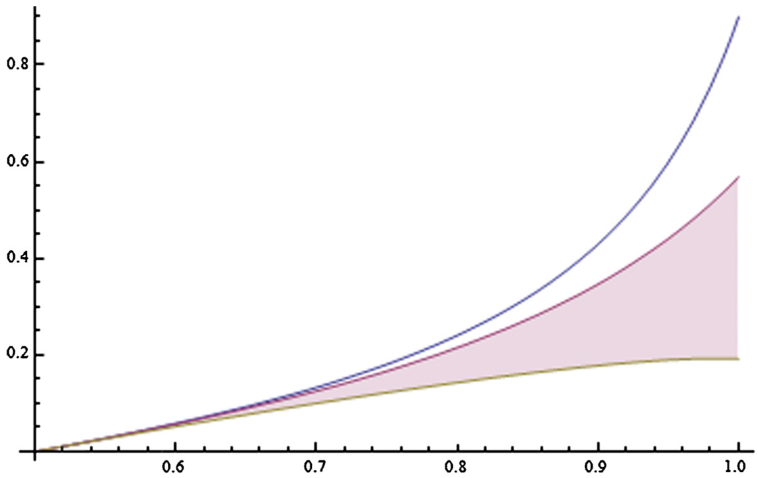

Fig. 6. Parameters (p, L) for which σ^* is an equilibrium.

The condition $k > h$ means that players must be responsive to good signals in a sufficiently strong way compared to their propensity to react to bad signals. Intuitively, the higher k and the smaller h, the easier it is to recoordinate on state NN from state UU. Indeed, once one player, say player 1, switches to N and cooperates, the other player will be likely to switch to N in the next period if k is large enough. If h is high as well however, then it is quite possible that the player who has initiated a move back to cooperation (i.e. player 1) returns to state U before player 2 switches to N, and recoordination does not occur.

In Fig. 6, the (p, L) for which σ^* is an equilibrium when $h = 0.1$, $k = 0.9$ and $q = 0.1$ are those in the shaded area.

The area between the upper curve and the lower curve shows the parameters (p, L) for which $L \in (\underline{L}, \bar{L})$, that is, for which a player does not gain by deviating to the strategy σ^C that plays C at all states, nor to the strategy σ^D that plays D at all states. The middle curve reflects the possibility that a player deviates to the strategy σ^{DC} that plays C at U and D at N. To ensure that σ^{DC} is not a profitable deviation, an additional upper constraint must hold, and this constraint is tighter than $\bar{L}$ for some parameters (h, k, q). Intuitively, σ^{DC} is a "manipulative" strategy in which a player cooperates to induce the other cooperate, and then defects, hoping that the other would not react too quickly: this strategy may be profitable precisely when the other is quickly forgiving (k high), and slow to react to bad signals (h small).

4. Further discussion

Evolution of mental systems. 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. 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, or reduce the set of parameters under which cooperation is possible. The question of evolution is discussed in more detail in Compte and Postlewaite (2009).

Extensions of the model.[26] We have focused on our leading example, which has two mental states. Compte and Postlewaite (2009) show how the basic insights carry over to mental systems with more mental states. That paper also investigates extensions of the basic model that include implicit communication.

Direct utility from being in a mental state. 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 transitions between states (going from upset to normal, for example).

Cooperation in larger groups. The basic structure and ideas in this paper can be extended to the case of many agents who are randomly matched. As is intuitive, the range of parameters for which cooperation is possible is smaller than in the two-person case because there is a longer time between a player's first defection and when he first meets opponents who do not cooperate as a result of his defection.[27]

Social norms. We have restricted attention to play in a prisoner's dilemma game to focus attention on the central ingredients of our model. It is straightforward to extend the basic idea to more general games, including asymmetric games. There may exist a "norm" that prescribes acceptable behavior for a wide variety of problems, with agents receiving noisy signals

[26] These extensions are discussed in more detail in Compte and Postlewaite (2009).

