Discussion Paper No. 1200

Reciprocity and Cooperation in Repeated Coordination Games: The Blurry Belief Approach

by

Alvaro Sandroni

November 1997

Reciprocity and Cooperation in Repeated Coordination Games: The Blurry Belief Approach

Alvaro Sandroni*
J. L. Kellogg Graduate School of Management
MEDS Department
Northwestern University
2001 Sheridan Road, Evanston, IL 60208
sandroni@nwu.edu

November 2, 1997

Abstract

Two long lived players play a repeated coordination game. Players do not specify a single (and correct) probability to each event. They have a vague notion about the evolution of the play, called blurry beliefs, which guide their behavior. General conditions that ensure cooperation are investigated.

Key words: Repeated Games, Learning, Cooperation, Bounded Rationality, Equilibrium Selection.

JEL classification: D83

*I had the benefit of extensive conversations with Ehud Kalai, whose ideas were fundamental in realizing this project. I would also like to thank Eddie Dekel, George Mailath, and Juuso Valimaki for useful suggestions. This paper began in a conversation with Rann Smorodinsky who helped me on many occasions. All errors are mine.

1. Introduction

Consider a repeated coordination game with two long-lived players and two actions available to each player at each stage. Assume that there is a single outcome which yields the highest payoffs for both players. Call the action associated with this outcome "to cooperate." Call the alternative action "not to cooperate." A fundamental question in game theory is whether cooperation can be derived by the principle of rational behavior (see Aumann and Sorin (89)). The standard way of modeling rational behavior is to impose equilibrium behavior. It is assumed that players' beliefs are correct and that players take best responses according to these beliefs. These assumptions are very demanding because players are assumed to know (or behave as if they knew) their opponents' strategies. However, equilibrium behavior does not necessarily ensure cooperation. In fact, the Folk Theorem shows that there are a large multiplicity of equilibria in repeated games. In particular, both the cooperative outcome and highly inefficient outcomes remain equilibria. For example, players will (not) cooperate, in all periods, if they believe that the other player will (not) cooperate regardless of past outcomes.

Clearly, the outcome of the game will depend upon players' beliefs about the evolution of the game. Hence, it is important to develop a theory of belief formation. In this paper, I focus on the question of belief representation. I do not assume that players specify a single probability to each event nor do I assume that players behave as if they knew their opponents' strategies; rather, I assume that players have some "notions" about how play might evolve, know their own payoffs and take best responses according to these vague notions. For example, consider a player who come to the game with the following notion: "If I cooperate then my opponent will be more inclined to cooperate." Is there an optimal action for this player? How can this player's belief be explicitly modeled? How would this belief influence (and be influenced by) the evolution of the game?

The notions which guide players' behavior are modelled by a "*blurry belief*." A blurry belief is a class of well-specified beliefs. The restrictions defining this class are derived from a general principle which, in turn, defines the players' incomplete notions about how play might evolve. The core question is whether there exists an optimal strategy independent of the specification of the belief in the class. If a strategy is optimal for all beliefs in a class, then this strategy is optimal for players maintaining the notions that this class represents.

The following example shows the subtlety of this analysis. Assume that each player believes that if he cooperates, in the first period, then the opponent will cooperate thereafter. This class of beliefs seems too restrictive to represent the notion that "if I cooperate then my opponent will be more inclined to do the same." However, does acceptance of such a principle ensure cooperation? Apparently, a sufficiently patient player who holds this view optimally should cooperate, regardless of what he thinks might happen if he does not cooperate in the first period, because, if he does cooperate, then any loss he might incur in the first period is compensated by the maximum payoffs obtained afterwards. Surprisingly, there exists a belief in this class such that, for all discount factors, both players optimally do not cooperate and eventually learn that the other player does not cooperate. The example that illustrates this point is called the "waiting players." In this example, both players believe that cooperation can also be triggered if the opponent cooperates and optimally they do not cooperate because they wait for the other player to cooperate first. Eventually, each player realizes that the other player will not cooperate in the near future with arbitrarily high probability.

The waiting players example shows that even if the class of beliefs seems restrictive it may not ensure cooperation. Clearly, it is necessary to assume other kinds of restrictions on players' beliefs. I assume that a player's belief about the probability that his opponent will cooperate depends upon the last outcome and the probability that he had assigned to his opponent's cooperating in the previous period. This restriction imposes a recursive structure on players' beliefs.

A player believes in *positive influence* if he believes that if he cooperates in the current period then it is more probable that the other player will cooperate in the next period than in this period. This class of beliefs represents the notion that "if I cooperate then my opponent will be more inclined to cooperate." As should be expected from the waiting players example, arbitrarily patient players, who believe in positive influence, optimally may not cooperate.

A player who believes in *negative influence* believes that if he does not cooperate in the current period then it is less probable that the opponent will cooperate in the next period than in this period. This class of beliefs represents the notion that "if I do not cooperate then my opponent will be less inclined to cooperate." A player believes in *reciprocity* if he believes in both positive and negative influence.[1]

The main result is that sufficiently patient players who believe in reciprocity optimally cooperate.[2] So, cooperation is derived from the principle of reciprocity.

[1]See Fehr, Gächter, and Kirchsteiger (97) for experimental work on reciprocity.

[2]Some additional restrictions on players' beliefs, of a technical nature, are also assumed.

Reciprocity and positive influence are heuristics which a player may use to guide his behavior. They are represented by a blurry belief which is a class of well-specified beliefs consistent with each principle. There are well-specified beliefs consistent with positive influence that leads to cooperation. However, cooperation does not necessarily follows from positive influence because, as shown by the waiting players example, there are beliefs which are consistent with positive influence that do not lead to cooperation. Hence, even if players accept this principle they may not optimally cooperate because they may make other considerations.

The main result in this paper can be viewed as an anti-folk theorem because a unique outcome is obtained in an infinitely repeated game with patient players. The basic assumption is that players believe in reciprocity. This does not mean that players believe in reciprocity and no other thought occurs to them. Players optimally cooperate for all well-specified beliefs in the class representing reciprocity. Some of these well-specified beliefs may belong to other classes representing other principles. Hence, sufficiently patient players who believe in reciprocity optimally cooperate even if they also make other considerations.

Cooperation is a Nash equilibrium play in coordination games. If, as in the case of this paper, the outcome of interest is a Nash equilibrium play, then blurry beliefs generate learning schemes that lead to a Nash equilibrium. However, as opposed to many results in the rational learning literature, no compatibility conditions between the beliefs and the true play, such as absolute continuity, have been assumed (see Kalai and Lehrer (93a) and (93b)). Thus, convergence to Nash equilibrium occurs solely because of the restrictions imposed on the exogenous variables i.e., beliefs, discount factors, and stage game payoffs.[3]

Blurry beliefs may be used in any game. For example, what principles lead to cooperation in the repeated prisoners' dilemma? This is an open and interesting question. However, a blurry belief probably should not be interpreted outside a given context. The interpretation and consequences of a blurry belief depend on the particular game being played. In particular, there is no reason to assume that reciprocity, as defined for coordination games, also leads to cooperation in any

[3]In the fictitious play literature, no compatibility conditions between beliefs and best responses are assumed. However, players believe that their past actions do not affect the future actions of the other player. Thus, long-run strategic considerations are ignored. So, these models are inappropriate when there are few, long-lived players as, for example, in the case of a firm and a worker who play the same game repeatedly. The same is true in evolutionary models where equilibrium selection results have also been obtained (see Matsui and Rob (91), Kandori, Mailath and Rob (93), Young (93)).

other game. However, although different principles may arise in different games, they may all be described in the framework of the rational learning model.

2. The Repeated Coordination Game

There are two players I and II. Player $i \in \{I, II\}$ has two possible actions given by the set $\sum = \{c, d\}$. The payoff function $u_i : \sum \to \Re$ is given by the payoff matrix

$$\begin{bmatrix} (I, II) & c & d \\ c & (c^1, c^2) & (w^1, w^2) \\ d & (z^1, z^2) & (d^1, d^2) \end{bmatrix}$$

where $c^i > \max\{z^i, w^i, d^i\}$, $i \in \{I, II\}$. A player (does not) cooperate if he plays (d) c. The outcome (c, c) is called cooperative.

Let $\sum^t$ be the set of all t-histories, $0 \le t \le \infty$. Let $H = \bigcup\limits_{t \geqslant 0} \sum^t$ be the set of all finite histories. Let $\Im_0 \subset ...\Im_t \subset ... \subset \Im$ be the filtration on $\sum^\infty$ where $\Im_t$ is the σ-algebra generated by all t-histories, and $\Im$ is the σ-algebra generated by the algebra $\bigcup\limits_{t \geqslant 0} \Im_t$.

Let Ψ be the set of all functions $g : H \to [0, 1]$. A behavior strategy $f_i \in \Psi$ describes the probability that player $i \in \{I, II\}$ will cooperate conditional on each finite history. Let $f = (f_I, f_{II})$ be the true behavior strategy profile. A well-specified belief $f_{-i} \in \Psi$ describes the probability that player i believes that the other player will cooperate, conditional on each finite history.[4] Player i's blurry belief is a set of well-specified beliefs $\Delta_{-i} \subset \Psi$.

Given a strategy profile q, let μ_q be the probability measure over play paths associated with q.[5] Given the strategy profile $q^i = (q, f_{-i})$, player i's discounted expected payoff is

$$V_i(q^i) = E^{\mu_{q^i}} \left\{ \sum_{r=0}^{\infty} \{(\beta_i)^r u_i\} \right\},$$

where β_i, $0 < \beta_i < 1$, is player i's discount factor and $E^{\mu_{q^i}}$ is the expectation operator associated with μ_{q^i}. The behavior strategy f_i is a best response to f_{-i} if for every strategy profile $q^i = (q, f_{-i})$, $V_i(f^i) - V_i(q^i) \ge 0$. The behavior strategy f_i is a best response to Δ_{-i} if f_i is a best response to all $f_{-i} \in \Delta_{-i}$.

[4] See Kuhn (53) for the description of players' beliefs.

[5] See Kalai and Lehrer (93) for details on the construction of this probability measure.

