Cooperation in Community Interaction without Information Flows

Parikshit Ghosh, Debraj Ray

The Review of Economic Studies, Volume 63, Issue 3 (Jul., 1996), 491-519.

Review of Economic Studies (1996) 63, 491–519
0034-6527/96/00210491$02.00

Cooperation in Community Interaction Without Information Flows

PARIKSHIT GHOSH
Boston University

and

DEBRAJ RAY
Boston University and Instituto de Análisis Económico (CSIC)

First version received October 1994; *final version accepted February* 1996 (*Eds.*)

We study cooperative behaviour in communities where the flow of information regarding past conduct is limited or missing. Players are initially randomly matched with no knowledge of each other's past actions; they endogenously decide whether or not to continue the repeated relationship. There is incomplete information regarding player types: a subset of the population is myopic, while the remainder have discount factors that permit cooperation, in principle. We define *social equilibrium* in such communities. Such equilibria are characterized by an initial testing phase, followed by cooperation if the test is successful. It is precisely the presence of myopic types that permit cooperation, by raising barriers to entry into new relationships. We examine the implications of increased patience, which takes two forms: an increase in the *number* of non-myopic types, and an increase in the *discount* factor of non-myopic types. These two notions turn out to have strikingly different implications for the degree of cooperation that can be sustained.

1. INTRODUCTION

In many spheres of social and economic interaction, the practice of cooperative, or "ethical" behaviour is widespread. Indeed, it may be argued that the proper functioning of a market system relies as much, if not more, on a high degree of trust and voluntary honouring of contracts, as on legal enforcement (or the compliance induced by the threat of legal punishment). However, such cooperative behaviour often seems to be at odds with selfish individual motives. In other words, apparently non-Nash strategic choices are common and, indeed, required for smooth transactions and even moderately efficient economic organization. Why do we see, in abundance, such actions of apparent selflessness?

The theory of repeated games provides an answer to this question by pointing out that many interactions are actually repeated, so that the threat of *future* retaliation may enforce cooperative behaviour. The culmination of this line of argument lies in the Folk Theorem (see, e.g. Fudenberg and Maskin (1986)), which says that *any* individually rational payoff vector can be sustained in a subgame-perfect equilibrium of a repeated game, *provided players have high enough discount factors.*[1] Hence, any degree of cooperation

1. To be precise, the theorem holds under a "full dimensionality condition"—the set of individually rational payoff vectors must have dimension equal to the number of players. Most games of interest usually satisfy this condition.

can be obtained as an equilibrium outcome when players are sufficiently patient. However, the standard theory refers to repeated games between a *fixed* set of players. In many interesting instances of repeated interaction, players play with changing opponents, the choice of opponents lying either with the players themselves, or with Nature, or to some extent with both. In such situations, the threat of future retaliation by a single opponent does not carry much bite, and as a result, mutually beneficial outcomes that are not one-shot Nash equilibria remain unexplained by this theory.

A group of papers on "random-matching games" (first introduced in Rosenthal (1979)) show that players may be able to cooperate or coordinate their actions even in settings where they play with changing partners. For example, Rosenthal and Landau (1979) demonstrate that for a particular bargaining game, players can avoid costly conflict through the evolution of "reputations" for players, and through suitable (incentive-compatible) "social norms" that prescribe behaviour based on these reputations. A recent paper by Kandori (1992)[2] greatly generalizes results on random matching games and extends the Folk Theorem to such games—he shows that (under certain conditions) any outcome that can be sustained in the two-player repeated game can also be sustained in the random matching version of the same stage game.

These papers assume, and their results rely on, a certain amount of information flow within the population or community from which players are drawn. Thus, a player, on meeting a new opponent, is assumed to obtain costlessly some relevant summary statistic of the opponent's past actions. While this is a valid assumption in some contexts (e.g. consumer credit markets in industrialized countries where a consumer's entire credit history is tracked on computer networks), in a large class of interesting and important situations, it is clearly not (e.g. in some informal markets in developing countries, where a defaulter can move to another village or town where his past crime will almost certainly be unknown). This motivates us to explore cooperation possibilities in community interaction *when information flows are absent*. Publicity and individual reputations work well as disciplining devices either in traditional, immobile societies, or in advanced societies with sophisticated information networks.[3] Most situations in currently developing countries fall in an intermediate zone, where sustainment of cooperative practices must rely on mechanisms other than blacklisting or social ostracism. In this paper, we propose one such mechanism.

Though cooperation possibilities seem very restricted at first glance in the setting we have described, we manage to identify a pair of fairly plausible conditions under which cooperative behaviour may be expected. First, we assume that matching is not completely random, players having the *option* of continuing to play old opponents.[4] Second, we assume that the population is non-homogeneous—while some players have a stake in the future (i.e. have discount factors greater than 0), there is a fraction of players who are myopic, and therefore prone to playing short-run best responses. We show that cooperative behaviour can emerge under these circumstances.

The presence of myopic types in the population gives patient players a scarcity value, which is utilized to sustain cooperation among non-myopic members. In a *social*

2. See also Okuno-Fujiwara and Postlewaite (1990).

3. This is also true in small groups, such as those of friends or colleagues.

4. *Voluntarily formed* long-term relationships can be seen all around us. People have life-long friendships, many marriages have silver jubilees and we often go to the same doctor over and over again. "Trusted business partner" or "reliable brand name"—these are commonly heard phrases. At the same time, acrimonious separations are not altogether rare either. Business partners often end up in court and many marriages end in divorce. Our model displays this co-existence of long-term bonds and quick separations. In fact it is this co-existence which justifies each phenomenon separately.

equilibrium (a precise definition of which follows in the next section), patient players are seen to offer an "experimental" level of cooperation to newly met opponents, reciprocation of which serves as a signal that the opponent is non-myopic. Pairs of patient players go on to form long-term relationships once they have successfully revealed their types to each other through such experimental cooperation. The equilibrium is characterized by the phenomenon of "gradual trust-building"—any long term relationship involves a low, initial level of cooperation (when players are uncertain about the other's type), which increases to a higher level when the initial phase is successfully passed without termination of the relationship. The character of play between two non-myopic members *in equilibrium* resembles that along an Abreu *punishment path*[5]—payoffs are low at the beginning, but increase over time. In the latter phase of a long term partnership, two non-myopic players can sustain a relatively high level of cooperation by disciplining each other through a *termination threat*—a threat to break off the relationship if the cooperative arrangement is not adhered to. The temptation to achieve temporary gains by deviating from cooperation unilaterally is then held in check by the cost of termination. It is precisely the presence of myopic types that contributes to this cost—a player who loses a non-myopic partner will find it somewhat difficult to find a new one, and even if one is found, the process of *gradual* build up of cooperation has to be started all over again.

The equilibrium we propose is robust against two kinds of deviation: first, no *individual* (non-myopic) player can achieve a higher payoff by deviating, and second, no *pair* of non-myopic players playing each other can propose a Pareto improving yet incentive compatible *joint* deviation (this is the criterion of "bilateral rationality" in Section 2). It is this criterion which makes the presence of myopic types crucial in justifying the phenomenon of *gradual* cooperation build-up, and ultimately in explaining cooperation in the absence of individual reputations. In Datta (1993), a model similar in flavour to ours (but without incomplete information about types) is constructed, and the existence of equilibria marked by slow cooperation build up is shown. However, such equilibria are immune only against individual deviations, and not against pairwise deviations, as we require them to be. In contexts where pre-play dialogue and communication between players is possible, it is difficult to see why this criterion should be ignored.

The main results of the paper consist in developing the implications of increasing the degree of patience in this framework. Note that in the present context, we may talk about *two* different notions of patience in the community—the discount factor of *individual* patient players, and the proportion of such patient players in the community. We find that going by the first notion, a rise in patience is unambiguously "good" as it leads to a rise in cooperation levels in *all* phases of a long-term relationship between a pair of patient players. This is quite in tune with results from standard theory. What is striking, however, is that going by the second notion of patience mentioned above, the results are almost exactly the opposite. A community which has a *greater* proportion of patient players is generally able to sustain *less* cooperation among its patient members; this is *always* true of the latter phase (the "friendship phase") of any long-term relationship, and sometimes true even of the former phase (the "stranger phase"). In this sense, "intrinsic cheats" exert a *positive* externality on those who are potentially cooperative. We show, however, that a greater abundance of patient players in the community can only *increase* the *expected payoff* of a patient member who has yet to form a long-term partnership with another patient player. In other words, the effect of falling cooperation levels is always more than offset by the increased *chance* of meeting a patient player in any given period.

5. See Abreu (1986, 1988).

The rest of the paper is organized as follows. In Section 2, we describe the basic model and discuss the concept of a social equilibrium. In Section 3, the properties of a social equilibrium are explored. In 3.1, we examine its existence and uniqueness; in 3.2, we prove the characteristic of *rising* cooperation levels within any long-term relationship; in 3.3, we note the effect of changes in the parameters (specifically, the *proportion* of patient people, and their *degree* of patience) on equilibrium cooperation levels. In Section 4, two examples with specific functional forms are presented, to make transparent some of the issues involved. Some extensions of the basic model are discussed as well. In Section 5, we relate our model to the existing literature. Section 6 concludes. Rigorous proofs for most of the propositions are in the Appendix.

2. THE MODEL

2.1. *Description of the game*

The stage game played bilaterally has common action set A and symmetric payoff functions for both players. In this paper we identify A with some compact interval $[0, \tilde{a}] \subset \mathbb{R}$. Assume that this game has a single (strictly) dominant strategy, which will be identified with the action 0. Normalize the payoff under the dominant strategy equilibrium to zero.

Let $\Pi(a, a')$ denote the payoff to a player in the stage game when he has chosen the action a, while his opponent has chosen a'. Since the action 0 is assumed to be a strictly dominant strategy for the stage game, it follows that $\Pi(0, a') > \Pi(a, a')$ for all $a, a' \neq 0$. To economize on notation, we introduce the following functions. Let $v(a)$ denote the common payoff when both players play the action a, let $l(a)$ be the best possible payoff to a player when his opponent plays a, and let $d(a)$ denote the payoff to a player who plays a but whose opponent plays the best response (action 0). In terms of the more primitive payoff function, $v(a) = \Pi(a, a)$, $d(a) = \Pi(0, a)$, and $l(a) = \Pi(a, 0)$. We maintain the following assumption on the payoff functions:

Assumption 1. The functions $v(a)$, $d(a)$ and $l(a)$ are continuous in a. In addition $v(a)$ is strictly increasing in a.