[27] There is a literature that analyzes the possibility of cooperation when players are repeatedly randomly matched (Kandori, 1992; Okuno-Fujiwara and Postlewaite, 1995), but in these models, some public information is or becomes available to players.

about whether their partner has followed the norm or not. Two-state mental systems will allow support of the norm in a manner similar to the cooperation that is possible in the model we analyze in this paper. Agents will follow a norm's prescriptions when they are in the "normal" state, and behave in their own self interest following observations that suggest their partner has violated the norm.

4.1. Related literature

Although we have emphasized the difficulty in supporting cooperation when signals are private, there are monitoring structures for which cooperation is relatively easy to sustain. This can be the case when 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 (2002) then 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 doesn't matter, because each can predict very accurately what other players will next do.[28] This is in sharp contrast to our example. First, the signals that players get are not helpful in predicting the signal received by the other player, and second, however accurate signals are, there are times (in state U for example) when a player may not be able to accurately predict what his opponent will do.[29]

As mentioned in the introduction, one branch of the repeated game literature aims at taking into account the complexity of strategies into account, assuming players use finite automata to implement their strategies.[30] The strategies that we consider can be represented as automata, yet we differ from this literature in several respects. First, in this literature players choose both the transition function and the mapping from states to actions, taking fixed only the number of states available given the automaton's size. In contrast, we take players' transition functions as fixed with players' choices being only the mapping from states to actions.[31] Second, to our knowledge, this literature does not consider games with private monitoring. Third, this literature used automata primarily as a tool to capture the complexity of a given strategy, while we emphasize the restriction on the *set* of strategies as a modelling device to capture the players' limited ability to tailor their behavior to the underlying parameters of the game. Fourth, our modeling strategy takes more seriously mental systems as being a plausible, if crude, model of the process by which players may interact, thereby shaping the restriction that we consider.

The model we study reduces to a dynamic game of a particular kind in which the underlying state is the profile of mental states.[32] Most of the literature in dynamic games assumes that in each period, there is a state that is known to both players,[33] while in our model the state is partially known: each player only knows his own mental state.

Following Mobius (2001), there is an active literature on favor/gift exchange between two players. The payoff structure of the stage game is analogous to the gift exchange model that we consider, and in some of these models players may receive relevant private information (for example about the opportunity to do favors as in Abdulkadiroglu and Bagwell, 2007).[34] These papers, however, assume public observations (i.e. whether a favor has been made) that allow players to coordinate future play.

Finally, in a recent paper, Romero (2010) provides an interesting example which, at least at a formal level, bears some resemblance to ours: there is a restriction to a limited set of automata that each individual may consider using (hence a limited set of strategies), and one of these automata (Win Stay Lose Shift) is an equilibrium in this restricted class.[35] This strategy facilitates recoordination when moves are simultaneous, but does poorly when players move in sequence: it would typically generate long sequences of miscoordinations after a bad signal.[36]

Appendix A

Proof of Proposition 1. (i) In what follows, we define $\mu = (1-q)/q$. Recall that ϕ_j^D denotes the long run probability that player 2 is in state j when 1 defects at all states and 2 plays σ_2^*. We have: $\phi_N^D = q + (1-q)(1-p)\phi_N^D$, implying that

[28] When one moves away from almost public signal monitoring, beliefs about each other's past histories may begin to matter, and even checking that a candidate strategy profile is an equilibrium becomes difficult, even when the strategy can be represented as an automaton with a small number of states. Phelan and Skrzypacz (2006) address this issue. See also Kandori and Obara (2007) for a related treatment.

[29] This is because even as p gets close to 1, the probability $\Pr(s_2 = U \mid s_1 = U) = \phi_{UU}^{\sigma}/(\phi_{UU}^{\sigma} + \phi_{UN}^{\sigma})$ remains bounded away from 0 and 1.

[30] The complexity of a player's strategy is defined to be the minimal size of a machine that can implement that strategy. See, e.g., Abreu and Rubinstein (1988) and Ben-Porath (1993).

[31] There has also been work in single-person decision making problems that is analogous to the papers using automata to capture complexity costs. See Wilson (2004) and Cover and Hellman (1970) for such models of single-person decision problems and Monte (2007) for a strategic treatment of such models. While we take agents' transition functions as fixed, the focus of this literature is on characterizing the optimal transition rule.