3. Reciprocity and Cooperation

Definition 1. *Players learn to cooperate if there exists a set $\Omega \in \Im$ such that $\mu_f(\Omega) = 1$, and given $f_{-i} \in \Delta_{-i}$, for every $s \in \Omega$, $s = (h, ...)$, $h \in \Sigma^t$, and $\varepsilon > 0$, there exists a period $\bar{t}$ such that for all $t \geq \bar{t}$, $f_{-i}(h) \geqslant 1 - \varepsilon$ and $f_i(h) \geqslant 1 - \varepsilon$.*

Players learn to cooperate if eventually players' beliefs and the behavior strategies are arbitrarily close to the Nash equilibrium in which players always cooperate.

Let the blurry belief ξ^l_{-i} be defined by $f_{-i} \in \xi^l_{-i}$ if and only if $f_{-i}(h) = 1$ for all finite histories $h \in H$ such that player i cooperated in the first l periods.

Definition 2. *Player i believes that he can induce the other player to cooperate by cooperating in the first l periods if player i holds the blurry belief ξ^l_{-i}.*

That is, player i believes that he can induce the other player to cooperate by cooperating in the first l periods if he thinks that if he cooperates in the first l periods, then the other player will respond by cooperating, with probability one, thereafter.

Example 1, below, shows that players who believe that they can induce the other player to cooperate by cooperating in the first period do not necessarily learn to cooperate regardless of how patient they are.

Example 1. *The waiting players.*

The payoffs of stage game are described by the payoff matrix

$$\begin{bmatrix} (I, II) & c & d \\ c & (9,9) & (0,8) \\ d & (8,0) & (8,8) \end{bmatrix}.$$

Let k be a natural number such that $\left(\frac{3}{4}\right)^k k^2 \sum\limits_{t=k+1}^{\infty} \frac{1}{t^2} < \frac{11}{4}$.[6] Player i believes that the other player will cooperate with probability 0.25 at period 1. If player i cooperates, at period 1, then he believes that the other player will cooperate

[6] The existence of k follows from $\left(\frac{3}{4}\right)^k k^2 \underset{k \to \infty}{\longrightarrow} 0$ and $\sum\limits_{t=k+1}^{\infty} \frac{1}{t^2} \underset{k \to \infty}{\longrightarrow} 0$.

thereafter. If player i does not cooperate, at period 1, but the other player cooperates, at any period, then he believes that the other player will cooperate thereafter. If player i does not cooperate, at period 1, and the other player does not cooperate until period t, then he believes that the other player will cooperate at period $t+1$ with probability γ_t, where $\gamma_t = 0.25$ if $t \leq k$, and $\gamma_t = 1 - \left(\frac{t-1}{t}\right)^2$ if $t > k$. By definition, both players believe that they can induce the other player to cooperate by cooperating in the first period.

If a player cooperates, at period 1, then he gets an expected discounted payoff which is smaller than or equal to

$$\frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t \, 9.$$

A player who only cooperates after the other player cooperates obtains the expected discounted payoff

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \left(\prod_{r=1}^{t} (1-\gamma_r) 8 + \left(1 - \prod_{r=1}^{t} (1-\gamma_r)\right) 9 \right) =$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - \sum_{t=1}^{\infty} (\beta_i)^t \left(\prod_{r=1}^{t} (1-\gamma_r) \right) =$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - \sum_{t=1}^{k} (\beta_i)^t \left(\left(\frac{3}{4}\right)^t \right) - \left(\frac{3}{4}\right)^k k^2 \sum_{t=k+1}^{\infty} (\beta_i)^t \left(\frac{1}{t^2}\right) >$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - 3 - \left(\frac{3}{4}\right)^k k^2 \sum_{t=k+1}^{\infty} \frac{1}{t^2} > \frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t \, 9.$$

Hence, both players optimally do not cooperate in the first period. In all other periods, both players believe that the other player will cooperate with probability smaller than 8/9 as long as they have not yet observed the other player cooperating. Therefore, both players optimally do not cooperate before they observe the other player cooperating. The true play will be (d, d) in every period, and each player will eventually believe that the other player will cooperate with probability arbitrarily close to the true probability (zero).

In the waiting players example, if either player had cooperated at period 1 then both players would have cooperated thereafter. Hence, both players correctly believed they could get maximum payoffs after period 1. However, they optimally chose to wait for the other player to cooperate first. Thus, both players end up waiting forever and the payoffs obtained ex-post are much lower than the expected payoffs, although players' beliefs over short-run events eventually become accurate.

In a static coordination game, players optimally cooperate if they believe that the other will cooperate, and players optimally do not cooperate if they believe that the other will not cooperate. But, in a repeated coordination game, the waiting players example shows that players may not cooperate even if they believe that the cooperative outcome may eventually be played. In fact, arbitrarily patient players may not optimally cooperate even if they believe that they will induce the other player to (not) cooperate forever if they (do not) cooperate in the first periods. For example, assume that each player believes that the other player will cooperate, with probability one, in the first period. Player $i \in \{I, II\}$ believes that if he does not cooperate in the first period then the other player will not cooperate thereafter, and if he cooperates in the first period then his belief will be as in the waiting players example. Then, player i believes that if he cooperates in the first two periods then the other player will cooperate thereafter. However, it is straightforward to show that, regardless of their discount factor, the optimal play will be (c, c) at period 1 and (d, d) thereafter.

In the waiting player example, both players believe that if they cooperate in the first period then they will induce the other player to cooperate, but if they do not cooperate in the first period then their actions no longer have any influence on the other player's future actions. Analogously, in the example given above, players also believe that their influence over the other player's actions will disappear if they play c followed by d. So, I consider a class of "recurrent" beliefs in which players' beliefs about their potential influence on the other player's actions do not change so abruptly.

Given a well-specified belief f_{-i}, let x_t^i be a $\Im_{t-1}-$measurable function representing player i's subjective probability, conditional on all information available at period $t-1$, that his opponent will cooperate at period t. That is, given a play path $s = (h, ...), h \in \Sigma^{t-1}$, $x_t^i(s) = f_{-i}(h)$. Assume that x_t^i follows the rule $x_t^i = g^i(x_{t-1}^i, a, b)$ where $a \in \{c, d\}$ and $b \in \{c, d\}$ are the actions taken, at period $t-1$, by player i and the opponent, respectively. Hence, player i's belief about the probability that the opponent will cooperate depends upon the current outcome and the probability that player i assigned last period to cooperation on the part of the opponent.

Player i's belief is well-specified given x_1^i and the functions g^i. Player i's blurry beliefs may now be defined by restrictions on x_1^i and g^i.

Let e^i and r^i be the functions $e^i(x) = xg^i(x,c,c) + (1-x)g^i(x,c,d)$ and $r^i(x) = xg^i(x,d,c) + (1-x)g^i(x,d,d)$. Assume that after observing the outcome at period $t-1$, player i decides to cooperate at period t. Then, he believes that the opponent will cooperate at period $t+1$ with probability $e^i(x_t^i)$. Analogously, if, at period $t-1$, player i decides not to cooperate at period t, then he believes that the opponent will cooperate at period $t+1$ with probability $r^i(x_t^i)$.

Let $V^i(x)$ be the expected discounted payoff of player i if he decides always to cooperate, and if he believes that his opponent will cooperate with probability x in the current period.

Definition 3. *Player i's belief is regular if V^i is a smooth, non-decreasing, and concave function.*

The function V^i is non-decreasing if the expected discounted payoff associated with the strategy "always cooperate" does not decrease when player i's probability that the opponent will cooperate in the current period increases.

The function V^i is concave if the expected discounted payoff associated with the strategy "always cooperate" when player i believes that the opponent will cooperate with probability $\lambda\hat{x} + (1-\lambda)\bar{x}$ in the current period is not smaller than the linear combination (using λ as weight) of the expected discounted payoffs associated with the same strategy when player i believes that in the current period the opponent will cooperate with probabilities $\hat{x}$ and $\bar{x}$, respectively. Lemma 1, below, provides sufficient conditions under which player i's beliefs are regular.

Lemma 1. *Assume that $g^i(x,c,c)$ and $g^i(x,c,d)$ are smooth functions of x. If $g^i(x,c,c) \geqslant g^i(x,c,d)$ and $g^i(x,c,c)$, $g^i(x,c,d)$ are non-decreasing functions of x then V^i is a smooth and non-decreasing function of x.*[7] *If, in addition, e^i and $g^i(x,c,d)$ are concave functions of x, then V^i is also a concave function of x.*[8]

[7] The assumption that $g^i(x,c,c) \geqslant g^i(x,c,d)$ means that player i's belief about the probability that the other player will cooperate is not smaller when the cooperative outcome is observed than when player i cooperates, but the other player doesn't.

[8] The assumption that e^i and $g^i(x,c,d)$ are concave functions of x are of a technical nature and the main result (proposition 1) is probably true without it, but a formal proof does not seem to follow from the techniques developed in this paper to solve the optimization problems.

Proof - See Appendix.

Definition 4. *Player i believes in positive influence if $e^i(x) > x$ whenever $x \in [0, 1)$, and $e^i(1) = 1$.*

Player i believes in positive influence if player i expects to make the opponent more inclined to cooperate by cooperating himself, in the sense that if player i cooperates then he believes that the opponent will be more likely to cooperate in the next period than in the current period. That is, at the beginning of period t, before the outcome is realized, player i believes that the other player will cooperate with probability x_t^i. Player i knows that whatever action he takes will not have any influence over the other player's decision in the current period, but knows his action might influence the other player's action in the next period. If player i believes in positive influence, then player i believes that by cooperating he will increase the chances that the other player will cooperate compared to the current odds.

If players believe in reciprocity and $g^i(x, c, c) \geqslant g^i(x, c, d)$ then $g^i(x, c, c) \geqslant x$. Hence, after observing the other players' action, players will become more confident that the other player will cooperate if they observe the cooperative outcome. However, if player i cooperates and the opponent doesn't then player i may or may not become more confident that the other player will cooperate.