We will often refer to a as a "cooperation level" so that higher cooperation levels are beneficial to both players. However, it is in the interest of each player *individually* to extract as much cooperation from the other as possible, while extending no cooperation himself (i.e. choosing action 0). The unique Nash equilibrium of this stage game is, therefore, both players choosing not to cooperate at all and getting zero payoffs. The stage game can hence be looked upon as an extended version of the Prisoner's Dilemma.

The stage game described above is played repeatedly by pairs of players in the population. Some of these are players continuing an old relationship. At each date, pairings occur between currently unmatched players. After the stage game has been played in every period, each player in any pair has an option: either to "continue" the previous relationship, or to "terminate" it. If both players agree to continue, they play the stage game between themselves once more. If, however, even one player decides to terminate, then their relationship is broken, and they return to the pool of unmatched players. At the next date, the same story is repeated.

We assume that at any date, the *pool of unmatched players* consists of a fraction π of non-myopic types. In this paper, we take π to be constant over time. This is done to keep the exposition simple. It is, strictly speaking, true in our model if at any date the

stock of fresh players is very large relative to the number playing the game. There is another interpretation of π as a steady state of an extended game, which we discuss in Section 4.2. Readers who feel uncomfortable about the assumption of a constant π might refer to that section before moving on.

Players come in two types. A fraction of the population is concerned about the future: these players have a discount factor $\delta \in (0, 1)$. The rest are "short-run players": they do not care about the future and play only myopic best responses (i.e. choose complete non-cooperation, or action 0) to their current opponent's action. In other words, these myopic players have a discount factor of zero. A player's "personal history" at any date is the record of his play in all previous periods (i.e. the actions taken by him and his, possibly various, opponents in the past, as well as the continuation/termination choices made by them at the end of each period). Since we assume that there is no information flow within the community, all the information a player has at the beginning or end of any period is his personal history up to that date. The strategy set of each player is, therefore, the set of all possible mappings from personal histories to the action set A, and to the set consisting of the continuation/termination options as well.

2.2. *Social norms and social equilibrium*

By a "social norm", we mean a strategy or behaviour rule which members of the community may follow. A behaviour rule should specify, contingent on a player's personal history, what actions he should choose and, further, when to terminate an on-going relationship and find a new opponent. An example of a social norm is the behaviour rule: "Always play full cooperation (play action $\tilde{a}$) and *never* terminate a relationship yourself". For reasons that are quite obvious, such a norm cannot be expected to gain wide acceptance. Our interest, therefore, lies in identifying norms that induce an *equilibrium* (in a sense to be made precise shortly) in the community at large: *given* that others are following the norm, it is best for an individual player (or a pair of players playing each other) to do likewise. The basic issue involved is this: can we construct sensible norms that involve some degree of cooperative play? If so, what are the features of such cooperation, and how does it respond to changes in the parameters of the model (e.g. the proportion of myopic types present)?

Two things need to be mentioned here. First, our model has a heterogeneous population: the same behaviour rule cannot be expected to be adhered to by both myopic and non-myopic types. The behaviour of myopic players is easy to describe—they always play the strictly dominant strategy of the stage game (i.e. non-cooperation, or, the action 0). Thus, whenever we talk of social norms or behaviour rules, the reference is to the set of non-myopic players alone.

Second, in concentrating on equilibria constituted by social norms, we essentially focus on symmetric equilibria (symmetric in the class of players of the same type) in pure strategies. For this reason, we have non-existence of equilibrium in our model for some parameter values (specifically, in cases where the proportion of patient players is "too high"; see Proposition 2). The problem of existence can be removed by allowing for mixed strategies, whether at the individual level or at the group level[6], and by relaxing the criterion of "bilateral rationality" (see below) that we stipulate for our concept of a "social equilibrium". However, we feel that equilibria constituted by simple and universally

6. As in the literature on evolutionary games, the probability weight on a strategy can be interpreted as the *proportion* of the population (of a given type) playing the specified strategy.

accepted behaviour conventions have a great deal of intuitive appeal, and are natural focal points we ought to study. This is further reinforced by our result (see Proposition 1) that when parameter values permit the existence of such an equilibrium, it is *unique*.

In what follows, we shall examine social norms which stipulate certain actions or cooperation levels and a "termination rule" which instructs a player to terminate a relationship if the stipulated cooperative action is not played by the other party. Social norms which, if accepted in the population at large, satisfy certain incentive constraints (to be mentioned below), will be said to constitute *social equilibria* for the game we discuss. There is a relationship in our model between individual actions and the expected payoff of an unmatched player in the society at large—each influences the other. Our notion of equilibrium calls for consistency and balance in this two-way feedback mechanism; the word "social" is a reminder to this effect.

It follows from our assumption regarding types that if a social norm prescribes some cooperation at any date, and this play is actually adhered to by a player, he reveals himself to be non-myopic. If this is the case for both the partners, it is natural to imagine that they will want to continue their bilateral relationship, since with myopic players all one can achieve is the dominant strategy equilibrium. Thus, it makes sense to look for a social equilibrium that divides each player's behaviour into two phases. The first phase, to be called Phase S, deals with play between *strangers*, two newly matched players. In phase S, these players "test" each other to check each other's type. Such a test must take the form of a stipulated, "experimental" level of cooperation—adherence to or deviation from this level will reveal a player's type. If the test is passed by both players, then a second phase of *friendship*, Phase F, is entered. In this phase, each non-myopic player seeks to cooperate, secure in the knowledge that the other player is non-myopic. Of course, if even one player fails the test, the players separate and pass into a repetition of Phase S with a new set of partners.

Behaviour in either phase must be individually incentive-compatible. For instance, in Phase F, it will be impossible, in general, to expect a level of cooperation that might prevail were the same two players to be interacting with each other in complete isolation. This is because there are bounds on the punishments that can be inflicted on the players; in particular, each player always has the option to deviate, terminate the relationship, and then enter into a new Phase S with another player. This leads us to propose the first criterion for a social equilibrium—the standard *individual* no-deviation condition:

Individual incentive constraint. A social norm is said to satisfy the individual incentive constraint if, given that other non-myopic players are believed to be following the norm, no individual non-myopic player can achieve a higher expected payoff by choosing a strategy other than that specified by the norm.

Of course, there are usually many equilibria that satisfy this condition, including trivial ones. For instance, the prescription that dictates perennial non-cooperation (choice of the dominant strategy outcome) certainly constitutes an equilibrium, and the division into phases S and F becomes essentially arbitrary. However, when one expands the notion of incentive-compatibility to include the partnership of two players, one might easily imagine that faced with such a trivial norm, two players will find it in their collective interests to test for non-myopia, proceeding thereafter to a cooperative phase if the test is passed. If many pairs of players do this, of course, then the earlier norm will soon be

eroded. It makes sense, therefore, to test these norms against some notion of *group rationality*, at least at the bilateral level. This leads us to propose a second criterion for a social equilibrium:

Bilateral rationality. A social norm is said to satisfy bilateral rationality if, given that all other players are following the norm, no matched pair of players, who have not deviated in their past arrangements with each other, can propose a joint change from the norm that both increases their expected payoffs and satisfies the individual incentive constraint.

Thus, for instance, perennial non-cooperation will generally fail bilateral rationality. However, observe that we restrict the application of bilateral rationality to situations where the players have not yet deviated on arrangements with each other. In short, we avoid issues of renegotiation-proofness[7] *within* a partnership in Phase F.

A word of caution is in order here. Application of the notion of bilateral rationality, *as described above*, may be problematic and involve a possible ambiguity when applied to Phase S, i.e. when players' types are not common knowledge.[8] The problem is this: should bilateral rationality dictate, subject to the individual incentive constraint, maximization of the expected payoff of the non-myopic type, or that of the myopic type, or some weighted average of the two? As will be clear from the mathematical formulation that follows, we make the first choice, i.e. we assume that a bilaterally rational behaviour rule for the S-Phase is one which achieves the constrained best expected payoff for the non-myopic type. A justification for this assumption is given by the following informal argument.

In "equilibrium", no non-myopic player will be interested in the deviation payoff; hence, any player who proposes to put some positive weight on the myopic player's payoff will necessarily reveal himself to be myopic.[9] Clearly, no player, regardless of his type, will make such a proposal, since it will destroy the other player's incentive to cooperate.[10]

Now, bilateral rationality enters in two ways: once in Phase S, and once in Phase F. Indeed, as we shall see, there is a tension between its effects in the two phases. Imagine that an equilibrium is in place, and denote by V^S and V^F the present discounted values to a non-myopic player in Phases S and F respectively, normalized as time averages by multiplying by $(1-\delta)$. Consider, first, a pair of players who have been fortunate to discover that they have non-myopic partners. They are in Phase F. Now that they are there, they wish to maximize cooperative possibilities (bilateral rationality). In other words, they solve the following problem:

$$\max_{a \in A} v(a) \tag{1}$$

7. See, e.g. Bernheim and Ray (1989) and Farrell and Maskin (1989). It is possible to extend the model to include renegotiation at this level using asymmetric strategies during Phase F. By moving along the constrained frontier, termination for one of the players can be mimicked. We do not consider these issues as in our opinion, they stray from the main point of the paper.

8. We are grateful to Yeon-Koo Che and an anonymous referee for bringing this point to our attention.

9. This line of argument is very much in the spirit of Kohlberg and Mertens' notion of strategic stability; see Kohlberg and Mertens (1986).

10. The problem here is quite subtle—it involves the issue of negotiation when there is asymmetric information regarding payoffs. In such a situation, the proposals that players make have possible information content, and are hence strategic variables themselves. We do not formalize this pre-game dialogue and negotiation process, nor do we claim to have resolved the general theoretical problem involved—that is beyond the modest aim of this paper. However, we believe that the formulation adopted here is sensible for this particular context, on the intuitive grounds that we have sketched above.

subject to the constraint

$$v(a) \geqq (1-\delta)d(a) + \delta V^S. \tag{2}$$

The constraint (2) is the individual incentive constraint: by deviating from a cooperative arrangement, a player obtains a current return of $d(a)$. The agreement is then broken off, so that he can now look forward to a present value of V^S after the rupture. Of course, δ may be so small that even the standard two-person repeated game permits no cooperation, but then this is just the trivial case where all players are effectively myopic. We avoid this by making the following assumption on parameters and payoff functions (remember that $v(0)$ is normalized to 0):

Assumption 2. There exists $a \neq 0$ such that

$$v(a) > (1-\delta)d(a).$$

Under this assumption, it follows that there exists $\hat{V}^S > 0$ such that a solution to (2) (and therefore (1)) exists if and only if $V^S \leqq \hat{V}^S$. Note, moreover, that as V^S increases, the value of cooperation in Phase F must *fall*, which is the source of the tension between Phases S and F.