[32] We thank Eilon Solan for this observation. Dynamic games have the property that the structure of the continuation game varies with some underlying state. The dynamic nature of the game would be reinforced if we attached payoffs to being in a particular mental state.

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

[34] Hauser and Hopenhayn (2008) analyze a similar model in continuous time.

[35] Win Stay Lose Shift is an automaton that plays C after $C\overline{y}$ or $D\underline{y}$, and D otherwise.

[36] After an initial bad signal $\underline{y}$ (unfortunately) received by player 1, say, the most likely sequence of signals would lead to persistent miscoordination:

player 1's play and signals	C	$\underline{y}$	D	$\underline{y}$	C	$\underline{y}$	D	$\overline{y}$	D	$\underline{y}$	C	$\underline{y}$	D
player 2's play and signals		C	$\underline{y}$	D	$\overline{y}$	D	$\underline{y}$	C	$\underline{y}$	D	$\overline{y}$	D	$\underline{y}$

$$\phi_N^D = \frac{q}{1-(1-q)(1-p)} = \frac{1}{1+\mu p}$$

Similarly we have $\phi_N^C = q + (1-q)p\phi_N^C$, implying that

$$\phi_N^C = \frac{q}{1-(1-q)p} = \frac{1}{1+\mu(1-p)}$$

Recall that ϕ_{ij} denote the long-run probability that player 1 is in state $i \in \{U, N\}$ while player 2 is in state j, under the candidate equilibrium σ^*. As we already explained, we have $\phi_{NN} = q + (1-q)p^2\phi_{NN}$ implying that

$$\phi_{NN} = \frac{q}{1-(1-q)p^2} = \frac{1}{1+\mu(1-p^2)}$$

Next, we similarly have $\phi_{UN} = (1-q)(1-p)\phi_{UN} + (1-q)p(1-p)\phi_{NN}$; implying that

$$\phi_{UN} = \frac{\mu p(1-p)}{1+\mu p}\phi_{NN}$$

Using $\bar{L} = \frac{\phi_N}{\phi_N^D} - 1$ and $\underline{L} = \frac{1-\phi_N}{1-\phi_N^C} - 1$, where $\phi_N = \phi_{UN} + \phi_{NN}$, one gets:

$$\bar{L} = \mu(2p-1)\phi_{NN} \text{ and } \underline{L} = \frac{p}{1+p\mu}\bar{L}$$

This shows that $\underline{L} < \bar{L}$. Besides, since ϕ_{NN} increases with p, and since $\frac{p}{1+p\mu}$ increases with p, both $\bar{L}$ and $\underline{L}$ are increasing functions of p.

(ii) To prove that $\sigma^* = (\sigma_1^*, \sigma_2^*)$ is an equilibrium, we need to check that the deviation to the strategy $\widetilde{\sigma}$ that plays D in N and C in U is not profitable. Call $\widetilde{\phi}$ the long run distribution over state profiles induced by $(\widetilde{\sigma}, \sigma_2^*)$. We have:

$$v_1(\widetilde{\sigma}, \sigma_2^*) = (1+L)\widetilde{\phi}_{NN} + \widetilde{\phi}_{UN} - L\widetilde{\phi}_{UU}.$$

We wish to compare the gain $v_1(\widetilde{\sigma}, \sigma_2^*)$ from following $\widetilde{\sigma}$ to the gain $v_1(\sigma^C, \sigma_2^*)$ from always cooperating. Since $v_1(\sigma^C, \sigma_2^*) = \phi_N^C - L\phi_U^C$, σ^C is better whenever (letting $\widetilde{\phi}_N = \widetilde{\phi}_{NN} + \widetilde{\phi}_{UN}$):

$$(\widetilde{\phi}_{NN} - \widetilde{\phi}_{UU} + \phi_U^C)L < \phi_N^C - \widetilde{\phi}_N$$

We show below that for any $L < \bar{L}$ the above inequality holds, implying that $\widetilde{\sigma}$ is not a profitable deviation when $L < \bar{L}$.