Definition 5. *Player i believes in negative influence if $r^i(x) < x$ whenever $x \in (0, 1]$, and $r^i(0) = 0$.*

Player i believes in negative influence if player i expects to make the opponent less inclined to cooperate by not cooperating himself, in the sense that if player i does not cooperate then he believes it is less likely that the opponent will cooperate in the next period than in the current period. Negative and positive influence are, of course, perfectly symmetric restrictions on players' beliefs.

Definition 6. *Player i believes in reciprocity if player i believes in both positive and negative influence.*

Relaxing the concavity assumption would make the results much more attractive because it is in the spirit of the blurry belief approach to consider classes of beliefs that are as general as possible and, more importantly, to make restrictions on beliefs that follows from principles that are easily interpretable. Unfortunately, I do not know how to dispose of this assumption.

Lemma 2. *If player i believes in positive influence, then for every $\varepsilon > 0$ there exists $\bar{\beta}$ such that if $\beta_i \geqslant \bar{\beta}$ then $V^i(0) \geqslant \frac{c^i - \varepsilon}{1-\beta_i}$.*

Proof - See Appendix.

Lemma 2 shows that if player i believes in positive influence and player i is sufficiently patient, then player i expects to obtain high expected discounted payoffs by cooperating in every period even in the extreme case that player i is sure that the opponent will not cooperate in the current period.

I now show the main result of this paper. Proposition 1, below, shows that cooperation can be derived from the principle of reciprocity, in the sense that patient players, whose beliefs are regular, learn to cooperate if they believe in reciprocity.

Proposition 1. *Assume that players' beliefs are regular. If players believe in reciprocity and players' discount factors are sufficiently high, then players learn to cooperate. Moreover, the play will be cooperative in all periods.*[9]

Proof - See Appendix.

Example 2. *Patient players who believe in positive influence, but not in negative influence, optimally may not cooperate.*

Consider the same payoff matrix as in the waiting players' example. Assume that $g^i(x, c, c) = g^i(x, c, d) = g^i(x, d, c) = 1$, and $g^i(x, d, d) = x_1^i = 0.25$. That is, each player believes that the other player will cooperate with probability 0.25 in the first period, and will continue to cooperate with probability 0.25 if (d, d) is observed last period. Otherwise, the other player will cooperate with probability one. If player i cooperates in the first period he obtains a discounted expected payoff equal to

$$\frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t \, 9.$$

[9]Proposition 1 also holds if instead of assuming that V^i is concave it is assumed that player $i \in \{I, II\}$ believes that if he does not cooperate then the probability of cooperation on the part of the opponent will not increase regardless of the action of the opponent i.e., $g^i(x, d, c) \leqslant x$ and $g^i(x, d, d) \leqslant x$.

A player who only cooperates after the other player cooperates obtains an expected discounted payoff equivalent equal to

$$8+\sum_{t=1}^{\infty}(\beta_i)^t\left((0.75)^t 8+\left(1-(0.75)^t\right)9\right)=$$

$$8+\sum_{t=1}^{\infty}(\beta_i)^t\,9-\sum_{t=1}^{\infty}(0.75\beta_i)^t>\frac{9}{4}+\sum_{t=1}^{\infty}(\beta_i)^t\,9.$$

Hence, both players do not cooperate in the first period. In the second period, players will face the same maximization problem as in the first period. By induction, it can be shown that the play will be (d,d) in every period.[10]

Patient players who believe in negative influence, but not in positive influence, may not optimally cooperate. For example, assume that $x_1^i=g^i(x,c,c)=g^i(x,c,d)=g^i(x,d,c)=g^i(x,d,d)=0$. Then, each player believes that the other player will not cooperate in every period, and optimally will not cooperate in every period.

Players who believe in reciprocity may not cooperate if they are not sufficiently patient. For example, assume that the payoffs are the same as in the waiting players' example. Let the beliefs be given by $x_1^i=g^i(x,c,c)=g^i(x,c,d)=1$ and $x_1^i=g^i(x,d,c)=g^i(x,d,d)=0$. Let the discount factor of each player be 0.25. If a player cooperates, at period 1, he gets an expected discounted payoff smaller than $\sum\limits_{t=1}^{\infty}(0.25)^t\,9<8$. If a player does not cooperate, then he gets a payoff greater than 8. By induction, it is easy to show that both players will not cooperate in every period.

It is interesting to consider an example in which the players' optimization problem can be solved directly. Let the payoffs be given by the matrix

$$\begin{bmatrix}(I,II) & c & d\\ c & (c^1,c^2) & (0,0)\\ d & (1,1) & (1,1)\end{bmatrix}$$

where $c^1>1$ and $c^2>1$.

Players' beliefs are given by $x_1^i=0$; $g^i(x,c,c)=g^i(x,c,d)=\alpha+(1-\alpha)x$; $g^i(x,d,c)=g^i(x,d,d)=\zeta x$; where $\alpha>0$ and $0\leqslant\zeta<1$.

[10]In this example, players' predictions are not accurate because players will not cooperate but they believe that the other player will cooperate with probability 0.25.

Let $e_t^i(x)$ be player i's subjective probability that the other player will cooperate t periods ahead. If player i cooperates in all periods, then $e_{t+1}^i(x) = e_t^i(\alpha+(1-\alpha)x)$ is the subjective probability that the other players will cooperate $t+1$ periods ahead. By definition, $e_1^i(x) = e^i(x) = \alpha+(1-\alpha)x$ is a linear function of x. If $e_t^i(x)$ is a linear function, then $e_{t+1}^i(x)$ is a linear function. Hence, by induction, if player i cooperates in all periods then $e_t^i(x)$ is a linear function of x for all t. The expected payoff of cooperating in all periods, $V^i(x)$, is equal to $c^i \sum\limits_{t=0}^{\infty} (\beta_i)^t e_t^i(x)$. Thus, $V^i(x)$ is a linear function of x. By definition, $V^i(x) = xc^i + \beta_i V^i(\alpha + (1-\alpha)x)$. With some algebra, it follows that $V^i(x) = \frac{c^i\beta_i\alpha}{(1-\beta_i)(1-\beta_i(1-\alpha))} + \frac{c^i x}{1-\beta_i(1-\alpha)}$. Therefore, the expected discounted payoff of playing d and then c forever is equal to $H^i(x) = 1 + c^i\beta_i\frac{\beta_i\alpha}{(1-\beta_i)(1-\beta_i(1-\alpha))} + c^i\beta_i\frac{\zeta x}{1-\beta_i(1-\alpha)}$. With some more algebra, it can be checked that $V^i(0) \geqslant H^i(0)$ holds if $\beta_i \geqslant \frac{1}{\alpha c^i+(1-\alpha)}$. Moreover, if $V^i(0) \geqslant H^i(0)$ then $V^i(x) \geqslant H^i(x)$ for every $x \in [0,1]$ because $\beta_i\zeta \leqslant 1$. By the principle of optimality, also called the one-shot principle, cooperation in every period is optimal if $V^i(x) \geqslant H^i(x)$ for every $x \in [0,1]$. Hence, both players will optimally cooperate in all periods if $\frac{1}{\alpha c^i+(1-\alpha)} \leqslant \beta_i < 1$. An open question is whether players optimally do not cooperate if $\beta_i < \frac{1}{\alpha c^i+(1-\alpha)}$.

Note that $\frac{1}{\alpha c^i+(1-\alpha)}$ approaches 1 if c^i approaches 1 or a approaches 0. However, if a is zero, then players optimally will not cooperate because both players believe that the other player will not cooperate with probability one in all periods.

Proposition 1 is a sharp result. The example above shows that proposition 1 would not be true if the assumption of positive influence were replaced by the weaker assumption "$e^i(x) \geqslant x$." The assumption of negative influence could be replaced by the weaker assumption "$r^i(x) \leqslant x$." The proof would be identical to the proof given in the appendix. However, a simple variation of example 2 shows that proposition 1 would not be true if negative influence were replaced by the weaker assumption "$r^i(x) > x$ for $x \leqslant \bar{x}$ and $r^i(x) \leqslant x$ for $x > \bar{x}$, $\bar{x} > 0$.

4. Conclusion

In this paper, it is shown that patient players who believe in reciprocity optimally cooperate. Reciprocity is represented by a class of well-specified beliefs.

Several important issues remain unresolved for models of this type. The difficulties emerge when we attempt to demonstrate that a certain action is optimal for all beliefs in a large class. The techniques used in this paper resolve some of these difficulties, but unfortunately many questions remain unanswered. For example, can the main result be proved without some of the regularity assumptions imposed on players' beliefs? Another important extension of the main result would be a full characterization of the outcomes of the game for all discount factors. That is, if players believe in reciprocity then what degree of patience is required to ensure that optimally they will cooperate? Is there a threshold such that if players' discount factors are above this level then optimally they will cooperate and if players' discount factors are below this level then optimally they will not cooperate?

5. Appendix

Assume that player i believes that the opponent will cooperate in the current period with probability x. Assume that player i decided to cooperate in all periods. Let $e_t^i(x)$ be the subjective probability that the opponent will cooperate t periods ahead. By definition,

$$e_0^i(x) = x, \;\; e_1^i(x) = e^i(x), \text{ and } V^i(x) = (c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t \, e_t^i(x) + \frac{w^i}{1-\beta_i}.$$

Moreover, by Bayes' rule,

$$e_{t+1}^i(x) = x e_t^i(g^i(x,c,c)) + (1-x) e_t^i(g^i(x,c,d)).$$

Proof of Lemma 1 - By definition, if the functions $g^i(x,c,c)$ and $g^i(x,c,d)$ are smooth then the functions e_t^i are also smooth. Then, V^i is a smooth function.

Assume, by induction, that $\frac{\partial}{\partial x} e_t^i(x) \geqslant 0$. Then, $\frac{\partial}{\partial x} e_{t+1}^i(x) =$

$$x \frac{\partial}{\partial x} e_t^i(g^i(x,c,c)) \frac{\partial}{\partial x} g^i(x,c,c) + (1-x) \frac{\partial}{\partial x} e_t^i(g^i(x,c,d)) \frac{\partial}{\partial x} g^i(x,c,d)$$

$$e_t^i(g^i(x,c,c)) - e_t^i(g^i(x,c,d) \geqslant 0.$$

Hence, e_t^i is a non-decreasing function for all t. Moreover, V^i is a non-decreasing function because V^i is a linear combination of the functions e_t^i.