As stated above, bilateral rationality also applies to Phase S, in the same way as it applies to Phase F. Strangers may be strangers, but that is no reason for partners in Phase S to make each other undergo unnecessarily unpleasant experiences (incentive compatible or not), only to prove their non-myopic nature. There may be (and indeed, often are) relatively more pleasant ways of doing the same thing. In other words, the "experimental" level of cooperation which two newly matched players would propose to each other should be chosen so as to maximize the *expected* payoff from such cooperation to a non-myopic player (which each of the players either is or pretends to be). So we may formalize Phase S behaviour in the following way:

$$\max_{a \in A, a \neq 0} \pi[(1-\delta)v(a) + \delta V^F] + (1-\pi)[(1-\delta)l(a) + \delta V^S] \tag{3}$$

subject to the constraint

$$\pi[(1-\delta)v(a) + \delta V^F] + (1-\pi)[(1-\delta)l(a) + \delta V^S] \geqq \pi(1-\delta)d(a) + \delta V^S. \tag{4}$$

To understand (3) and (4), observe first that myopic players always play best responses so there is no need to consider directly the behaviour of such players. Now consider a pair formed by two strangers. During their initial association, they might either choose to play the dominant strategy (in which case no information is conveyed regarding their true type) or they might agree on an incentive-compatible course of play, which, if validated, will reveal their types. Observe that for a non-myopic player in Phase S, the present value of an agreement to play a is precisely that given by the maximand in (3). With probability π the agreement is upheld and play then proceeds to phase F in the next period. With the remaining probability, however, his partner is myopic and proceeds to deviate. Current payoff to the agent is then only $l(a)$, followed by a return to Phase S. This explains the maximand in (3).

However, for a particular value of (3) to be achievable, it must be that the agreement a is incentive-compatible for a non-myopic player. Such a player might entertain a deviation from a. If he does, his best payoff is given precisely by the RHS of (4). With probability π, he enjoys a current payoff of $d(a)$ (with the remaining probability $1-\pi$, both deviate to the dominant strategy from which the payoff is zero), followed by a return to Phase S.

This deviation payoff should not exceed the expected present value of the agreement, which is exactly what is expressed by (4). Thus, what we capture by (3) and (4) is the idea that strangers attempt to "cooperate" to the maximum extent possible, subject to individual incentive constraints.[11]

Phase S may display the unfortunate feature that with strangers there is not any value to current cooperation, whether incentive-compatible or not. This is especially so if the percentage of myopic types is very high. We will rule out this case by making the assumption that a small amount of cooperation is always possible (though we do not presume, *ex ante*, that such cooperation will always be incentive compatible):

Assumption 3. For all $\varepsilon > 0$, there is $a < \varepsilon$ such that $\pi v(a) + (1-\pi) l(a) > 0$.[12]

We are now in a position to define a *social equilibrium*. It is a collection of actions (a^S, a^F) and payoffs (V^S, V^F) such that

1. a^F solves (1) subject to (2).
2. a^S solves (3) subject to (4).
3. V^F equals the maximum value attained in (1), and
4. V^S equals the maximum value attained in (3).

3. PROPERTIES OF THE SOCIAL EQUILIBRIUM

3.1. *Existence and uniqueness of a social equilibrium*

A social equilibrium, as defined above, need not exist, though under some additional assumptions it is possible to completely characterize those situations where existence is guaranteed. On the other hand, uniqueness is guaranteed under general conditions: there can be at most one social equilibrium. The rest of this subsection is devoted to elaborating on these points, as well as to collecting some observations that will be useful in the analysis to follow.

It will be convenient for the exposition to construct a mapping, a fixed point of which describes a social equilibrium. To this end, we consider different values of V^S, to be denoted by x. Recall from the discussion following (2) that if $x \in [0, \hat{V}^S]$, problem (1) is well defined. So restrict x in what follows to lie in this interval. For each such x, we may define $V^F(x)$ as the solution to problem (1) (subject to (2), with x in place of V^S). Note that $V^F(x) > x$: simply inspect (2) and recall that for each $a \neq 0$, $d(a) > v(a)$.

11. A non-myopic player may consider, in the S Phase, deviating from the proposed action a to some other action a', where $0 < a' < a$, thereby increasing his single-period expected payoff, while still signalling a non-myopic type. If such a deviation occurs, then by the intuitive criterion (see Cho and Kreps (1987), and also Kohlberg and Mertens (1986)) or other similar equilibrium refinement notions, the opponent should believe that the deviant is non-myopic with probability 1. Termination is thereafter in the interest of neither player, if both happen to be non-myopic, in other words, problems of renegotiation crop up. This problem can be overcome by constructing punishment phases *within* Phase F, with asymmetric strategies and payoffs, in the manner suggested in footnote 7. Alternatively, one may assume that each player, whether cooperative or non-cooperative, has a small probability of committing a mistake in choosing his desired action, and in the case of a mistake, his choice has a distribution over the action set that is identical for all agents. On seeing a stranger choosing an action $a' \neq a$, therefore, a player does not revise his probability beliefs about this opponent's type, and is therefore indifferent between playing him again and playing some other randomly drawn player. The tie-breaking assumption may be made that he terminates relationship in such cases.

12. This assumption is satisfied automatically if, for instance, $l'(0) = 0$, so that small levels of cooperation do not impose a first-order loss in the presence of a deviant. More generally, $v'(0)/|l'(0)| = \infty$ is sufficient as well.

Armed with the values x and $V^F(x)$, turn to the S-phase problem in (3). We may rewrite that problem by changing the notation appropriately and by manipulating the expressions a bit:

$$\max_{a\in A, a\neq 0} (1-\delta)[\pi v(a)+(1-\pi)l(a)]+\delta[\pi V^F(x)+(1-\pi)x] \tag{5}$$

subject to the constraint

$$[d(a)-v(a)]-\frac{1-\pi}{\pi}l(a)\leqq\frac{\delta}{1-\delta}[V^F(x)-x]. \tag{6}$$

Note that for small a, the condition (6) must always hold, because at $a=0$ (the dominant strategy), all LHS values equal zero, and the RHS value is strictly positive, as argued above. We have assumed, moreover, that cooperation is indeed possible for small values of a. Therefore, a solution exists to this problem. Denote by $\phi(x)$ the maximum value of (5) thus attained: this is the mapping that we wish to examine.

Note that social equilibria bear a one-to-one relationship with the set of fixed points of ϕ. The following observation reveals that $\phi(\cdot)$ is always flatter than the 45°-line. The uniqueness of equilibrium is then a standard corollary.

Lemma 1. *Let $x>x'$. Then $\phi(x)-\phi(x')<x-x'$.*

Proof. Define $z(x)\equiv\max_{a\in A, a\neq 0}\pi v(a)+(1-\pi)l(a)$, subject to (6). Then, using (5),

$$\phi(x)=(1-\delta)z(x)+\delta[\pi V^F(x)+(1-\pi)x].$$

Observe from problem (1) that $V^F(x)$ is a non-increasing function of x, and therefore, so is $z(x)$ (inspect the constraint (6) and observe that it becomes tighter as x goes up). Therefore, for $x>x'$,

$$\begin{aligned}\phi(x)-\phi(x')&=(1-\delta)[z(x)-z(x')]+\delta\pi[V^F(x)-V^F(x')]+\delta(1-\pi)[x-x']\\&\leqq\delta(1-\pi)(x-x')\\&<x-x',\end{aligned}$$

which completes the proof. ‖

Lemma 1 yields the following easy corollary.

Proposition 1. *A social equilibrium, if it exists, must be unique.*

Existence is a different matter altogether. The problem is that the function ϕ has jumps in it, not to mention the fact that it fails to be well defined once x crosses the threshold value $\hat{V}^S$. These jumps are always "downward" (by Lemma 1): see Figure 1.

There is a subclass of interesting models for which we can offer necessary and sufficient conditions for existence of a social equilibrium. Make the following set of assumptions on the curvature of the payoff, gain and loss functions. This assumption will be utilized in the following proposition (Proposition 2), and in Proposition 5A, but nowhere else.

Assumption C. The function $v(a)$ is strictly concave, the function $l(a)$ is concave, and the function $d(a)$ is convex.

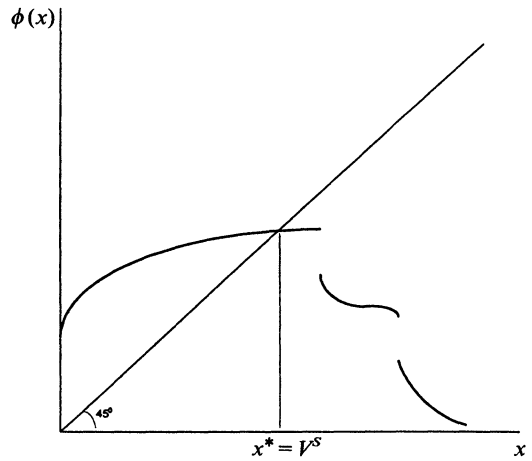

FIGURE 1
Jumps in the ϕ function

These assumptions imply that the payoff increases with the action vector, but at a decreasing rate, while the deviation payoff and the (absolute value of the) loss increases at an increasing rate.[13]

We shall now provide necessary and sufficient conditions for the existence of a social equilibrium in this model. While the conditions are placed directly on the parameters of the model, they are nevertheless a bit cumbersome, so we will need some notation. Later, we provide an intuitive description.

First denote by a^1 the (unique) maximizer of the function $v(a)-(1-\delta)d(a)$ on the set A. Next, note that by Assumption (C),

$$h(a) \equiv [d(a)-v(a)]-\frac{1-\pi}{\pi}l(a)$$

is an increasing function, 0 at 0. Let a^2 denote the (unique) value of a for which $h(a)$ equals $d(a^1)-v(a^1)$. Finally, let a^3 denote the (unique) maximizer of the strictly concave function $\pi v(a)+(1-\pi)l(a)$ on A.

Now consider the following condition

Condition E. If $a^3 \leqq a^2$, then

$$\pi v(a^3)+(1-\pi)l(a^3) \leqq \left(\frac{1}{\delta}+\pi\right)v(a^1)+\left(1-\frac{1}{\delta}-\pi\right)d(a^1). \tag{7}$$

Otherwise

$$\delta\pi d(a^2) \leqq v(a^1)-(1-\delta)d(a^1). \tag{8}$$

13. It is also easy to check that under Assumption C, the functions are monotone in a, in particular, $d(a)$ is increasing and $l(a)$ is decreasing in a. This follows from the fact that all three functions have a value 0 when evaluated at 0, $d(a)$ is positive and $l(a)$ is negative for all $a>0$.