To do that, we define $L^* = (\phi_N^C - \widetilde{\phi}_N)/(\widetilde{\phi}_{NN} - \widetilde{\phi}_{UU} + \phi_U^C)$ and compute the long run probabilities $\widetilde{\phi}$. These probabilities satisfy:

$$\widetilde{\phi}_{NN} = q + (1-q)p(1-p)\widetilde{\phi}_{NN} = \frac{1}{1+\mu(1-p(1-p))}$$

$$\widetilde{\phi}_{NU} = (1-q)[(1-p)\widetilde{\phi}_{NU} + p^2\widetilde{\phi}_{NN}] = \frac{\mu p^2}{1+\mu p}\widetilde{\phi}_{NN}$$

$$\widetilde{\phi}_{UN} = (1-q)[p\widetilde{\phi}_{UN} + (1-p)^2\widetilde{\phi}_{NN}] = \frac{\mu(1-p)^2}{1+\mu(1-p)}\widetilde{\phi}_{NN}$$

$$\widetilde{\phi}_{UU} = 1 - \widetilde{\phi}_{NN} - \widetilde{\phi}_{NU} - \widetilde{\phi}_{UN}$$

Using $1/\phi_N^C = 1 + \mu(1-p)$ and $1/\phi_{NN} = 1 + \mu(1-p^2)$, one easily obtains:

$$\frac{\phi_N^C - \widetilde{\phi}_N}{\phi_N^C\widetilde{\phi}_{NN}} = \mu(2p-1) \text{ and} \tag{A.1}$$

$$\frac{\widetilde{\phi}_{NN} + \widetilde{\phi}_{NU}}{\phi_N^C\widetilde{\phi}_{NN}} - \frac{1}{\phi_{NN}} = \frac{\mu p}{1+\mu p}(2p-1) \tag{A.2}$$

Next we use $\widetilde{\phi}_{NN} - \widetilde{\phi}_{UU} + \phi_U^C = \widetilde{\phi}_N - \phi_N^C + \widetilde{\phi}_{NN} + \widetilde{\phi}_{NU}$, and Eqs. (A.1) and (A.2) to write:

$$\begin{aligned}\frac{\mu(2p-1)}{L^*} - \frac{\mu(2p-1)}{\bar{L}} &= \mu(2p-1)\left[\frac{\widetilde{\phi}_{NN} + \widetilde{\phi}_{NU}}{\phi_N^C - \widetilde{\phi}_N} - 1\right] - \frac{1}{\phi_{NN}} \\ &= \frac{\widetilde{\phi}_{NN} + \widetilde{\phi}_{NU}}{\phi_N^C\widetilde{\phi}_{NN}} - \frac{1}{\phi_{NN}} - \mu(2p-1) \\ &= -\frac{1-p+\mu p}{1+\mu p}\mu(2p-1) < 0. \qquad \square\end{aligned}$$

Proof of Proposition 2. For each player i, the feasible strategies are σ_i^C (cooperate in both states), σ_i^D (defect in both states), σ_i^* (cooperate in N and defect in U – this induces Tit-for-Tat), and $\widehat{\sigma}_i$ (the manipulative strategy that cooperates in U and defects in N). The value for player 1 from following σ_i when player 2 follows σ_2 is denoted $v(\sigma_1, \sigma_2)$. We will show that $v(\sigma_1^*, \sigma_2^*)$ can be expressed as an average of the values $v(\sigma, \sigma_2^*)$ for $\sigma \in \{\sigma_1^C, \sigma_1^D, \widehat{\sigma}_1\}$. From this we shall conclude that (σ_1^*, σ_2^*) may only be an equilibrium if all strategies yield the same payoff.

In a given period, player 1's expected payoff only depends on his current action a and his previous action a'. Define this payoff by $w^*_{a'a}$. There are 4 such values. For example, $w^*_{CC} = p + (1-p)(-L)$ and $w^*_{DD} = (1-p)(1+L)$. Any strategy σ by player 1 generates a weighted average of these payoffs, where the weights correspond to the long-run frequencies of each pair $a'a$. Denote by $\alpha^\sigma_{a'a}$ these weights when player 1 follows σ. For player 1, the value $v(\sigma, \sigma_2^*)$ from following σ can thus be written:

$$v(\sigma, \sigma_2^*) = \sum_{a',a} \alpha^\sigma_{a'a} w^*_{a'a} \tag{A.3}$$

When player 1 follows σ_1^C, only α^σ_{CC} is positive, so $v(\sigma_1^C, \sigma_2^*) = w^*_{CC}$. Similarly, $v(\sigma_1^D, \sigma_2^*) = w^*_{DD}$. When player 1 follows σ_1^* or $\widehat{\sigma}_1$, any CD must eventually be followed by a DC, so in either case we must have: $\alpha^\sigma_{CD} = \alpha^\sigma_{DC}$. Defining $w^* \equiv \frac{1}{2}(w^*_{CD} + w^*_{DC})$, (A.3) may be rewritten as:

$$v(\sigma, \sigma_2^*) = \alpha^\sigma_{CC} w^*_{CC} + \alpha^\sigma_{DD} w^*_{DD} + (1 - \alpha^\sigma_{CC} - \alpha^\sigma_{DD}) w^* \tag{A.4}$$

For the particular payoff structure assumed, one can check that $w^* = 1/2$.[37] We thus have:

$$\max w^*_{CC}, w^*_{DD} \geq (w^*_{CC} + w^*_{DD})/2 = 1/2 = w^* \tag{A.5}$$

which immediately implies that (σ_1^*, σ_2^*) may only be an equilibrium if (A.5) holds with an equality, requiring $p = 1 - 1/(2(1+L))$. But the result holds for more general payoff structures. One may use (A.4) to write:

$$v(\sigma_1^*, \sigma_2^*) = \rho_C v(\sigma_1^C, \sigma_2^*) + \rho_D v(\sigma_1^D, \sigma_2^*) + (1 - \rho_C - \rho_D) v(\widehat{\sigma}_1, \sigma_2^*)$$
$$\text{with } \rho_a = \frac{\alpha^{\sigma_1^*}_{a,a}(1 - \alpha^{\widehat{\sigma}_1}_{a'a'}) - \alpha^{\widehat{\sigma}_1}_{a,a}(1 - \alpha^{\sigma_1^*}_{a'a'})}{1 - \alpha^{\widehat{\sigma}_1}_{aa} - \alpha^{\widehat{\sigma}_1}_{a'a'}} \text{ for } a = C, D \text{ and } a' \neq a$$

The coefficients ρ_C, ρ_D and $1 - \rho_C - \rho_D$ are all positive because when $p > 1/2$, we have[38]:

$$\alpha^{\sigma_1^*}_{CC} > \alpha^{\widehat{\sigma}_1}_{CC} \text{ and } \alpha^{\sigma_1^*}_{DD} > \alpha^{\widehat{\sigma}_1}_{DD}$$

The strategy σ_1^* may thus only be optimal if all strategies yield the same payoffs.

To complete the proof, we need to check that no other profile involving cooperation can be an equilibrium. It is obvious that σ^C cannot be part of an equilibrium. Assume now that player 2 plays $\widehat{\sigma}_2$, and define as above the payoffs $\widehat{w}_{a'a}$ obtained by player 1. It is easy to check that $\widehat{w}_{DD} = p(1+L)$ is the largest such payoff, hence that σ^D is the best response. □

Proof of Proposition 3. Proposition 3 is an immediate corollary of a stronger result which we now state, and which applies to any mental system where transitions depend only on the signal received. Any mental system induces probabilities that a given player will switch from one state to the other, as a function of the action (C or D) played by the other player. For transitions from N to U, we let:

$$p_C = \Pr(N \to U \mid C) \text{ and } p_D = \Pr(N \to U \mid D)$$

and for transitions from U to N:

$$q_C = \Pr(U \to N \mid C) \text{ and } q_D = \Pr(U \to N \mid D).$$

The following proposition holds:

Proposition 4. *Assume $p_D > p_C$ and $q_C > q_D$. Then $\underline{L} < \bar{L}$ if and only if $p_C + q_C > p_D + q_D$.*

When $p_D > p_C$ and $q_C > q_D$, the mental system has a Tit-for-Tat flavor: a player tends to become upset when his opponent defects, and forgiving when his opponent cooperates. Proposition 4 says that a necessary condition for such a mental system to support cooperation is that forgiveness induced by cooperation $(q_C - q_D)$ is stronger than the deterioration

[37] We have $w^*_{CD} = p(1+L)$, $w^*_{DC} = 1 - p - pL$, thus $w^* = 1/2$. We also have $w^*_{CC} = p + (1-p)(-L)$ and $w^*_{DD} = (1-p)(1+L)$ so $(w^*_{CC} + w^*_{DD})/2 = 1/2 = w^*$.