Assume, by the induction assumption that $\frac{\partial^2}{\partial x^2} e_t^i(x) \leqslant 0$. Then, $\frac{\partial^2}{\partial x^2} e_{t+1}^i(x) =$

$$x\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2} g^i(x,c,c) + x\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x} g^i(x,c,c)\right)^2 +$$

$$2\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial}{\partial x} g^i(x,c,c) - 2\frac{\partial}{\partial x} e_t^i(g^i(x,c,d))\frac{\partial}{\partial x} g^i(x,c,d) +$$

$$(1-x)\frac{\partial}{\partial x} e_t^i(g^i(x,c,d))\frac{\partial^2}{\partial x^2} g^i(x,c,d) + (1-x)\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x} g^i(x,c,d)\right)^2 \leqslant$$

$$x\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x} g^i(x,c,c)\right)^2 + (1-x)\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x} g^i(x,c,d)\right)^2 +$$

$$x\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2} g^i(x,c,c) + 2\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial}{\partial x} g^i(x,c,c) +$$

$$-2\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial}{\partial x} g^i(x,c,d) + (1-x)\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2} g^i(x,c,d) =$$

$$x\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x} g^i(x,c,c)\right)^2 + (1-x)\frac{\partial^2}{\partial x^2} e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x} g^i(x,c,d)\right)^2 +$$

$$\frac{\partial}{\partial x} e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2} e^i(x) \leqslant 0.$$

Hence, e_t^i is a concave function for all t, and V^i is a concave function because V^i is a linear combination of the functions e_t^i.

q.e.d.

Proof of Lemma 2 - Let $g^i(x,c,c)$ and $g^i(x,c,d)$ be functions such that positive influence is satisfied. Let f_{-i} be the well specified belief associated with $g^i(x,c,c)$ and $g^i(x,c,d)$ and arbitrary functions $g^i(x,d,c)$ and $g^i(x,d,d)$ and an arbitrary number x_1^i.

As defined before, let x_t^i be an $\Im_{t-1}-$measurable function representing player i's subjective probability that his opponent will cooperate in the current period. That is, given a play path $s = (h, ...)$, $h \in \Sigma^{t-1}$, $x_t^i(s) = f_{-i}(h)$. Let y_t^i be x_{t+1}^i.

Let l_i be the strategy "always cooperate". That is, $l_i(h) = 1$ for all $h \in H$. Let l^i be the strategy profile $l^i = (l_i, f_{-i})$. Let μ_{l^i} be the probability measure associated with l^i. That is, μ_{l^i} describes how player i thinks that the play will evolve if he always cooperates and holds the well specified belief f_{-i}.

By assumption, if player i cooperates in every period then

$$E^{\mu_{l^i}}\{y_t^i/\Im_{t-1}\} = e^i(y_{t-1}^i) \geqslant y_{t-1}^i.$$

Hence, y_t^i is a bounded positive supermartigale. By the theorem of the convergence of positive supermartingales, there exists a random variable y^i such that

$$y_t^i \underset{t\to\infty}{\to} y^i \ a.s. \ \mu_{l^i} \text{ and } E^{\mu_{l^i}}\{y_t^i\} \underset{t\to\infty}{\to} E^{\mu_{l^i}}\{y^i\}.$$

Hence,

$$E^{\mu_{l^i}}\{e^i(y^i) - y^i\} = \lim_{t\to\infty} E^{\mu_{l^i}}\{e^i(y_t^i) - y_t^i\} = \lim_{t\to\infty} E^{\mu_{l^i}}\{y_{t+1}^i - y_t^i\} = 0.$$

But, $e^i(y^i) - y^i > 0$ if $y^i < 1$. Thus, $y^i = 1$ $a.s.$ μ_{l^i}. Therefore,

$$x_t^i \underset{t\to\infty}{\to} 1 \ a.s. \ \mu_{l^i} \text{ and } E^{\mu_{l^i}}\{x_t^i\} \underset{t\to\infty}{\to} 1.$$

Assume that player i believed that the opponent would cooperate with probability zero in the first period. That is, assume that $x_1^i = 0$. In this case $e_t^i(0) = E^{\mu_{l^i}}\{x_t^i\}$. The limit above holds for every initial condition x_1^i. In particular, it holds for $x_1^i = 0$. Hence, $e_t^i(0) \underset{t\to\infty}{\to} 1$.

Fix $\varepsilon > 0$. Let ε_1 be $(c^i - w^i)\varepsilon$. Let $\bar{t}$ be large enough such that if $t \geqslant \bar{t}$ then $e_t^i(0) - 1 + \varepsilon_1 > \frac{\varepsilon_1}{2}$. Let $\bar{\beta}$ be large enough such that if $\beta_i \geqslant \bar{\beta}$ then $\sum\limits_{t=\bar{t}+1}^{\infty} (\beta_i)^t \frac{\varepsilon_1}{2} \geqslant \bar{t}$. By definition,

$$\sum_{t=0}^{\bar{t}} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant \sum_{t=0}^{\bar{t}} (\beta_i)^t (-1) \geqslant -\bar{t}$$

and

$$\sum_{t=\bar{t}+1}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant \sum_{t=\bar{t}+1}^{\infty} (\beta_i)^t \frac{\varepsilon_1}{2} \geqslant \bar{t}.$$

So,

$$\sum_{t=0}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant 0.$$

By definition,

$$V^i(0) = (c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t e_t^i(0) + \frac{w^i}{1-\beta_i} =$$

$$(c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) + \frac{w^i}{1-\beta_i} + (c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t (1 - \varepsilon_1) \geqslant$$

$$\frac{w^i}{1-\beta_i} + \frac{(c^i - w^i)(1-\varepsilon_1)}{1-\beta_i} = \frac{c^i + (w^i - c^i)\varepsilon_1}{1-\beta_i} = \frac{c^i - \varepsilon}{1-\beta_i}.$$

q.e.d.

Proof of Proposition 1 - I first show that both players will optimally cooperate in all periods. By the principle of optimality, it suffices to show that the expected discounted payoff of cooperating in all periods, $V^i(x)$, is greater than the expected discounted payoff of playing d and then c in all periods which is given by

$$H^i(x) = xz^1 + (1-x)d^1 + \beta^i \left(xV^i(g(x,d,c) + (1-x)V^i(g(x,d,d)\right).$$

By the concavity of V^i,

$$H^i(x) \leqq xz^1 + (1-x)d^1 + \beta^i V^i \left(x(g(x,d,c) + (1-x)(g(x,d,d)\right).$$

By the monotonicity of V^i, and the assumption that player i believes in negative influence,

$$H^i(x) \leqq xz^1 + (1-x)d^1 + \beta^i V^i(x).$$

Let ε be small enough such that $xz^1 + (1-x)d^1 < c^i - \varepsilon$. Then, by lemma 2, if β^i is large enough then,

$$H^i(x) < c^i - \varepsilon + \beta^i V^i(x) \leqslant V^i(x).$$

Hence, both players optimally cooperate in every period. By assumption, $g^i(x, c, c) > x$ if $x < 1$ and $g^i(1, c, c) = 1$. Thus, if the cooperative outcome occurs in every period then both players' beliefs about the probability that the opponent will cooperate will increase every period and eventually both players will believe that the opponent will cooperate with arbitrarily high probability.

q.e.d.

References

[1] Anderlini, L. and H. Sabourian (1995), "Cooperation and Effective Computability," *Econometrica*, 63, 6, 137-1400.

[2] Aumann, R. and S. Sorin (1989), "Cooperation and Bounded Recall," *Games and Economic Behavior*, 1, 5-39.

[3] Battigalli, P. and J. Watson (1997), "On Reputation Refinements With Heterogeneous Beliefs," *Econometrica*, 65, 2, 369-374.

[4] Crawford, V. (1991) "An 'Evolutionary' Interpretation of Van Huyck, Battalio and Beil's Experimental Results on Coordination," *Games and Economic Behavior,* 3, 25-59.

[5] Crawford, V. (1995) "Adaptive Dynamics in Coordination Games," *Econometrica*, 63, 1, 103-143.

[6] Crawford, V. and H. Haller (1990), "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica*, 58, 571-595.

[7] Goyal, S. and M. Janssen (1996) "Can We Rationally Learn to Coordinate," *Theory and Decision*, 40, 29-49.

[8] Güth, W. (1995) "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives," *International Journal of Game Theory*, 24, 323-344.

[9] Kalai, E. and E. Lehrer (1993a) "Rational Learning Leads to Nash equilibrium," *Econometrica*, 61, 1021-1045.

[10] Kalai, E. and E. Lehrer (1993b) "Subjective Equilibrium in Repeated Games", *Econometrica*, 61, 1231-1240.

[11] Kandori, M., Mailath, G. and R. Rob (1993), "Learning, Mutation and Long Run Equilibria in Games," *Econometrica*, 61, 29-56.

[12] Kuhn, H. (1953), "Extensive Games and the Problem of Information," *Annals of Mathematics Studies* (1953), 137-156.

[13] Lagunoff, R. (1997) "On the Evolution of Pareto Optimal Behavior in Repeated Coordination Problems," mimeo, Georgetown University.

[14] Matsui, A. (1991), "Cheap Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245-258.

[15] Matsui, A and R. Rob (1991) "Evolution and Rationality in Societal Games," CARESS Working Paper #91-31, University of Pennsylvania.

[16] Fehr, E., Gächter, S., and G. Kirchsteiger (1997) "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65, 4, 833-860.

[17] Watson, J. (1993) "A Reputation Refinement Without Equilibrium," *Econometrica*, 61, 199-205.

[18] Young, P. (1993) "The Evolution of Conventions," *Econometrica*, 61, 57-84.