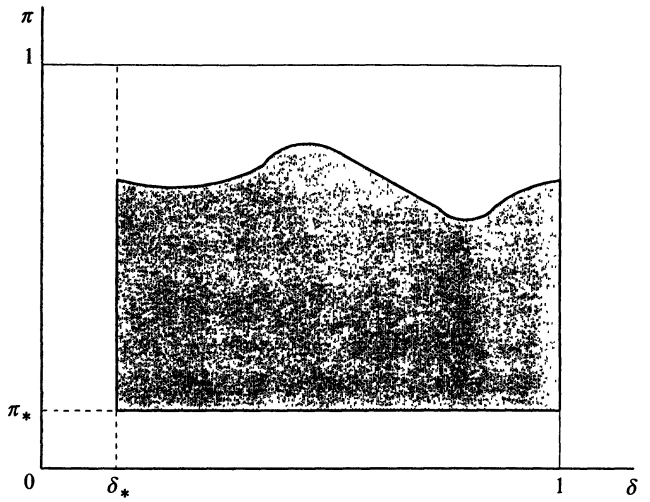

FIGURE 2
Parameters satisfying Condition E

We can now state

Proposition 2. *Under the additional Assumption* C, *a social equilibrium exists if and only if Condition* E *holds.*

Proof. See Appendix. ||

As the condition appears far from transparent, a few remarks on what it implies may not be out of place. To understand the kind of restriction that is implied, suppose first that Condition E is satisfied for some value of π, say π', and consider a lower value, say π''. Then Condition E continues to be satisfied (the details are relegated to a footnote[14]). Thus it is a *high* proportion of non-myopic individuals that puts the existence of social equilibrium at risk, an issue to which we return in the next section. Intuitively, a high proportion of non-myopic individuals permits a relatively large present value payoff in Phase S, destroying the possibilities of cooperation in Phase F, and leading to non-existence.

We may put together this observation with Assumptions 2 and 3, which imply lower bounds on the parameters δ and π. The joint effect of Assumptions 2 and 3, and Condition E, are summarized in Figure 2.

The values δ_* and π_* are lower bounds imposed to satisfy Assumptions 2 and 3. The curve represents the upper limit of π, for each δ, that is consistent with existence. In line with the argument made above, it follows that the shaded area represents the parametric zone that permits the existence of a social equilibrium.

14. Suppose that the first part of Condition E held for π'. Then it continues to hold for π'', because the LHS of (7) comes down, a^1 is independent of π, and the RHS of (7) is a decreasing function of π. Likewise, if the condition $a^3 > a^2$, and (8), held before, *and* the condition $a^3 > a^2$ continues to hold at π'', then (8) must also hold for π'' (because a^2 is easily seen to be increasing in π). It only remains to consider the case where the reduction in π induces a switch from part 1 of (E) to part 2, but this is easily handled by simply noting that conditions (7) and (8) are exactly the same when $a^3 = a^2$.

In particular, it should be observed that the upper bound on π is strictly less than one, and that *this effect persists even if we were to take the discount factor of non-myopic individuals to unity*. To see this, pass to the limit in Condition E as $\delta \to 1$. Note first that for δ close enough to 1, the maximizer a^1 of $v(a) - (1-\delta)d(a)$ must equal the maximum possible action $\tilde{a}$ in A. Now imagine, contrary to our claim, that there is a sequence of values of π (as $\delta \to 1$) satisfying (E), and converging to 1. In this case, the values of a^2 and a^3 *also* converge to $\tilde{a}$, as casual observation of their definitions will easily reveal. Consider a sub-sequence such that either (7) holds throughout or (8) holds throughout. In the former case, passing to the limit in (7), we see that $v(\tilde{a}) \leqq 2v(\tilde{a}) - d(\tilde{a})$, a contradiction because $d(\tilde{a}) > v(\tilde{a})$. In the latter, passing to the limit in (8), we see that $d(\tilde{a}) \leqq v(\tilde{a})$, again a contradiction.

Moreover, it is possible (though Figure 2 as drawn does not reflect this feature) that certain values of the discount factor are not compatible with equilibrium even if higher and lower values are so compatible. This would occur if the bounding curve were to dip below the lower bounds described by δ_* and π_*.

This interpretation of Condition E also shows us that under some assumptions, the concept that we are studying is far from vacuous. For instance, if $l'(0) = 0$, then π_* equals 0 as well, and there always exists a (connected) set of (δ, π) such that an equilibrium exists.[15] Section 4.1 discusses an example in detail.

In Sections 3.2 and 3.3, we shall have more to say about the dependence of social equilibrium on the value of π, as well as about the effects of low discounting.

3.2. *Equilibrium values in Phases S and F*

The first observation that we make about social equilibrium is that Phase F must involve higher per-period value than Phase S, *even if cooperation were to occur in Phase S*. Of course, if we omit the emphasized clause in the previous sentence, this observation must be trivially correct. After all, if V^S were not to fall short of V^F, there is no way that cooperation can be sustained in Phase F. Individuals would deviate without any fear of punishment. The point is, however, that the basic agreement chosen in Phase S cannot dominate (in value) the agreement of Phase F, even supposing that the former agreement is adhered to. In other words, the cooperation level stipulated for the S phase will usually be lower than that for the F phase.

Proposition 3. *In any social equilibrium, $v(a^F) \geqq v(a^S)$ (hence $a^F \geqq a^S$), with strict inequality holding whenever $v(a^F)$ is not the maximum possible symmetric payoff in the game.*

Proof. See Appendix. ‖

Proposition 3 establishes that a social equilibrium is characterized by a testing phase in which players are "cautious", so that there is less to gain in this phase. If this phase is passed successfully, then one moves into a phase of greater cooperation.

An additional observation is relevant here. Note that the extent of cooperation in the testing phase is limited by the presence of myopic types in the population. It is this

15. More generally, it will be the case that $\pi_* = -l'(0)/(v'(0) - l'(0))$. Then, if $v(\tilde{a}) > -l'(0)/(v'(0) - l'(0))d(\tilde{a})$, there exists a range $[\hat{\delta}, 1)$ and $\pi(\delta) > \pi_*$ on $[\hat{\delta}, 1)$ such that for all $\delta \in [\hat{\delta}, 1)$ and all $\pi \in (\pi_*, \pi(\delta))$, an equilibrium exists.

limitation, coupled with the intrinsic uncertainty about an opponent's type in Phase S, that permits higher levels of cooperation in Phase F. The presence of myopic types might therefore act as an inducement to cooperation (see Proposition 4 below). It is this presence which rationalizes "slow cooperation build-up", and by driving a wedge between the payoffs to be expected in the F and the S phases, ultimately justifies any kind of cooperation itself. In a community where *every* person is equally patient, an equilibrium with slow trust-building can yet be imposed (see Datta (1993)), but in such an equilibrium, every *pair* of players has an incentive to renegotiate and move to the higher cooperation level right from the start, as long as *others* stick to the norm. This, of course, destroys the norm itself. Hence, it is our insistence on the criterion of bilateral rationality that makes the presence of myopic members a crucial driving force in the model.

3.3. *Changes in the proportion of non-myopic types*

Let us explore the last observation of the previous subsection a bit further. One way to do so is to increase parametrically the percentage of non-myopic types in the economy, and examine the effects of this change on the extent of cooperation. Once we see the intuition that the presence of impatient types is what serves to maintain cooperation, the outcome is exactly as expected. Evaluated at the old values in phases S and F, it is now possible, in phase S, to obtain a higher present value return. This occurs because the chances of being "cheated" in phase S are reduced. But this, in turn, means, that the maximum degree of cooperation that one can hope for in phase F must *fall*. This, of course, feeds back to reduce the value in phase S, which includes phase F as a future contingency. But standard "stability-type" arguments should be sufficient to convince one that this latter reduction cannot compensate for the initial increase. We may summarize all this as

Proposition 4. *An increase in the proportion of non-myopic types, π, must decrease the equilibrium present value in phase F (i.e. V^F) and raise it in Phase S (i.e. V^S) (provided that an equilibrium exists in both these cases). The cooperation level in phase F (i.e. a^F) must also go down.*

Proof. See Appendix. ‖

Proposition 4 argues that long-term cooperation in *on-going* relationships falls with an increase in the number of patient types in the community. On the other hand, the proposition also says that when patient players are more abundant in the population, each such player should expect a higher *expected* payoff when in phase S. Nothing can be inferred from this about whether the actual *cooperation level* in phase S goes up or falls in response to the rise in π. In fact, as π increases, a rise in V^S is consistent with the possibility that a^S falls, as long as the fall is not too large. Indeed, we are able to show that the effect on the S phase cooperation level can go either way. In particular, if non-myopic players are very patient (have discount factors close to 1), then an increase in the fraction of such players in the population may lead to a decrease in the cooperation level *even in the S Phase* (and hence in both phases together).

Proposition 4A. *Assume that $d(a)$ is non-decreasing in a. Consider $\pi_0 \in (0, 1)$ such that for all values of δ close enough to 1, an equilibrium exists whenever π is within ε of π_0, for some $\varepsilon > 0$. Further, suppose that in every such equilibrium, the incentive constraint in* (4)

is binding. Then, for δ sufficiently close to one, a small increase in the value of π_0 will lead to a decrease in the equilibrium value of a^S.

Proof. See Appendix. ‖

To confirm that the condition under which the result stated in Proposition 4A is obtained is not vacuous, see Example 2 in Section 4.1 below.

Observe that we can have two notions of "patience" in the framework of our model. One is the *size* of the non-myopic population. The other notion of patience is, of course, the standard one—the discount factor of *each* non-myopic individual. We have examined the effect of an increase in patience in the community on equilibrium cooperation levels and expected payoffs, going by the first notion mentioned above. We now turn to the same investigation, following the second notion of patience.