[38] Intuitively, the manipulative strategy induces more frequent switches. It can actually be checked that $\alpha^{\widehat{\sigma}}_{a',a} = 1/4$ for all $a'a$, and that $\alpha^{\sigma^*}_{CC} = \alpha^{\sigma^*}_{DD} > 1/4$.

induced by defection $(p_D - p_C)$. Proposition 3 is a corollary of Proposition 4 because given our assumption about the mental system, we have:

$$p_C = (1-q)(1-p)h \text{ and } p_D = (1-q)ph$$
$$q_C = q + (1-q)pk \text{ and } q_D = q + (1-q)(1-p)k$$

so when $p > 1/2$ and $0 < h < k$, all the conditions of Proposition 4 are satisfied. □

Proof. We need to compute ϕ_N^D, ϕ_N^C and ϕ_N and check whether and when $\Delta \equiv \frac{\phi_N}{1-\phi_N} - \frac{\phi_N^D}{1-\phi_N^C}$ is positive. The long-run probabilities ϕ_N^C and ϕ_N^D are easy to compute. We have $\phi_N^C = (1-p_C)\phi_N^C + q_C(1-\phi_N^C)$, which yields:

$$\phi_N^C = \frac{q_C}{q_C + p_C}.$$

Similarly, we have $\phi_N^D = (1-p_D)\phi_N^D + q_D(1-\phi_N^D)$, implying that:

$$\phi_N^D = \frac{q_D}{q_D + p_D}.$$

To compute $\phi_N = \phi_{NN} + \phi_{UN}$, we have to find a probability vector $\phi = (\phi_{NN}, \phi_{NU}, \phi_{UN}, \phi_{UU})$ which is fixed point of:

$$\phi = \phi.M \text{ where } M = \begin{pmatrix} (1-p_C)^2 & (1-p_C)p_C & p_C(1-p_C) & (p_C)^2 \\ (1-p_D)q_C & (1-p_D)(1-q_C) & p_Dq_C & p_D(1-q_C) \\ q_C(1-p_D) & p_Dq_C & (1-p_D)(1-q_C) & p_D(1-q_C) \\ (q_D)^2 & q_D(1-q_D) & (1-q_D)q_D & (1-q_D)^2 \end{pmatrix}. \tag{A.6}$$

Define $X = 2 - p_C - q_D$ and $Y = (p_D - p_C)q_C + p_D(q_C - q_D)$. Note that both X and Y are positive under the assumption of the Proposition. One can check that the vector $y = (y_{NN}, y_{NU}, y_{UN}, y_{UU})$ defined by:

$$y_{NN} = q_D(q_CX - Y)$$
$$y_{NU} = y_{UN} = q_Dp_CX$$
$$y_{UU} = p_C(p_DX - Y)$$

solves (A.6), so $\phi = \alpha y$ for some normalizing constant α. Thus we have:

$$\frac{\phi_N}{1-\phi_N} = \frac{y_{NN} + y_{NU}}{y_{UN} + y_{UU}} = \frac{q_D}{p_C}\frac{(p_C + q_C)X - Y}{(p_D + q_D)X - Y} \text{ and}$$
$$\frac{\phi_N^D}{1-\phi_N^C} = \frac{q_D}{p_C}\frac{p_C + q_C}{p_D + q_D}$$

Since X and Y are positive, the difference Δ is positive if and only if $p_C + q_C > p_D + q_D$. □

References

Abdulkadiroglu, Atila, Bagwell, Kyle, 2007. Trust, reciprocity and favors in cooperative relationships. Mimeo, Columbia University.

Abreu, D., 1986. Extremal equilibria of oligopolistic supergames. J. Econ. Theory 39, 191–225.

Abreu, D., Rubinstein, A., 1988. The structure of Nash equilibrium in repeated games with finite automata. Econometrica 56 (6), 1259–1281.