Discussion Paper No. 1200

Reciprocity and Cooperation in Repeated Coordination Games: The Blurry Belief Approach

by

Alvaro Sandroni

November 1997

Reciprocity and Cooperation in Repeated Coordination Games: The Blurry Belief Approach

Alvaro Sandroni*
J. L. Kellogg Graduate School of Management
MEDS Department
Northwestern University
2001 Sheridan Road, Evanston, IL 60208
sandroni@nwu.edu

November 2, 1997

Abstract

Two long lived players play a repeated coordination game. Players do not specify a single (and correct) probability to each event. They have a vague notion about the evolution of the play, called blurry beliefs, which guide their behavior. General conditions that ensure cooperation are investigated.

Key words: Repeated Games, Learning, Cooperation, Bounded Rationality, Equilibrium Selection.

JEL classification: D83

*I had the benefit of extensive conversations with Ehud Kalai, whose ideas were fundamental in realizing this project. I would also like to thank Eddie Dekel, George Mailath, and Juuso Valimaki for useful suggestions. This paper began in a conversation with Rann Smorodinsky who helped me on many occasions. All errors are mine.

1. Introduction

Consider a repeated coordination game with two long-lived players and two actions available to each player at each stage. Assume that there is a single outcome which yields the highest payoffs for both players. Call the action associated with this outcome "to cooperate." Call the alternative action "not to cooperate." A fundamental question in game theory is whether cooperation can be derived by the principle of rational behavior (see Aumann and Sorin (89)). The standard way of modeling rational behavior is to impose equilibrium behavior. It is assumed that players' beliefs are correct and that players take best responses according to these beliefs. These assumptions are very demanding because players are assumed to know (or behave as if they knew) their opponents' strategies. However, equilibrium behavior does not necessarily ensure cooperation. In fact, the Folk Theorem shows that there are a large multiplicity of equilibria in repeated games. In particular, both the cooperative outcome and highly inefficient outcomes remain equilibria. For example, players will (not) cooperate, in all periods, if they believe that the other player will (not) cooperate regardless of past outcomes.

Clearly, the outcome of the game will depend upon players' beliefs about the evolution of the game. Hence, it is important to develop a theory of belief formation. In this paper, I focus on the question of belief representation. I do not assume that players specify a single probability to each event nor do I assume that players behave as if they knew their opponents' strategies; rather, I assume that players have some "notions" about how play might evolve, know their own payoffs and take best responses according to these vague notions. For example, consider a player who come to the game with the following notion: "If I cooperate then my opponent will be more inclined to cooperate." Is there an optimal action for this player? How can this player's belief be explicitly modeled? How would this belief influence (and be influenced by) the evolution of the game?

The notions which guide players' behavior are modelled by a "*blurry belief.*" A blurry belief is a class of well-specified beliefs. The restrictions defining this class are derived from a general principle which, in turn, defines the players' incomplete notions about how play might evolve. The core question is whether there exists an optimal strategy independent of the specification of the belief in the class. If a strategy is optimal for all beliefs in a class, then this strategy is optimal for players maintaining the notions that this class represents.

The following example shows the subtlety of this analysis. Assume that each player believes that if he cooperates, in the first period, then the opponent will cooperate thereafter. This class of beliefs seems too restrictive to represent the notion that "if I cooperate then my opponent will be more inclined to do the same." However, does acceptance of such a principle ensure cooperation? Apparently, a sufficiently patient player who holds this view optimally should cooperate, regardless of what he thinks might happen if he does not cooperate in the first period, because, if he does cooperate, then any loss he might incur in the first period is compensated by the maximum payoffs obtained afterwards. Surprisingly, there exists a belief in this class such that, for all discount factors, both players optimally do not cooperate and eventually learn that the other player does not cooperate. The example that illustrates this point is called the "waiting players." In this example, both players believe that cooperation can also be triggered if the opponent cooperates and optimally they do not cooperate because they wait for the other player to cooperate first. Eventually, each player realizes that the other player will not cooperate in the near future with arbitrarily high probability.

The waiting players example shows that even if the class of beliefs seems restrictive it may not ensure cooperation. Clearly, it is necessary to assume other kinds of restrictions on players' beliefs. I assume that a player's belief about the probability that his opponent will cooperate depends upon the last outcome and the probability that he had assigned to his opponent's cooperating in the previous period. This restriction imposes a recursive structure on players' beliefs.

A player believes in *positive influence* if he believes that if he cooperates in the current period then it is more probable that the other player will cooperate in the next period than in this period. This class of beliefs represents the notion that "if I cooperate then my opponent will be more inclined to cooperate." As should be expected from the waiting players example, arbitrarily patient players, who believe in positive influence, optimally may not cooperate.

A player who believes in *negative influence* believes that if he does not cooperate in the current period then it is less probable that the opponent will cooperate in the next period than in this period. This class of beliefs represents the notion that "if I do not cooperate then my opponent will be less inclined to cooperate." A player believes in *reciprocity* if he believes in both positive and negative influence.[1]

The main result is that sufficiently patient players who believe in reciprocity optimally cooperate.[2] So, cooperation is derived from the principle of reciprocity.

[1] See Fehr, Gächter, and Kirchsteiger (97) for experimental work on reciprocity.

[2] Some additional restrictions on players' beliefs, of a technical nature, are also assumed.

Reciprocity and positive influence are heuristics which a player may use to guide his behavior. They are represented by a blurry belief which is a class of well-specified beliefs consistent with each principle. There are well-specified beliefs consistent with positive influence that leads to cooperation. However, cooperation does not necessarily follows from positive influence because, as shown by the waiting players example, there are beliefs which are consistent with positive influence that do not lead to cooperation. Hence, even if players accept this principle they may not optimally cooperate because they may make other considerations.

The main result in this paper can be viewed as an anti-folk theorem because a unique outcome is obtained in an infinitely repeated game with patient players. The basic assumption is that players believe in reciprocity. This does not mean that players believe in reciprocity and no other thought occurs to them. Players optimally cooperate for all well-specified beliefs in the class representing reciprocity. Some of these well-specified beliefs may belong to other classes representing other principles. Hence, sufficiently patient players who believe in reciprocity optimally cooperate even if they also make other considerations.

Cooperation is a Nash equilibrium play in coordination games. If, as in the case of this paper, the outcome of interest is a Nash equilibrium play, then blurry beliefs generate learning schemes that lead to a Nash equilibrium. However, as opposed to many results in the rational learning literature, no compatibility conditions between the beliefs and the true play, such as absolute continuity, have been assumed (see Kalai and Lehrer (93a) and (93b)). Thus, convergence to Nash equilibrium occurs solely because of the restrictions imposed on the exogenous variables i.e., beliefs, discount factors, and stage game payoffs.[3]

Blurry beliefs may be used in any game. For example, what principles lead to cooperation in the repeated prisoners' dilemma? This is an open and interesting question. However, a blurry belief probably should not be interpreted outside a given context. The interpretation and consequences of a blurry belief depend on the particular game being played. In particular, there is no reason to assume that reciprocity, as defined for coordination games, also leads to cooperation in any

[3]In the fictitious play literature, no compatibility conditions between beliefs and best responses are assumed. However, players believe that their past actions do not affect the future actions of the other player. Thus, long-run strategic considerations are ignored. So, these models are inappropriate when there are few, long-lived players as, for example, in the case of a firm and a worker who play the same game repeatedly. The same is true in evolutionary models where equilibrium selection results have also been obtained (see Matsui and Rob (91), Kandori, Mailath and Rob (93), Young (93)).

other game. However, although different principles may arise in different games, they may all be described in the framework of the rational learning model.

2. The Repeated Coordination Game

There are two players I and II. Player $i \in \{I, II\}$ has two possible actions given by the set $\sum = \{c, d\}$. The payoff function $u_i : \sum \to \Re$ is given by the payoff matrix

$$\begin{bmatrix} (I, II) & c & d \\ c & (c^1, c^2) & (w^1, w^2) \\ d & (z^1, z^2) & (d^1, d^2) \end{bmatrix}$$

where $c^i > \max\{z^i, w^i, d^i\}$, $i \in \{I, II\}$. A player (does not) cooperate if he plays (d) c. The outcome (c, c) is called cooperative.

Let $\sum^t$ be the set of all t-histories, $0 \le t \le \infty$. Let $H = \bigcup_{t \geqslant 0} \sum^t$ be the set of all finite histories. Let $\Im_0 \subset ...\Im_t \subset ... \subset \Im$ be the filtration on $\sum^\infty$ where $\Im_t$ is the σ-algebra generated by all t-histories, and $\Im$ is the σ-algebra generated by the algebra $\bigcup_{t \geqslant 0} \Im_t$.

Let Ψ be the set of all functions $g : H \to [0, 1]$. A behavior strategy $f_i \in \Psi$ describes the probability that player $i \in \{I, II\}$ will cooperate conditional on each finite history. Let $f = (f_I, f_{II})$ be the true behavior strategy profile. A well-specified belief $f_{-i} \in \Psi$ describes the probability that player i believes that the other player will cooperate, conditional on each finite history.[4] Player i's blurry belief is a set of well-specified beliefs $\Delta_{-i} \subset \Psi$.

Given a strategy profile q, let μ_q be the probability measure over play paths associated with q.[5] Given the strategy profile $q^i = (q, f_{-i})$, player i's discounted expected payoff is

$$V_i(q^i) = E^{\mu_{q^i}} \left\{ \sum_{r=0}^{\infty} \{(\beta_i)^r u_i\} \right\},$$

where β_i, $0 < \beta_i < 1$, is player i's discount factor and $E^{\mu_{q^i}}$ is the expectation operator associated with μ_{q^i}. The behavior strategy f_i is a best response to f_{-i} if for every strategy profile $q^i = (q, f_{-i})$, $V_i(f^i) - V_i(q^i) \ge 0$. The behavior strategy f_i is a best response to Δ_{-i} if f_i is a best response to all $f_{-i} \in \Delta_{-i}$.

[4]See Kuhn (53) for the description of players' beliefs.

[5]See Kalai and Lehrer (93) for details on the construction of this probability measure.