3.4. *Changes in the discount factor*

Like an increase in π, raising the discount factor has the effect of increasing the expected value in phase S. Equilibrium present value in that phase rises, *after* normalizing for the increase in the discount factor. Curiously enough, and *unlike* the counterpart result for changes in π, equilibrium present value in Phase F must also increase. It is perhaps worth emphasizing why we find the second observation surprising. After all, it is generally the case (in standard models) that increases in the discount factor make cooperation more likely. However, in the framework under consideration, there are two opposing effects. The first is the natural farsightedness induced by a rising discount factor: individuals value the gains from a single deviation less, and this induces a greater tendency towards cooperation. But there is a second effect: cooperation in Phase S is enhanced too, and this *reduces* the effective punishment in the case of a deviation. It turns out that in this model, the former effect always dominates the latter. An inspection of the proof below will show that the argument establishing this domination is quite subtle.[16]

Proposition 5. *Assume an equilibrium exists throughout. An increase in the discount factor δ must raise (normalized) equilibrium present value in Phase S (i.e. V^S). This is also true of the equilibrium present value (V^F) as well as the cooperation level (a^F) in phase F provided that the a^F is not already at the maximum level of $\tilde{a}$, in which case V^F and a^F are left unchanged. Further, in the limit, as $\delta \to 1$, the value of a^F converges to $\tilde{a}$, the maximum cooperation level.*[17]

Proof. See Appendix. ‖

Once again, we have said nothing about the movement in a^S in response to changes in δ in the above proposition. By imposing additional structure on the payoff functions (i.e. under Assumption C), we can prove the following

16. It is, in fact, unclear to us whether this property will survive extensions to other models in this general class of games.

17. We are grateful to Kunal Sengupta for pointing out the result on the limit behaviour of a^F, and for suggesting the proof.

Proposition 5A. *Assume an equilibrium exists throughout. If Assumption* C *holds, there exists* $\delta^* \in (0, 1]$ *such that an increase in* δ *leads to an increase in* a^S *for* $\delta < \delta^*$. *If* $\delta \geqq \delta^*$, *a further increase in* δ *has no effect on the value of* a^S *in equilibrium.*

Proof. See Appendix. ‖

Note how Propositions 5 and 5A stand in stark contrast to the observations in Propositions 4 and 4A. There, increasing the *number* of patient types certainly reduces phase F cooperation and may reduce cooperation in phase S as well. In the current case, however, neither a^S nor a^F can fall, and generally will rise. Moreover, a folk theorem type of result holds—as non-myopic players become infinitely patient, they can sustain the maximal cooperation level in the F Phase, *irrespective of the proportion of myopic players in the population.*

4. EXAMPLES AND EXTENSIONS

4.1. *Examples*

We first present two examples. The first one satisfies all the assumptions of the paper, including Assumption C, over the entire parameter range. We show how social equilibrium may be computed, and in the process describe Condition E. The second example is designed to illustrate the possibility that was asserted in Proposition 4C—with a rise in the fraction of patient players in the population, cooperation levels in Phase S may go down.

Example 1. Consider pairs of players who have to supply inputs to each other's production process. Thus, individual i, to produce an output of value a, needs some input from another player j, the cost of the necessary amount of input to the latter being $\frac{1}{2}a^2$. Player j has symmetric requirements, to be procured from player i. Thus the payoff function between any two players is given by $\Pi(a, a') = a' - \frac{1}{2}a^2$.

The corresponding reduced form functions are then as follows: $v(a) = a - \frac{1}{2}a^2$, $d(a) = a$, and $l(a) = -\frac{1}{2}a^2$.

Note that irrespective of incentive constraints, players will never find it optimal to strike a deal of choosing a outside the closed interval $[0, 1]$, so we take $\tilde{a} = 1$. Note that for all $(\delta, \pi) \in (0, 1)^2$, Assumptions 1–3 and Condition C are satisfied.

To compute the social equilibrium, we construct the mapping $\phi(x)$, and find its fixed point. Thus, we set $V^S = x$ in the F Phase problem, and note that $a^F(x)$ is the larger of at most two solutions to the following equation (which is obtained by making the incentive constraint bind).

$$\delta a - \tfrac{1}{2}a^2 = \delta x.$$

On explicitly solving, we get

$$a^F(x) = \delta\left[1 + \sqrt{1 - \frac{2x}{\delta}}\right]. \tag{9}$$

Note that a real solution exists only for $x \leqq \delta/2$, so that the fixed point must lie in the range $[0, \delta/2]$. For $x \in [0, \delta/2]$, we see that

$$V^F(x) = x + \frac{1-\delta}{2\delta} a^F(x)^2.$$

Using the functional forms of $v(a)$, $d(a)$ and $l(a)$, the S Phase maximization problem can be expressed as follows

$$\max_{a \in [0,\,1]} [\pi a - \tfrac{1}{2}a^2]$$

subject to

$$a^2 \leqq \pi a^F(x)^2.$$

It is easy to see that the solution is given by

$$a^S(x) = \min \{\pi, \sqrt{\pi} a^F(x)\}. \tag{10}$$

Putting these values in the expression for V^S, we obtain the mapping ϕ:

$$\phi(x) = (1-\delta)[\pi a^S(x) - \tfrac{1}{2}a^S(x)^2] + \tfrac{1}{2}(1-\delta)\pi a^F(x)^2 + \delta x \tag{11}$$

We will use (11) to directly check for existence. Note that ϕ is continuous on $[0, \delta/2]$. So the necessary and sufficient condition for existence is the condition

$$\phi(\delta/2) \leqq \delta/2. \tag{12}$$

Two cases will be considered:

Case 1. $\pi \leqq \delta^2$. In this case, $a^F(\delta/2) = \delta$ (by (9)) and so $a^S(\delta/2) = \pi$, by (10). Using this information in (12), an equilibrium exists if and only if

$$\pi^2 + \delta^2 \pi - \delta \leqq 0.$$

Solving this inequality, we observe that in this case

$$\pi \leqq \tfrac{1}{2}[\sqrt{\delta^4 + 4\delta} - \delta^2] \tag{13}$$

is necessary and sufficient for existence.

Case 2. $\pi > \delta^2$. Then $a^F(\delta/2) = \delta$, as before, but $a^S(\delta/2) = \delta\sqrt{\pi}$. Using this information in (12), we obtain after simplification the condition

$$\pi \leqq (\tfrac{1}{2})^{2/3}. \tag{14}$$

The combination of (13) and (14) yields an explicit form[18] for Condition E:

$$\begin{aligned} &\pi \leqq (\tfrac{1}{2})^{2/3} && \text{if } \delta \leqq (\tfrac{1}{2})^{1/3}, \\ &\pi \leqq \tfrac{1}{2}[\sqrt{\delta^4 + 4\delta} - \delta^2] && \text{if } \delta > (\tfrac{1}{2})^{1/3}. \end{aligned} \tag{15}$$

18. To derive this condition, it will be useful to take note of the fact that $\tfrac{1}{2}[\sqrt{\delta^4 + 4\delta} - \delta^2] \geqq \delta^2$ if and only if $\delta \leqq (\tfrac{1}{2})^{1/3}$. Now, a systematic consideration of the two zones $\delta \leqq (\tfrac{1}{2})^{1/3}$ and $\delta > (\tfrac{1}{2})^{1/3}$, along with (13) and (14), yields (15).

Simple numerical evaluation of (15) shows that the critical value of π permitting existence is flat at approximately 63% as long as δ does not exceed approximately 0·8. Thereafter, the critical value of π drops steadily, though marginally, to approximately 61·8%.

Example 2. This example illustrates Proposition 4A, which argues that cooperation levels in Phase S might decline with an increase in π. Consider the following payoff function: $\Pi(a, a') = (1+\lambda)a' - a$, where $a \in [0, 1]$ and $\lambda > 0$.

Then the reduced-form functions are as follows: $v(a) = \lambda a$, $d(a) = (1+\lambda)a$, and $l(a) = -a$.

For cooperation to be obtained in the repeated game, we need $\lambda a > (1-\delta)(1+\lambda)a$ for $a > 0$. This implies the following lower bound on the discount factor:

$$\delta > \frac{1}{1+\lambda} \tag{16}$$

which is just Assumption 2. To ensure that π satisfies Assumption 3, we need $\pi\lambda a - (1-\pi)a \geqq 0$ for $a > 0$. This implies that

$$\pi \geqq \frac{1}{1+\lambda}. \tag{17}$$

The assumed linearity of the payoff structure implies that phase F cooperation must be maximal, if at all feasible. That is, $a^F = 1$ in a social equilibrium with cooperation. Then the maximum value of V^S that is compatible with some cooperation in Phase S is given by the solution to $\lambda = (1-\delta)(1+\lambda) + \delta V^S$, which yields

$$\hat{V}^S = \frac{1}{\delta}[\lambda - (1-\delta)(1+\lambda)]. \tag{18}$$

Thus, $\phi(x)$ drops to zero at $x = \hat{V}^S$. The search for a social equilibrium, provided one exists, has to be restricted to the domain $x \in [0, \hat{V}^S]$. For any x in this domain, $a^F(x) = 1$, and hence, $V^F(x) = \lambda$. Now suppose $a^S(x) < 1$, i.e. partial cooperation occurs in Phase S. (We shall shortly derive the parametric restrictions under which this will be the case.) Putting $V^F(x) = \lambda$ in the S Phase incentive constraint, and observing that this constraint must be binding for the value of a^S to be less than 1, we derive

$$a^S(x) = \frac{\delta\pi(\lambda - x)}{1-\delta}. \tag{19}$$

Inserting the values of $a^F(x)$ and $a^S(x)$ in the expression for V^S, we obtain the expression for $\phi(x)$ as follows:

$$\phi(x) = (1-\delta)\left[\frac{\delta\pi^2\lambda(\lambda - x)}{1-\delta} - \frac{\delta\pi(1-\pi)(\lambda - x)}{1-\delta}\right] + \delta\pi(\lambda - x) + \delta x.$$

The condition $\phi(x^*) = x^*$ describes a social equilibrium, so that

$$x^* = \frac{\delta\pi^2\lambda(1+\lambda)}{(1-\delta) + \delta\pi^2(1+\lambda)}. \tag{20}$$

Now for the restrictions that permit (20) to be derived. We need, first, that $x^* \leqq \hat{V}^S$, which implies, using (18) and (20), that

$$\delta\pi^2(1+\lambda) \leqq \lambda\delta - (1-\delta). \tag{21}$$

Finally, the above assumed that $a^S < 1$. To find conditions that guarantee this, put $a^F = a^S = 1$, and calculate the value of V^S. If this value is greater than $\hat{V}^S$, then $a^S = 1$ is not possible in the social equilibrium, and we are done. Under $a^F = a^S = 1$,

$$V^S = \frac{\pi\lambda - (1-\delta)(1-\pi)}{1-\delta(1-\pi)},$$

and the requirement that this value must exceed $\hat{V}^S$ yields

$$\delta(1-\pi)(2+\lambda) < 1. \tag{22}$$

We may substitute $x = x^*$ in (19) to obtain the equilibrium value of a^S, provided the parametric restrictions in (21) and (22) are satisfied.

$$a^S = \frac{\delta\pi\lambda}{(1-\delta)+\delta\pi^2(1+\lambda)}.$$

For a^S to be decreasing in π, we need that the derivative of the right-hand side in the above expression be negative, which amounts to the condition

$$\pi^2 > \frac{1-\delta}{\delta(1+\lambda)}. \tag{23}$$

Thus, for any values of the parameters satisfying the inequalities (16), (17), (21), (22) and (23), it will be true that a social equilibrium exists and that a^S falls with a small increase in π.