Altman, E., Avrachenkov, K., Marquez, R., Miller, G., 2005. Zero-sum constrained stochastic games with independent state processes. Math. Methods Operations Res. 62, 375–386.

Aumann, R., 1981. Survey of repeated games. In: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Bibliograph. Inst., Zurich, pp. 11–42.

Bagwell, K., 1995. Commitment and observability in games. Games Econ. Behav. 8, 271–280.

Ben-Porath, E., 1990. The complexity of computing a best response automaton in repeated games with mixed strategies. Games Econ. Behav. 2, 1–12.

Ben-Porath, E., 1993. Repeated games with finite automata. J. Econ. Theory 59, 17–32.

Cohen, D., Nisbett, R.E., Bowdle, B., Schwartz, N., 1996. Insult, aggression, and the southern culture of honor: an 'experimental ethnography'. J. Pers. Soc. Psychol. 70, 945–960.

Compte, O., 1998. Communication in repeated games with imperfect private monitoring. Econometrica 66, 597–626.

Compte, O., 2002. On sustaining cooperation without public observations. J. Econ. Theory 102, 106–150.

Compte, O., Postlewaite, A., 2008. Repeated relationships with limits on information processing. Mimeo, University of Pennsylvania.

Compte, O., Postlewaite, A., 2009. The extent of cooperation. Mimeo, University of Pennsylvania.

Compte, O., Postlewaite, A. 2012. Strategy restrictions and limited knowledge. Mimeo.

Compte, O., Postlewaite, A. 2013. Belief free equilibria. Mimeo, University of Pennsylvania.

Cover, T., Hellman, M., 1970. Learning with finite memory. Ann. Math. Stat. 41, 765–782.

Ely, J.C., Välimäki, J., 2002. A robust folk theorem for the Prisoner's dilemma. J. Econ. Theory 102, 84–105.

Gilboa, I., 1988. The complexity of computing best-response automata in repeated games. J. Econ. Theory 45, 342–352.

Kalai, E., Stanford, W., 1988. Finite rationality and interpersonal complexity in repeated games. Econometrica 56, 397–410.

Kalai, E., Samet, D., Stanford, W., 1988. A note on reactive equilibria in the discounted Prisoner's dilemma and associated games. Int. J. Game Theory 17, 177–186.
Hauser, C., Hopenhayn, H., 2008. Trading favors: optimal exchange and forgiveness. Mimeo, UCLA.
Kandori, M., 1992. Social norms and community enforcement. Rev. Econ. Stud. 59, 63–80.
Kandori, M., Matsushima, H., 1998. Private observation, communication and collusion. Econometrica 66, 627–652.
Kandori, M., Obara, I., 2007. Finite state equilibria in dynamic games. Mimeo.
Mailath, G., Morris, S., 2002. Repeated games with almost-public monitoring. J. Econ. Theory 102, 189–228.
Mobius, M., 2001. Trading favors. Mimeo, Harvard University.
Monte, D., 2007. Reputation and bounded memory: a theory of inertia and skepticism. Mimeo.
Neyman, A., 1985. Bounded complexity justifies cooperation in the finitely repeated Prisoner's dilemma. Econ. Letters 19, 227–229.
Neyman, A., 1998. Finitely repeated games with finite automata. Math. Oper. Res. 23, 513–552.
Okuno-Fujiwara, M., Postlewaite, A., 1995. Social norms in matching games. Games Econ. Behav. 9, 79–109.
Phelan, C., Skrzypacz, A. 2006. Private monitoring with infinite histories. Mimeo.
Piccione, M., 2002. The repeated Prisoner's dilemma with imperfect private monitoring. J. Econ. Theory 102, 70–83.
Romero, J., 2010. Bounded rationality in repeated games. Mimeo, Caltech.
Rubinstein, A., 1986. Finite automata play the repeated Prisoner's dilemma. J. Econ. Theory 39, 83–96.
Sekiguchi, T., 1997. Efficiency in repeated Prisoner's dilemma with private monitoring. J. Econ. Theory 76, 345–361.
Wilson, A., 2004. Bounded memory and biases in information processing. Mimeo, University of Chicago.