3. Reciprocity and Cooperation

Definition 1. *Players learn to cooperate if there exists a set $\Omega \in \Im$ such that $\mu_f(\Omega) = 1$, and given $f_{-i} \in \Delta_{-i}$, for every $s \in \Omega$, $s = (h, \ldots)$, $h \in \Sigma^t$, and $\varepsilon > 0$, there exists a period $\bar{t}$ such that for all $t \geq \bar{t}$, $f_{-i}(h) \geqslant 1 - \varepsilon$ and $f_i(h) \geqslant 1 - \varepsilon$.*

Players learn to cooperate if eventually players' beliefs and the behavior strategies are arbitrarily close to the Nash equilibrium in which players always cooperate.

Let the blurry belief ξ^l_{-i} be defined by $f_{-i} \in \xi^l_{-i}$ if and only if $f_{-i}(h) = 1$ for all finite histories $h \in H$ such that player i cooperated in the first l periods.

Definition 2. *Player i believes that he can induce the other player to cooperate by cooperating in the first l periods if player i holds the blurry belief ξ^l_{-i}.*

That is, player i believes that he can induce the other player to cooperate by cooperating in the first l periods if he thinks that if he cooperates in the first l periods, then the other player will respond by cooperating, with probability one, thereafter.

Example 1, below, shows that players who believe that they can induce the other player to cooperate by cooperating in the first period do not necessarily learn to cooperate regardless of how patient they are.

Example 1. *The waiting players.*

The payoffs of stage game are described by the payoff matrix

$$\begin{bmatrix} (I, II) & c & d \\ c & (9,9) & (0,8) \\ d & (8,0) & (8,8) \end{bmatrix}.$$

Let k be a natural number such that $\left(\frac{3}{4}\right)^k k^2 \sum\limits_{t=k+1}^{\infty} \frac{1}{t^2} < \frac{11}{4}$.[6] Player i believes that the other player will cooperate with probability 0.25 at period 1. If player i cooperates, at period 1, then he believes that the other player will cooperate

[6]The existence of k follows from $\left(\frac{3}{4}\right)^k k^2 \underset{k\to\infty}{\longrightarrow} 0$ and $\sum\limits_{t=k+1}^{\infty} \frac{1}{t^2} \underset{k\to\infty}{\longrightarrow} 0$.

thereafter. If player i does not cooperate, at period 1, but the other player cooperates, at any period, then he believes that the other player will cooperate thereafter. If player i does not cooperate, at period 1, and the other player does not cooperate until period t, then he believes that the other player will cooperate at period $t+1$ with probability γ_t, where $\gamma_t = 0.25$ if $t \leq k$, and $\gamma_t = 1 - \left(\frac{t-1}{t}\right)^2$ if $t > k$. By definition, both players believe that they can induce the other player to cooperate by cooperating in the first period.

If a player cooperates, at period 1, then he gets an expected discounted payoff which is smaller than or equal to

$$\frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t \, 9.$$

A player who only cooperates after the other player cooperates obtains the expected discounted payoff

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \left(\prod_{r=1}^{t} (1-\gamma_r) 8 + \left(1 - \prod_{r=1}^{t} (1-\gamma_r)\right) 9 \right) =$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - \sum_{t=1}^{\infty} (\beta_i)^t \left(\prod_{r=1}^{t} (1-\gamma_r) \right) =$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - \sum_{t=1}^{k} (\beta_i)^t \left(\left(\frac{3}{4}\right)^t \right) - \left(\frac{3}{4}\right)^k k^2 \sum_{t=k+1}^{\infty} (\beta_i)^t \left(\frac{1}{t^2}\right) >$$

$$8 + \sum_{t=1}^{\infty} (\beta_i)^t \, 9 - 3 - \left(\frac{3}{4}\right)^k k^2 \sum_{t=k+1}^{\infty} \frac{1}{t^2} > \frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t \, 9.$$

Hence, both players optimally do not cooperate in the first period. In all other periods, both players believe that the other player will cooperate with probability smaller than 8/9 as long as they have not yet observed the other player cooperating. Therefore, both players optimally do not cooperate before they observe the other player cooperating. The true play will be (d, d) in every period, and each player will eventually believe that the other player will cooperate with probability arbitrarily close to the true probability (zero).

In the waiting players example, if either player had cooperated at period 1 then both players would have cooperated thereafter. Hence, both players correctly believed they could get maximum payoffs after period 1. However, they optimally chose to wait for the other player to cooperate first. Thus, both players end up waiting forever and the payoffs obtained ex-post are much lower than the expected payoffs, although players' beliefs over short-run events eventually become accurate.

In a static coordination game, players optimally cooperate if they believe that the other will cooperate, and players optimally do not cooperate if they believe that the other will not cooperate. But, in a repeated coordination game, the waiting players example shows that players may not cooperate even if they believe that the cooperative outcome may eventually be played. In fact, arbitrarily patient players may not optimally cooperate even if they believe that they will induce the other player to (not) cooperate forever if they (do not) cooperate in the first periods. For example, assume that each player believes that the other player will cooperate, with probability one, in the first period. Player $i \in \{I, II\}$ believes that if he does not cooperate in the first period then the other player will not cooperate thereafter, and if he cooperates in the first period then his belief will be as in the waiting players example. Then, player i believes that if he cooperates in the first two periods then the other player will cooperate thereafter. However, it is straightforward to show that, regardless of their discount factor, the optimal play will be (c, c) at period 1 and (d, d) thereafter.

In the waiting player example, both players believe that if they cooperate in the first period then they will induce the other player to cooperate, but if they do not cooperate in the first period then their actions no longer have any influence on the other player's future actions. Analogously, in the example given above, players also believe that their influence over the other player's actions will disappear if they play c followed by d. So, I consider a class of "recurrent" beliefs in which players' beliefs about their potential influence on the other player's actions do not change so abruptly.

Given a well-specified belief f_{-i}, let x_t^i be a $\Im_{t-1}-$measurable function representing player i's subjective probability, conditional on all information available at period $t-1$, that his opponent will cooperate at period t. That is, given a play path $s = (h, ...), h \in \Sigma^{t-1}$, $x_t^i(s) = f_{-i}(h)$. Assume that x_t^i follows the rule $x_t^i = g^i(x_{t-1}^i, a, b)$ where $a \in \{c, d\}$ and $b \in \{c, d\}$ are the actions taken, at period $t-1$, by player i and the opponent, respectively. Hence, player i's belief about the probability that the opponent will cooperate depends upon the current outcome and the probability that player i assigned last period to cooperation on the part of the opponent.

Player i's belief is well-specified given x_1^i and the functions g^i. Player i's blurry beliefs may now be defined by restrictions on x_1^i and g^i.

Let e^i and r^i be the functions $e^i(x) = xg^i(x, c, c) + (1 - x)g^i(x, c, d)$ and $r^i(x) = xg^i(x, d, c) + (1 - x)g^i(x, d, d)$. Assume that after observing the outcome at period $t - 1$, player i decides to cooperate at period t. Then, he believes that the opponent will cooperate at period $t + 1$ with probability $e^i(x_t^i)$. Analogously, if, at period $t - 1$, player i decides not to cooperate at period t, then he believes that the opponent will cooperate at period $t + 1$ with probability $r^i(x_t^i)$.

Let $V^i(x)$ be the expected discounted payoff of player i if he decides always to cooperate, and if he believes that his opponent will cooperate with probability x in the current period.

Definition 3. *Player i's belief is regular if V^i is a smooth, non-decreasing, and concave function.*

The function V^i is non-decreasing if the expected discounted payoff associated with the strategy "always cooperate" does not decrease when player i's probability that the opponent will cooperate in the current period increases.

The function V^i is concave if the expected discounted payoff associated with the strategy "always cooperate" when player i believes that the opponent will cooperate with probability $\lambda\hat{x} + (1 - \lambda)\bar{x}$ in the current period is not smaller than the linear combination (using λ as weight) of the expected discounted payoffs associated with the same strategy when player i believes that in the current period the opponent will cooperate with probabilities $\hat{x}$ and $\bar{x}$, respectively. Lemma 1, below, provides sufficient conditions under which player i's beliefs are regular.

Lemma 1. *Assume that $g^i(x, c, c)$ and $g^i(x, c, d)$ are smooth functions of x. If $g^i(x, c, c) \geqslant g^i(x, c, d)$ and $g^i(x, c, c)$, $g^i(x, c, d)$ are non-decreasing functions of x then V^i is a smooth and non-decreasing function of x.*[7] *If, in addition, e^i and $g^i(x, c, d)$ are concave functions of x, then V^i is also a concave function of x.*[8]

[7] The assumption that $g^i(x, c, c) \geqslant g^i(x, c, d)$ means that player i's belief about the probability that the other player will cooperate is not smaller when the cooperative outcome is observed than when player i cooperates, but the other player doesn't.

[8] The assumption that e^i and $g^i(x, c, d)$ are concave functions of x are of a technical nature and the main result (proposition 1) is probably true without it, but a formal proof does not seem to follow from the techniques developed in this paper to solve the optimization problems.

Proof - See Appendix.

Definition 4. *Player i believes in positive influence if $e^i(x) > x$ whenever $x \in [0, 1)$, and $e^i(1) = 1$.*

Player i believes in positive influence if player i expects to make the opponent more inclined to cooperate by cooperating himself, in the sense that if player i cooperates then he believes that the opponent will be more likely to cooperate in the next period than in the current period. That is, at the beginning of period t, before the outcome is realized, player i believes that the other player will cooperate with probability x_t^i. Player i knows that whatever action he takes will not have any influence over the other player's decision in the current period, but knows his action might influence the other player's action in the next period. If player i believes in positive influence, then player i believes that by cooperating he will increase the chances that the other player will cooperate compared to the current odds.

If players believe in reciprocity and $g^i(x, c, c) \geqslant g^i(x, c, d)$ then $g^i(x, c, c) \geqslant x$. Hence, after observing the other players' action, players will become more confident that the other player will cooperate if they observe the cooperative outcome. However, if player i cooperates and the opponent doesn't then player i may or may not become more confident that the other player will cooperate.

Definition 5. *Player i believes in negative influence if $r^i(x) < x$ whenever $x \in (0, 1]$, and $r^i(0) = 0$.*

Player i believes in negative influence if player i expects to make the opponent less inclined to cooperate by not cooperating himself, in the sense that if player i does not cooperate then he believes it is less likely that the opponent will cooperate in the next period than in the current period. Negative and positive influence are, of course, perfectly symmetric restrictions on players' beliefs.