The reader can check that the values $\lambda = 2$, $\delta \geqq 0{\cdot}9$ and $9/16 < \pi^2 < 17/27$ are consistent with all the above restrictions.

4.2. *Some extensions*

4.2.1. *Endogenous proportion of non-myopic types*

One problematic assumption in the model we have presented is the constancy of π—the fraction of non-myopic players in the pool of unmatched players who are subject to random matching. For obvious reasons, the rates at which the two types exit from the pool of unmatched players are different in equilibrium. Apart from the fact that players' strategies both within and outside a long-term partnership would necessarily become non-stationary in this situation, the incentive to cooperate with a stranger may be impossible to sustain beyond a certain point.

Our assumption of the constancy of π, however, was made for convenience of exposition. Our model may be viewed as studying the *steady state* of an extended model, in which (at least one of) two new features are present—the population grows at a rate n, and there is an exogenous probability θ that a "good" partnership breaks up for reasons outside the scope of the model. Under these circumstances, the composition of the pool

evolves according to the following difference equation:

$$\pi_{t+1}=1-\frac{(1+n)(1-\pi^*)}{(1-\pi^*)+n+\theta[1-(1-\pi^*)\pi_t]+(1-\pi^*)\pi_t} \tag{24}$$

where π_t denotes the fraction of non-myopic players in the pool of unmatched players at date t, and π^* denotes the *genetic* proportion of non-myopic players, relevant for the initial population as well as any batch of fresh arrivals.

To understand (24), normalize date t population to unity, and note that the total measure of myopic players at date $t+1$, who will all be in the unmatched pool, is $(1+n)(1-\pi^*)$. It remains to explain the denominator of (24), the total population of unmatched players at date $t+1$. This is done by counting four components: (i) all the myopic players at time t, (ii) all fresh arrivals, (iii) all breakups at date t, and (iv) all the date t non-myopic players who were matched with myopic types. The accounting details are tedious but simple, and are omitted.

It is easy to see that *if* (24) is followed over time, the proportion π_t must converge to a unique, strictly positive steady state,[19] which we may express as a function of the parameters: $\pi=\pi(\pi^*, n, \theta)$. It is straightforward to check that π is increasing in all three of its arguments.

The analysis of equilibrium norms and behaviour in the *steady state* of this extended model will then run exactly as in the model we have presented, with fixed π. The expression for V^F and V^S and the incentive constraints will have to be modified slightly to take into account the break-up probability θ. We do not conduct such an analysis, but it is easy to see that the results will be similar. Since π is increasing in π^*, the genetic fraction, the comparative static results of an increase in π can be interpreted as the comparative statics of an increase in π^*, which is a primitive of the model.

However, it must be pointed out that this approach does not offer much of a clue as to the nature of the *non-stationary* equilibrium, when π_t is away from its steady state value. The reason is that (24) is constructed on the presumption that the social equilibrium takes a particular form. We have not been able to verify whether this presumption holds up during the "transition dynamics" of the model, and leave this case as an interesting open question.

4.2.2. *The case of more than two types*

Imagine that there are a large number of possible discount factors in the population, not merely two. In that case, instead of a single Phase S, there will be in general several such phases, with each phase characterized by the degree of mutual knowledge that players have about each other. Observe that in the natural generalization of our equilibrium, such knowledge would be increasing in these phases, thus permitting successively higher levels of cooperation as we move from one phase to the next.

19. To see this, note first that the needed end-point restrictions are satisfied: $\pi_{t+1}>\pi_t$ if π_t is small, with the opposite inequality holding if π_t is close to one. Next, observe that

$$\frac{d\pi_{t+1}}{d\pi_t}=\frac{(1+n)(1-\pi^*)^2(1-\theta)}{[(1-\pi^*)+n+\theta+(1-\pi^*)(1-\theta)\pi_t]^2}>0.$$

Moreover,

$$\frac{d\pi_{t+1}}{d\pi_t}<\frac{(1+n)(1-\pi^*)^2(1-\theta)}{[(1-\pi^*)(1+n)+\theta+(1-\pi^*)(1-\theta)\pi_t]^2}<\frac{(1+n)(1-\pi^*)^2(1-\theta)}{[(1-\pi^*)(1+n)]^2}<1.$$

This verifies the assertion.

Thus far the analysis is straightforward. However, it will also be necessary to describe a "stopping rule" for each type, describing the minimum (revealed) discount factor of the partner that he would be content with. This creates an interesting (though by no means insurmountable) complication that we leave as another open question. Such an extended model will accord more closely with reality, where the process of trust-building is often painstakingly slow and gradual.

5. RELATED LITERATURE

5.1. *Theoretical*

There is a large literature on cooperation problems, particularly in the context of interaction in large populations. For a theory based on the hypothesis of discernible emotional pre-dispositions, see Akerlof (1993). Frank (1988) elaborates on this theory and generalizes it, giving it an evolutionary flavour. For a discussion in evolutionary biology on "the battle of the sexes" (the problem of cooperation between male and female parents over the care of offspring), see Dawkins (1989, Chapter 9).

Among papers within the framework of rational agents, Kandori (1992) suggests, in the context of random matching games, a process of "contagious defection" to be sparked off by a single deviation, that helps to discipline individual players. As pointed out there, the process works for small groups, but not for large populations, which are of central concern here. Ellison (1994) elaborates further on the disciplining effect of "contagious" punishments in an anonymous population with random matching.

Datta (1993) takes a broadly similar approach to ours, by endogenizing quit decisions and proposing termination threats and paths of rising cooperation as the elements that sustain cooperative practices in the absence of information flows within a community. Our difference from Datta, as pointed out earlier, lies in our stress on population heterogeneity and the uncertainty arising from incomplete information regarding types, and in our insistence on "bilateral rationality" as an equilibrium selection criterion. Matsushima (1990) studies a similar setting with the 2×2 Prisoner's Dilemma as the stage game, and shows that cooperation is possible in equilibrium. However, since there is no incomplete information regarding types in Matsushima's model, the equilibrium proposed fails to satisfy "bilateral rationality".[20]

The work of Sobel (1985) is also relevant in this context. He constructs a model of repeated interaction between a "sender" and a "receiver", where the sender's interest may be either completely coincident with the receiver's or completely opposed, but the receiver is uncertain about the sender's type (i.e. payoff function). The model illustrates the importance of incomplete information regarding other players' payoffs in generating the phenomenon of gradual trust building. However, the emphasis in our paper is on the use of such paths as implicit punishments. In fact, in Sobel's model, uncertainty regarding the other player's interests *creates* the cooperation problem in the first place, whereas in ours, that uncertainty helps to *solve* the problem for players who are potentially cooperative.

20. While preparing the final version of this paper, we also became aware of the related work of Kranton (1995) and Watson (1995). Kranton, like us, emphasizes the role of incomplete information in creating norms that involve the gradual build-up of cooperation robust to pairwise deviations, and quotes sociological evidence to illustrate the incidence of long term relationships and group formation. Watson considers similar issues. He compares the value of "starting small" as a device to signal a cooperative type to other signalling mechanisms such as "money burning", and comes to the conclusion that the former can, in many situations, be more effective than the latter.

There is also a literature in political science which emphasizes how an "exit" option can, in many contexts, be more practicable and effective than the "deviation" option used as part of a punishment strategy in standard repeated Prisoner's Dilemma games. For example, Vanberg and Congleton (1992) show, by conducting Axelrod-type tournaments,[21] that programmes which make use of the "exit" option to punish deviant behaviour often fare better than those which exercise the "deviate" option, the reason for such evolutionary fitness being that it imposes less cost on the punisher. Tullock (1985) also expresses ideas along similar lines. However, the payoff from exit is taken to be *exogenously* given in these papers, and our contribution lies in endogenizing this value, and in showing how different factors and parameters (such as the composition of the population) can influence this value, and how that in turn determines the level of cooperation achieved by cooperative members.

There are other papers, tied to more specific contexts, that are relevant to our work. The foremost among these is Shapiro and Stiglitz (1984), which explains how involuntary unemployment may arise as an equilibrium phenomenon out of firms' efforts to remove workers' incentives to shirk, in a situation where workers may be fired but not stigmatized. This story can be accommodated in a formal game-theoretic model by making the *number* of relationships a player can form a choice variable, and making the assumption that the possible marginal payoff from a relationship is strictly diminishing in the number of relationships. In contrast, our model takes the number of relationships a player can form to be exogenous, and focuses on a *different* mechanism through which cooperative behaviour can be sustained in a similar set-up—this mechanism derives its strength from endogenously formed long-term bonds with rising cooperation paths which lend weight to termination threats.

The two-way feedback in our model between individual decisions and the "social climate" is also a feature in some other recent papers. Sah (1992) discusses a similar link between individuals' choice of crime and their estimates of detection probabilities. Tirole (1996) examines issues of corruption and bribe-taking in society, with imperfect individual reputations, and shows the possibility of multiple equilibria involving both high and low levels of corruption.

Finally we should mention that the stage game discussed in this paper has a specific structure, which may not be most suitable in analysing all kinds of social interaction. More to the point, the stage game here has the feature of *two-sided* moral hazard, and by imposing symmetry in payoffs and in equilibrium strategies, we effectively give equal bargaining power to each party in any relationship. In a related paper (Ghosh and Ray (1995)), we analyze a one-sided moral hazard problem in the context of competitive informal credit markets, with limited access to borrower credit histories. We show that a similar mechanism as here acts to discipline agents, though additional issues arise (e.g. over the division of surplus) and some effects of parametric changes are reversed.

5.2. *Empirical*

As was mentioned in the introduction to this paper, informal rural credit markets in many developing countries seem to present real life situations that fit closely (though not exactly—see the discussion in the last paragraph of the previous section) our paradigm. Moneylenders often face the difficult task of ensuring repayment of their loans in the absence of access to legal remedy or individual reputations. Aleem (1993) and Bell (1993)

21. See Axelrod (1984).

discuss case studies from India and Pakistan that bear out some of the crucial features of our model. Both authors claim that default rates with informal moneylenders are far lower than that with formal sector banks, which suggests some subtle incentive mechanism at work in the informal credit sector. Both authors note and lay stress on the phenomenon of *segmentation* of the rural credit market—moneylenders build up a more or less fixed group of clients through repeated interaction, and are usually extremely reluctant to take on new ones. Even when they do, a new client is initially screened by giving him a small "testing loan" not enough to cover his needs. Loan sizes are later increased conditional on timely repayment of previous loans (see Aleem (1993)).[22] This accords exactly with the features of long-term bonding and *gradual* increase of cooperation levels in the interaction between two given players that are present in our model.

6. CONCLUSION

We have presented a model of repeated pairwise interaction within a large population of players, where the stage game between any two players has the characteristics of a Prisoner's Dilemma and hence, poses a problem of cooperation. We have shown that social norms that prescribe cooperative behaviour can be sustained in equilibrium in the presence of partner-switching options for players and in the presence of a fraction of "myopic players" who play only short-run best responses to their opponent's strategy. The equilibrium proposed is robust against both individual as well as pairwise deviations. We showed that the equilibrium is always marked by *gradual* cooperation build-up—cooperation levels *rise* over time in any long-term partnership. We also follow a common theme in the repeated games literature and examine the effect of an increase in levels of patience in the community—both in the sense of an increase in the discount factor of *individual* non-myopic players, and in the sense of an increase in the *proportion* of non-myopic players in the community. We find that in the first sense, an increase in patience leads to a rise in equilibrium cooperation levels. However, going by the second notion of patience, we find that an increase in patience leads to a *fall* in long-term cooperation levels within a given partnership, and sometimes in cooperation levels all through. This is in sharp contrast to the theory of repeated games between a *fixed* set of players, where more patience is seen to be unambiguously "good" for the prospect of cooperation.

APPENDIX

Proof of Proposition 2. Recall the construction of $\hat{V}^S$ and ϕ. We first observe that under the additional assumption (C), $\phi(x)$ is continuous for all $x \in [0, \hat{V}^S]$. To see this, first consider problem (1) (subject to (2)), rewritten here as $\max_{a \in A} v(a)$, subject to $v(a) \geqq (1-\delta)d(a) + \delta x$. Using Assumption C, it is easy to check that $v(a)$ is either $v(\tilde{a})$, or is given by the larger of (at most) two solutions to the equality version of the constraint (if $\tilde{a}$ is not a feasible choice). This is continuous in x, so $V^F(x)$ is continuous.

Turn now to the problem described in (5) (subject to (6)). Recalling the definition of $h(a)$, (6) is just the condition that $h(a) \leqq \delta[v^F(x) - x]/(1-\delta)$. Let $a(x)$ be the maximum value of a that satisfies this constraint. Because the objective in (5) is concave in a and because $h(a)$ is increasing, it follows that the solution to (5) (subject to (6)) is $a^S(x) \equiv \min\{a^2, a(x)\}$. This is continuous in x. Consequently,

$$\phi(x) = (1-\delta)[\pi v(a^S(x)) + (1-\pi)l(a^S(x))] + \delta[\pi V^F(x) + (1-\pi)x]$$

22. These points need to be qualified and further examined. For instance, a small initial loan may not be a test for patience, but of the capacity to repay. A specific analysis of this issue would lead to a different model, though the overall ideas will be similar. We are grateful to Karla Hoff for pointing this out to us.

is continuous in x. Therefore, by Lemma 1, it follows that ϕ has a fixed point (equivalently, a social equilibrium exists) if and only if

$$\phi(\hat{V}^S) \leqq \hat{V}^S. \tag{25}$$

The remainder of the proof simply consists in checking that (25) is equivalent to Condition E. We begin by noting that

$$\hat{V}^S = \frac{1}{\delta} \max_{a \in A} [v(a) - (1-\delta)d(a)] = \frac{1}{\delta}[v(a^1) - (1-\delta)d(a^1)]$$

so that

$$\begin{aligned} V^F(\hat{V}^S) - \hat{V}^S &= v(a^1) - \frac{1}{\delta}[v(a^1) - (1-\delta)d(a^1)] \\ &= \frac{1-\delta}{\delta}[d(a^1) - v(a^1)]. \end{aligned}$$

Consequently, if $x = \hat{V}^S$, the problem (5) (subject to (6)) may be expressed as

$$\max_{a \in A, a \neq 0} (1-\delta)[\pi v(a) + (1-\pi)l(a)] + [1 - \pi(1-\delta)]v(a^1) - (1-\pi)(1-\delta)d(a^1) \tag{26}$$

subject to the constraint

$$h(a) \leqq d(a^1) - v(a^1), \tag{27}$$

and note that the maximum value attained in (26) is exactly $\phi(\hat{V}^S)$. Now recall a^2 and a^3 as defined for Condition E. Observe that the solution to (26), subject to (27), is precisely a^3 if $a^3 \leqq a^2$, and a^2 otherwise (this uses the observations that $\pi v(a) + (1-\pi)l(a)$ is concave in a and that $h(a)$ is increasing in a). In the former case, (25) is equivalent to the condition

$$\begin{aligned} &(1-\delta)[\pi v(a^3) + (1-\pi)l(a^3)] + [1 - \pi(1-\delta)]v(a^1) - (1-\pi)(1-\delta)d(a^1) \\ &\leqq \frac{1}{\delta}[v(a^1) - (1-\delta)d(a^1)] \end{aligned}$$

which on simplication is equivalent to (7), the first part of Condition E.

Finally, in the latter case ($a^3 > a^2$), using the constraint (27) with equality, (25) is equivalent to the condition

$$\begin{aligned} &(1-\delta)[\pi d(a^2) - \pi\{d(a^1) - v(a^1)\}] + [1 - \pi(1-\delta)]v(a^1) - (1-\pi)(1-\delta)d(a^1) \\ &\leqq \frac{1}{\delta}[v(a^1) - (1-\delta)d(a^1)], \end{aligned}$$

which on simplification is equivalent to condition (8), the second part of Condition E. This completes the proof. ‖

Proof of Proposition 3. Observe, first, that for all $a \neq 0$, we have $l(a) < 0$. This is a consequence of our assumption that there is a single (strict) dominant strategy, and our payoff normalization, so that for all $a \neq 0$, $l(a) = \Pi(a, 0) < \Pi(0, 0) = 0$.

If $v(a^F)$ is the maximum possible feasible payoff, then of course we are done. So look at the remaining case. Suppose the proposition is false. Then $v(a^S) \geqq v(a^F)$. But then, because $v(\cdot)$ is increasing, it must be the case that the constraint (2) with a^S in place of a^F holds with the opposite (weak) inequality: $v(a^S) \leqq (1-\delta)d(a^S) + \delta V^S$ (otherwise this contradicts the fact that a^F solves (1) subject to (2)). Using this information,

$$\begin{aligned} (1-\delta)[d(a^S) - v(a^S)] &\geqq \delta[v(a^S) - V^S] \\ &\geqq \delta[v(a^F) - V^S] \\ &= \delta[V^F - V^S] \\ &> \delta[V^F - V^S] + \frac{1-\pi}{\pi} l(a^S)(1-\delta) \end{aligned}$$

where the last inequality uses the fact that in any social equilibrium, it must be the case that $a^S \neq 0$, and the observation that $l(a) < 0$ for all $a \neq 0$. Inspection of this inequality reveals that it runs counter to (4). We therefore have a contradiction, and the proposition is established. ‖

Proof of Proposition 4. Carry π explicitly in the definition of ϕ, writing it as $\phi(x, \pi)$. Observe that $\hat{V}^S$ and, more generally, $V^F(x)$ is independent of the value of π. Now fix x, and examine the problem (5) (subject to (6)) as π increases. Recalling that $l(a)$ is negative, we see that the constraint (6) is relaxed, and moreover, the value of the objective (5) is increased for each feasible choice of a. It follows, therefore, that $\phi(x, \pi)$ is increasing in π. Using Lemma 1, and noting that $V^F = v(a^F)$, the proposition follows. ‖

Proof of Proposition 4A. Since (4) is assumed to be binding,

$$d(a^S) - v(a^S) - \frac{1-\pi}{\pi} l(a^S) = \frac{\delta}{1-\delta}[V^F - V^S].$$

The left-hand side above is bounded. Hence, so is the right-hand side. We conclude that for fixed $\pi^0 \in (0, 1)$, $(V^F - V^S) \to 0$ as $\delta \to 1$.

Next, we claim that as long as (4) is binding in equilibrium,

$$\pi d(a^S) = V^S. \tag{28}$$

To see this, recall that the left-hand side of (4) is equal to V^S in equilibrium, so that if (4) is binding, $V^S = \pi(1-\delta)d(a^S) + \delta V^S$. Manipulating this yields (28).

Now consider an increase in the value of π by η, where $\eta < \varepsilon$. Take a value of δ sufficiently close to 1, so that $V^F - V^S < \eta d(a^S)$ at $\pi = \pi_0$. By Proposition 4, as π increases from π_0 to $\pi_0 + \eta$, the value of V^F is lower, so that the increment in V^S must be less than $\eta d(a^S)$. Let V_0^S and a_0^S be the values of V^S and a^S respectively when $\pi = \pi_0$, and V_η^S and a_η^S be the corresponding values when $\pi = \pi_0 + \eta$. Then, by (28),

$$(\pi_0 + \eta)d(a_\eta^S) - \pi_0 d(a_0^S) = V_\eta^S - V_0^S$$

which implies

$$(\pi_0 + \eta)[d(a_\eta^S) - d(a_0^S)] + \eta d(a_0^S) < \eta d(a_0^S)$$

from which it follows that $d(a_\eta^S) < d(a_0^S)$. Since $d(a)$ is nondecreasing, we conclude that $a_\eta^S < a_0^S$. ‖

Proof of Proposition 5. Let $0 < \delta < \hat{\delta} < 1$. Denote by (V^S, V^F, a^S, a^F) and $(\hat{V}^S, \hat{V}^F, \hat{a}^S, \hat{a}^F)$ the two equilibria. We must show that $\hat{V}^S > V^S$ and $\hat{V}^F \geqq V^F$, with strict inequality holding if $V^F < v(\bar{a})$.

We start by establishing the proposition for Phase S.