Definition 6. *Player i believes in reciprocity if player i believes in both positive and negative influence.*

Relaxing the concavity assumption would make the results much more attractive because it is in the spirit of the blurry belief approach to consider classes of beliefs that are as general as possible and, more importantly, to make restrictions on beliefs that follows from principles that are easily interpretable. Unfortunately, I do not know how to dispose of this assumption.

Lemma 2. *If player i believes in positive influence, then for every $\varepsilon > 0$ there exists $\bar{\beta}$ such that if $\beta_i \geqslant \bar{\beta}$ then $V^i(0) \geqslant \frac{c^i - \varepsilon}{1-\beta_i}$.*

Proof - See Appendix.

Lemma 2 shows that if player i believes in positive influence and player i is sufficiently patient, then player i expects to obtain high expected discounted payoffs by cooperating in every period even in the extreme case that player i is sure that the opponent will not cooperate in the current period.

I now show the main result of this paper. Proposition 1, below, shows that cooperation can be derived from the principle of reciprocity, in the sense that patient players, whose beliefs are regular, learn to cooperate if they believe in reciprocity.

Proposition 1. *Assume that players' beliefs are regular. If players believe in reciprocity and players' discount factors are sufficiently high, then players learn to cooperate. Moreover, the play will be cooperative in all periods.*[9]

Proof - See Appendix.

Example 2. *Patient players who believe in positive influence, but not in negative influence, optimally may not cooperate.*

Consider the same payoff matrix as in the waiting players' example. Assume that $g^i(x, c, c) = g^i(x, c, d) = g^i(x, d, c) = 1$, and $g^i(x, d, d) = x_1^i = 0.25$. That is, each player believes that the other player will cooperate with probability 0.25 in the first period, and will continue to cooperate with probability 0.25 if (d, d) is observed last period. Otherwise, the other player will cooperate with probability one. If player i cooperates in the first period he obtains a discounted expected payoff equal to

$$\frac{9}{4} + \sum_{t=1}^{\infty} (\beta_i)^t 9.$$

[9]Proposition 1 also holds if instead of assuming that V^i is concave it is assumed that player $i \in \{I, II\}$ believes that if he does not cooperate then the probability of cooperation on the part of the opponent will not increase regardless of the action of the opponent i.e., $g^i(x, d, c) \leqslant x$ and $g^i(x, d, d) \leqslant x$.

A player who only cooperates after the other player cooperates obtains an expected discounted payoff equivalent equal to

$$8+\sum_{t=1}^{\infty}(\beta_i)^t\left((0.75)^t 8+\left(1-(0.75)^t\right)9\right)=$$

$$8+\sum_{t=1}^{\infty}(\beta_i)^t 9-\sum_{t=1}^{\infty}(0.75\beta_i)^t>\frac{9}{4}+\sum_{t=1}^{\infty}(\beta_i)^t 9.$$

Hence, both players do not cooperate in the first period. In the second period, players will face the same maximization problem as in the first period. By induction, it can be shown that the play will be (d, d) in every period.[10]

Patient players who believe in negative influence, but not in positive influence, may not optimally cooperate. For example, assume that $x_1^i = g^i(x, c, c) = g^i(x, c, d) = g^i(x, d, c) = g^i(x, d, d) = 0$. Then, each player believes that the other player will not cooperate in every period, and optimally will not cooperate in every period.

Players who believe in reciprocity may not cooperate if they are not sufficiently patient. For example, assume that the payoffs are the same as in the waiting players' example. Let the beliefs be given by $x_1^i = g^i(x, c, c) = g^i(x, c, d) = 1$ and $x_1^i = g^i(x, d, c) = g^i(x, d, d) = 0$. Let the discount factor of each player be 0.25. If a player cooperates, at period 1, he gets an expected discounted payoff smaller than $\sum_{t=1}^{\infty}(0.25)^t 9 < 8$. If a player does not cooperate, then he gets a payoff greater than 8. By induction, it is easy to show that both players will not cooperate in every period.

It is interesting to consider an example in which the players' optimization problem can be solved directly. Let the payoffs be given by the matrix

$$\begin{bmatrix} (I, II) & c & d \\ c & (c^1, c^2) & (0, 0) \\ d & (1, 1) & (1, 1) \end{bmatrix}$$

where $c^1 > 1$ and $c^2 > 1$.

Players' beliefs are given by $x_1^i = 0$; $g^i(x, c, c) = g^i(x, c, d) = \alpha + (1-\alpha)x$; $g^i(x, d, c) = g^i(x, d, d) = \zeta x$; where $\alpha > 0$ and $0 \leqslant \zeta < 1$.

[10] In this example, players' predictions are not accurate because players will not cooperate but they believe that the other player will cooperate with probability 0.25.

Let $e_t^i(x)$ be player i's subjective probability that the other player will cooperate t periods ahead. If player i cooperates in all periods, then $e_{t+1}^i(x) = e_t^i(\alpha+(1-\alpha)x)$ is the subjective probability that the other players will cooperate $t+1$ periods ahead. By definition, $e_1^i(x) = e^i(x) = \alpha+(1-\alpha)x$ is a linear function of x. If $e_t^i(x)$ is a linear function, then $e_{t+1}^i(x)$ is a linear function. Hence, by induction, if player i cooperates in all periods then $e_t^i(x)$ is a linear function of x for all t. The expected payoff of cooperating in all periods, $V^i(x)$, is equal to $c^i \sum_{t=0}^{\infty} (\beta_i)^t e_t^i(x)$. Thus, $V^i(x)$ is a linear function of x. By definition, $V^i(x) = xc^i + \beta_i V^i(\alpha + (1-\alpha)x)$. With some algebra, it follows that $V^i(x) = \frac{c^i\beta_i\alpha}{(1-\beta_i)(1-\beta_i(1-\alpha))} + \frac{c^i x}{1-\beta_i(1-\alpha)}$. Therefore, the expected discounted payoff of playing d and then c forever is equal to $H^i(x) = 1 + c^i\beta_i\frac{\beta_i\alpha}{(1-\beta_i)(1-\beta_i(1-\alpha))} + c^i\beta_i\frac{\zeta x}{1-\beta_i(1-\alpha)}$. With some more algebra, it can be checked that $V^i(0) \geqslant H^i(0)$ holds if $\beta_i \geqslant \frac{1}{\alpha c^i+(1-\alpha)}$. Moreover, if $V^i(0) \geqslant H^i(0)$ then $V^i(x) \geqslant H^i(x)$ for every $x \in [0,1]$ because $\beta_i\zeta \leqslant 1$. By the principle of optimality, also called the one-shot principle, cooperation in every period is optimal if $V^i(x) \geqslant H^i(x)$ for every $x \in [0,1]$. Hence, both players will optimally cooperate in all periods if $\frac{1}{\alpha c^i+(1-\alpha)} \leqslant \beta_i < 1$. An open question is whether players optimally do not cooperate if $\beta_i < \frac{1}{\alpha c^i+(1-\alpha)}$.

Note that $\frac{1}{\alpha c^i+(1-\alpha)}$ approaches 1 if c^i approaches 1 or a approaches 0. However, if a is zero, then players optimally will not cooperate because both players believe that the other player will not cooperate with probability one in all periods.

Proposition 1 is a sharp result. The example above shows that proposition 1 would not be true if the assumption of positive influence were replaced by the weaker assumption "$e^i(x) \geqslant x$." The assumption of negative influence could be replaced by the weaker assumption "$r^i(x) \leqslant x$." The proof would be identical to the proof given in the appendix. However, a simple variation of example 2 shows that proposition 1 would not be true if negative influence were replaced by the weaker assumption "$r^i(x) > x$ for $x \leqslant \bar{x}$ and $r^i(x) \leqslant x$ for $x > \bar{x}$, $\bar{x} > 0$.

4. Conclusion

In this paper, it is shown that patient players who believe in reciprocity optimally cooperate. Reciprocity is represented by a class of well-specified beliefs.

Several important issues remain unresolved for models of this type. The difficulties emerge when we attempt to demonstrate that a certain action is optimal for all beliefs in a large class. The techniques used in this paper resolve some of these difficulties, but unfortunately many questions remain unanswered. For example, can the main result be proved without some of the regularity assumptions imposed on players' beliefs? Another important extension of the main result would be a full characterization of the outcomes of the game for all discount factors. That is, if players believe in reciprocity then what degree of patience is required to ensure that optimally they will cooperate? Is there a threshold such that if players' discount factors are above this level then optimally they will cooperate and if players' discount factors are below this level then optimally they will not cooperate?

5. Appendix

Assume that player i believes that the opponent will cooperate in the current period with probability x. Assume that player i decided to cooperate in all periods. Let $e^i_t(x)$ be the subjective probability that the opponent will cooperate t periods ahead. By definition,

$$e^i_0(x) = x,\ e^i_1(x) = e^i(x), \text{ and } V^i(x) = (c^i - w^i)\sum_{t=0}^{\infty}(\beta_i)^t\, e^i_t(x) + \frac{w^i}{1-\beta_i}.$$

Moreover, by Bayes' rule,

$$e^i_{t+1}(x) = xe^i_t(g^i(x,c,c)) + (1-x)e^i_t(g^i(x,c,d)).$$

Proof of Lemma 1 - By definition, if the functions $g^i(x,c,c)$ and $g^i(x,c,d)$ are smooth then the functions e^i_t are also smooth. Then, V^i is a smooth function.

Assume, by induction, that $\frac{\partial}{\partial x}e^i_t(x) \geqslant 0$. Then, $\frac{\partial}{\partial x}e^i_{t+1}(x) =$

$$x\frac{\partial}{\partial x}e^i_t(g^i(x,c,c))\frac{\partial}{\partial x}g^i(x,c,c) + (1-x)\frac{\partial}{\partial x}e^i_t(g^i(x,c,d))\frac{\partial}{\partial x}g^i(x,c,d)$$

$$e^i_t(g^i(x,c,c)) - e^i_t(g^i(x,c,d) \geqslant 0.$$

Hence, e^i_t is a non-decreasing function for all t. Moreover, V^i is a non-decreasing function because V^i is a linear combination of the functions e^i_t.

Assume, by the induction assumption that $\frac{\partial^2}{\partial x^2}e_t^i(x) \leqslant 0$. Then, $\frac{\partial^2}{\partial x^2}e_{t+1}^i(x) =$

$$x\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2}g^i(x,c,c) + x\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x}g^i(x,c,c)\right)^2 +$$

$$2\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial}{\partial x}g^i(x,c,c) - 2\frac{\partial}{\partial x}e_t^i(g^i(x,c,d))\frac{\partial}{\partial x}g^i(x,c,d)+$$

$$(1-x)\frac{\partial}{\partial x}e_t^i(g^i(x,c,d))\frac{\partial^2}{\partial x^2}g^i(x,c,d) + (1-x)\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x}g^i(x,c,d)\right)^2 \leqslant$$

$$x\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x}g^i(x,c,c)\right)^2 + (1-x)\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x}g^i(x,c,d)\right)^2 +$$

$$x\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2}g^i(x,c,c) + 2\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial}{\partial x}g^i(x,c,c)+$$

$$-2\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial}{\partial x}g^i(x,c,d) + (1-x)\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2}g^i(x,c,d) =$$

$$x\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,c))\left(\frac{\partial}{\partial x}g^i(x,c,c)\right)^2 + (1-x)\frac{\partial^2}{\partial x^2}e_t^i(g^i(x,c,d))\left(\frac{\partial}{\partial x}g^i(x,c,d)\right)^2 +$$

$$\frac{\partial}{\partial x}e_t^i(g^i(x,c,c))\frac{\partial^2}{\partial x^2}e^i(x) \leqslant 0.$$

Hence, e_t^i is a concave function for all t, and V^i is a concave function because V^i is a linear combination of the functions e_t^i.

q.e.d.

Proof of Lemma 2 - Let $g^i(x,c,c)$ and $g^i(x,c,d)$ be functions such that positive influence is satisfied. Let f_{-i} be the well specified belief associated with $g^i(x,c,c)$ and $g^i(x,c,d)$ and arbitrary functions $g^i(x,d,c)$ and $g^i(x,d,d)$ and an arbitrary number x_1^i.

As defined before, let x_t^i be an $\Im_{t-1}$−measurable function representing player i's subjective probability that his opponent will cooperate in the current period. That is, given a play path $s = (h, ...)$, $h \in \Sigma^{t-1}$, $x_t^i(s) = f_{-i}(h)$. Let y_t^i be x_{t+1}^i.

Let l_i be the strategy "always cooperate". That is, $l_i(h) = 1$ for all $h \in H$. Let l^i be the strategy profile $l^i = (l_i, f_{-i})$. Let μ_{l^i} be the probability measure associated with l^i. That is, μ_{l^i} describes how player i thinks that the play will evolve if he always cooperates and holds the well specified belief f_{-i}.

By assumption, if player i cooperates in every period then

$$E^{\mu_{l^i}}\{y_t^i/\Im_{t-1}\} = e^i(y_{t-1}^i) \geqslant y_{t-1}^i.$$

Hence, y_t^i is a bounded positive supermartigale. By the theorem of the convergence of positive supermartingales, there exists a random variable y^i such that

$$y_t^i \underset{t\to\infty}{\to} y^i \ a.s. \ \mu_{l^i} \text{ and } E^{\mu_{l^i}}\{y_t^i\} \underset{t\to\infty}{\to} E^{\mu_{l^i}}\{y^i\}.$$

Hence,

$$E^{\mu_{l^i}}\{e^i(y^i) - y^i\} = \lim_{t\to\infty} E^{\mu_{l^i}}\{e^i(y_t^i) - y_t^i\} = \lim_{t\to\infty} E^{\mu_{l^i}}\{y_{t+1}^i - y_t^i\} = 0.$$

But, $e^i(y^i) - y^i > 0$ if $y^i < 1$. Thus, $y^i = 1$ $a.s.$ μ_{l^i}. Therefore,

$$x_t^i \underset{t\to\infty}{\to} 1 \ a.s. \ \mu_{l^i} \text{ and } E^{\mu_{l^i}}\{x_t^i\} \underset{t\to\infty}{\to} 1.$$

Assume that player i believed that the opponent would cooperate with probability zero in the first period. That is, assume that $x_1^i = 0$. In this case $e_t^i(0) = E^{\mu_{l^i}}\{x_t^i\}$. The limit above holds for every initial condition x_1^i. In particular, it holds for $x_1^i = 0$. Hence, $e_t^i(0) \underset{t\to\infty}{\to} 1$.

Fix $\varepsilon > 0$. Let ε_1 be $(c^i - w^i)\varepsilon$. Let $\bar{t}$ be large enough such that if $t \geqslant \bar{t}$ then $e_t^i(0) - 1 + \varepsilon_1 > \frac{\varepsilon_1}{2}$. Let $\bar{\beta}$ be large enough such that if $\beta_i \geqslant \bar{\beta}$ then $\sum_{t=\bar{t}+1}^{\infty} (\beta_i)^t \frac{\varepsilon_1}{2} \geqslant \bar{t}$. By definition,

$$\sum_{t=0}^{\bar{t}} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant \sum_{t=0}^{\bar{t}} (\beta_i)^t (-1) \geqslant -\bar{t}$$

and

$$\sum_{t=\bar{t}+1}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant \sum_{t=\bar{t}+1}^{\infty} (\beta_i)^t \frac{\varepsilon_1}{2} \geqslant \bar{t}.$$

So,

$$\sum_{t=0}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) \geqslant 0.$$

By definition,

$$V^i(0) = (c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t \, e_t^i(0) + \frac{w^i}{1-\beta_i} =$$

$$(c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t \left(e_t^i(0) - 1 + \varepsilon_1\right) + \frac{w^i}{1-\beta_i} + (c^i - w^i) \sum_{t=0}^{\infty} (\beta_i)^t (1 - \varepsilon_1) \geqslant$$

$$\frac{w^i}{1-\beta_i} + \frac{(c^i - w^i)(1-\varepsilon_1)}{1-\beta_i} = \frac{c^i + (w^i - c^i)\varepsilon_1}{1-\beta_i} = \frac{c^i - \varepsilon}{1-\beta_i}.$$

q.e.d.

Proof of Proposition 1 - I first show that both players will optimally cooperate in all periods. By the principle of optimality, it suffices to show that the expected discounted payoff of cooperating in all periods, $V^i(x)$, is greater than the expected discounted payoff of playing d and then c in all periods which is given by

$$H^i(x) = xz^1 + (1-x)d^1 + \beta^i \left(xV^i(g(x,d,c) + (1-x)V^i(g(x,d,d)\right).$$

By the concavity of V^i,

$$H^i(x) \leqq xz^1 + (1-x)d^1 + \beta^i V^i \left(x(g(x,d,c) + (1-x)(g(x,d,d)\right).$$

By the monotonicity of V^i, and the assumption that player i believes in negative influence,

$$H^i(x) \leqq xz^1 + (1-x)d^1 + \beta^i V^i(x).$$

Let ε be small enough such that $xz^1 + (1-x)d^1 < c^i - \varepsilon$. Then, by lemma 2, if β^i is large enough then,

$$H^i(x) < c^i - \varepsilon + \beta^i V^i(x) \leqslant V^i(x).$$

Hence, both players optimally cooperate in every period. By assumption, $g^i(x, c, c) > x$ if $x < 1$ and $g^i(1, c, c) = 1$. Thus, if the cooperative outcome occurs in every period then both players' beliefs about the probability that the opponent will cooperate will increase every period and eventually both players will believe that the opponent will cooperate with arbitrarily high probability.

q.e.d.

References

[1] Anderlini, L. and H. Sabourian (1995), "Cooperation and Effective Computability," *Econometrica*, 63, 6, 137-1400.

[2] Aumann, R. and S. Sorin (1989), "Cooperation and Bounded Recall," *Games and Economic Behavior*, 1, 5-39.

[3] Battigalli, P. and J. Watson (1997), "On Reputation Refinements With Heterogeneous Beliefs," *Econometrica*, 65, 2, 369-374.

[4] Crawford, V. (1991) "An 'Evolutionary' Interpretation of Van Huyck, Battalio and Beil's Experimental Results on Coordination," *Games and Economic Behavior*, 3, 25-59.

[5] Crawford, V. (1995) "Adaptive Dynamics in Coordination Games," *Econometrica*, 63, 1, 103-143.

[6] Crawford, V. and H. Haller (1990), "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica*, 58, 571-595.

[7] Goyal, S. and M. Janssen (1996) "Can We Rationally Learn to Coordinate," *Theory and Decision*, 40, 29-49.

[8] Güth, W. (1995) "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives," *International Journal of Game Theory*, 24, 323-344.

[9] Kalai, E. and E. Lehrer (1993a) "Rational Learning Leads to Nash equilibrium," *Econometrica*, 61, 1021-1045.

[10] Kalai, E. and E. Lehrer (1993b) "Subjective Equilibrium in Repeated Games", *Econometrica*, 61, 1231-1240.

[11] Kandori, M., Mailath, G. and R. Rob (1993), "Learning, Mutation and Long Run Equilibria in Games," *Econometrica*, 61, 29-56.

[12] Kuhn, H. (1953), "Extensive Games and the Problem of Information," *Annals of Mathematics Studies* (1953), 137-156.

[13] Lagunoff, R. (1997) "On the Evolution of Pareto Optimal Behavior in Repeated Coordination Problems," mimeo, Georgetown University.

[14] Matsui, A. (1991), "Cheap Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245-258.

[15] Matsui, A and R. Rob (1991) "Evolution and Rationality in Societal Games," CARESS Working Paper #91-31, University of Pennsylvania.

[16] Fehr, E., Gächter, S., and G. Kirchsteiger (1997) "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65, 4, 833-860.

[17] Watson, J. (1993) "A Reputation Refinement Without Equilibrium," *Econometrica*, 61, 199-205.

[18] Young, P. (1993) "The Evolution of Conventions," *Econometrica*, 61, 57-84.