In what follows, we will carry δ explicitly in all relevant functions. By Lemma 1, it will suffice to prove that $\phi(V^S, \hat{\delta}) > \phi(V^S, \delta)$. For ease of notation, let $x = V^S$. Note first, from problem (1), that $V^F(x, \hat{\delta}) > V^F(x, \delta)$, so that in particular

$$\frac{\hat{\delta}}{1-\hat{\delta}}[V^F(x, \hat{\delta}) - x] \geqq \frac{\delta}{1-\delta}[V^F(x, \delta) - x].$$

Therefore the constraint in (6) is not tighter as δ changes to $\hat{\delta}$. Consequently, denoting by $\bar{a}^S$ the solution to (5) (subject to (6)) at $\hat{\delta}$, we see that

$$\pi v(\bar{a}^S) + (1-\pi)l(\bar{a}^S) \geqq \pi v(a^S) + (1-\pi)l(a^S). \tag{29}$$

Now inspect the new maximum value in (5). Note first that $v(a^S) \leqq v(a^F) = V^F(x, \delta)$ (Proposition 3) and $l(a^S) < 0 \leqq x$, so that

$$\begin{aligned}(1-\hat{\delta})[\pi v(a^S) + (1-\pi)l(a^S)] + \hat{\delta}[\pi V^F(x, \delta) + (1-\pi)x] \\ > (1-\delta)[\pi v(a^S) + (1-\pi)l(a^S)] + \delta[\pi V^F(x, \delta) + (1-\pi)x].\end{aligned} \tag{30}$$

Moreover, as already noted, $V^F(x, \hat{\delta}) > V^F(x, \delta)$. Combining this last observation with (29) and (30), we see that

$$\begin{aligned}\phi(x, \hat{\delta}) &= (1-\hat{\delta})[\pi v(\tilde{a}^S) + (1-\pi) l(\tilde{a}^S)] + \hat{\delta}[\pi V^F(x, \hat{\delta}) + (1-\pi)x] \\ &> (1-\delta)[\pi v(a^S) + (1-\pi) l(a^S)] + \delta[\pi V^F(x, \delta) + (1-\pi)x] \\ &= \phi(x, \delta),\end{aligned}$$

which completes the first part of the proof.

We now turn to the proposition for Phase F. Because $v(\cdot)$ is assumed to be strictly increasing, we see at once from (1) and (2) that a^F is the *maximum* value (over $[0, \tilde{a}]$) satisfying the inequality

$$v(a^F) \geqq (1-\delta) d(a^F) + \delta V^S \tag{31}$$

for V^S given at the equilibrium value. Now observe from (5) that in equilibrium,

$$V^S = \frac{(1-\delta)[\pi v(a^S) + (1-\pi) l(a^S)] + \delta \pi v(a^F)}{1-\delta(1-\pi)}. \tag{32}$$

Combining (31) and (32) and simplifying, we see that

$$v(a^F) \geqq \frac{1-\delta(1-\pi)}{\delta}[d(a^F) - v(a^F)] + [\pi v(a^S) + (1-\pi) l(a^S)]. \tag{33}$$

The following lemma will be needed.

Lemma 2. *If (a^F, a^S) are equilibrium values under δ, then, given a^S, a^F is the maximum value of a in $[0, \tilde{a}]$ satisfying* (33).[23]

Proof. Suppose not. Then there is $a' > a^F$ (so $v(a') > v(a^F)$) such that

$$v(a') \geqq \frac{1-\delta(1-\pi)}{\delta}[d(a') - v(a')] + [\pi v(a^S) + (1-\pi) l(a^S)]. \tag{34}$$

Define

$$V'^S \equiv \frac{(1-\delta)[\pi v(a^S) + (1-\pi) l(a^S)] + \delta \pi v(a')}{1-\delta(1-\pi)}. \tag{35}$$

Then, from (34) and (35) (following exactly the opposite route to arrive at (33) from (31) and (32)),

$$v(a') \geqq (1-\delta) d(a') + \delta V'^S.$$

But it is obvious that $V'^S > V^S$. Consequently,

$$v(a') \geqq (1-\delta) d(a') + \delta V^S,$$

which contradicts the fact that a^S is the maximal solution to (31), given V^S. This proves the lemma. ‖

We now return to the main proof. First, recall the constraint (6). In any equilibrium,

$$[d(a^S) - v(a^S)] - \frac{1-\pi}{\pi} l(a^S) \leqq \frac{\delta}{1-\delta}[V^F - V^S]. \tag{36}$$

Note that if $a^F < \tilde{a}$, then (31) holds with equality. In that case, combining (31) (with equality) and (36), we see that

$$[d(a^S) - v(a^S)] - \frac{1-\pi}{\pi} l(a^S) \leqq d(a^F) - v(a^F). \tag{37}$$

First consider the case $V^F < v(\tilde{a})$. Suppose, contrary to the statement of the proposition, that $\hat{V}^F \leqq V^F$. Then, we see that (37) holds for both (a^F, a^S) and $(\hat{a}^F, \hat{a}^S)$.

We claim that $d(\hat{a}^F) - v(\hat{a}^F) \leqq d(a^F) - v(a^F)$.

23. Note well that this claim is different from (31). (31) is true for a *given* value of V^S. The point is that $v(a^F)$ also appears in the formula for V^S, which was used to obtain (33).

Suppose not. Then

$$d(\hat{a}^F) - v(\hat{a}^F) > d(a^F) - v(a^F). \tag{38}$$

Using (33) for the case of $\hat{\delta}$, we see that

$$v(\hat{a}^F) \geqq \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}} [d(\hat{a}^F) - v(\hat{a}^F)] + [\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S)]. \tag{39}$$

Using (38) and $\hat{V}^F \leqq V^F$ in (39),

$$v(a^F) > \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}} [d(a^F) - v(a^F)] + [\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S)]. \tag{40}$$

But using the fact that $v(\cdot)$ is increasing, and that $v(\cdot)$ and $d(\cdot)$ are continuous, (40) implies that there exists a such that $v(a) > v(a^F) \geqq v(\hat{a}^F)$ and

$$v(a) \geqq \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}} [d(a) - v(a)] + [\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S)],$$

which contradicts Lemma 2 for the equilibrium under $\hat{\delta}$. This proves the claim.

Therefore, recalling that (37) holds for both equilibria in the case under consideration, we see that

$$\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S) \leqq \pi v(a^S) + (1-\pi) l(a^S). \tag{41}$$

Use (41) and the fact that $\hat{\delta} > \delta$ in (33) to obtain

$$v(a^F) > \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}} [d(a^F) - v(a^F)] + [\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S)].$$

But then, just as before, there exists a with $v(a) > v(a^F) \geqq v(\hat{a}^F)$ and

$$v(a) \geqq \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}} [d(a) - v(a)] + [\pi v(\hat{a}^S) + (1-\pi) l(\hat{a}^S)],$$

which contradicts Lemma 2, and completes the proof in this case.

Finally, suppose that $V^F = v(\tilde{a})$. The proof in this case follows the same lines with some variations. Suppose, contrary to the statement of the proposition, that $\hat{V}^F < V^F$. Then (37) holds for the equilibrium under $\hat{\delta}$:

$$[d(\hat{a}^S) - v(\hat{a}^S)] - \frac{1-\pi}{\pi} l(\hat{a}^S) \leqq d(\hat{a}^F) - v(\hat{a}^F), \tag{42}$$

while (36) holds for the equilibrium under δ.

We claim that $d(\hat{a}^F) - v(\hat{a}^F) \leqq (\delta/1-\delta)[V^F - V^S]$.

To prove the claim, first use (2) to observe that

$$V^F = v(a^F) \geqq (1-\delta) + \delta V^S$$

so that

$$\frac{\delta}{1-\delta} [V^F - V^S] \geqq d(a^F) - v(a^F). \tag{43}$$

Given (43), then, it will suffice to establish that

$$d(\hat{a}^F) - v(\hat{a}^F) \leqq d(a^F) - v(a^F).$$

Suppose on the contrary that

$$d(\hat{a}^F) - v(\hat{a}^F) > d(a^F) - v(a^F). \tag{44}$$

Using (44), along with our supposition that $v(\hat{a}^F) < v(a^F)$, in the version of (33) for $\hat{\delta}$ (see (39)), we get (40) again. But because $v(a^F) > v(\hat{a}^F)$, (40) contradicts Lemma 2 for the equilibrium under $\hat{\delta}$. This proves the claim.

To complete the proof, use the claim above, together with (36) and (42) to obtain (41) again. Use (41) and the fact that $\hat{\delta} > \delta$ in (33) to obtain, just as before,

$$v(a^F) > \frac{1-\hat{\delta}(1-\pi)}{\hat{\delta}}[d(a^F) - v(a^F)] + [\pi v(\hat{a}^S) + (1-\pi)l(\hat{a}^S)]. \tag{45}$$

But (45) contradicts Lemma 2 for the equilibrium under $\hat{\delta}$, and the proof is complete for the first part of the Proposition.

The proof for the limit behaviour of a^F is as follows. Suppose, on the contrary, that as $\delta \to 1$, a^F converges to some value $a^* < \tilde{a}$. Then, by bilateral rationality, it must be true that for any $\delta < 1$, $v(\tilde{a}) < (1-\delta)d(\tilde{a}) + \delta V^S$. However, Proposition 3 implies that $V^S < V^F = v(a^F)$, so that the above inequality can be rewritten as $v(\tilde{a}) < (1-\delta)d(\tilde{a}) + \delta v(a^F)$. Taking the limit of the right-hand side as δ goes to 1, and noting that $\lim_{\delta \to 1} a^F = a^*$, we have $v(\tilde{a}) \leq v(a^*)$, which, given that $v(a)$ is strictly increasing, contradicts our assumption that $a^* < \tilde{a}$. ‖

Proof of Proposition 5A. The following lemma will be useful in the proof.

Lemma 3. *Suppose Assumption* C *holds. Then, for given values of* V^F, V^S, π *and* δ, *the feasible set in phase S defined by* (4) *is a non-empty closed interval* $[0, \bar{a}]$, *where* $\bar{a}$ *can be written explicitly as a function of given values:* $\bar{a} = \bar{a}(V^F, V^S, \pi, \delta)$.

Moreover, let $\psi^S(V^F, V^S, \pi, \delta)$ *denote the value of a that maximizes* (3) *subject to* (4). Then

$$\psi^S(V^F, V^S, \pi, \delta) = \min\{a^*(\pi), \bar{a}(V^F, V^S, \pi, \delta)\},$$

where $a^*(\pi) = \arg\max_{a \in A}[\pi v(a) + (1-\pi)l(a)]$.

Proof. First, note that maximizing (3) subject to (4) is equivalent to:

$$\max_{a \in A} \pi v(a) + (1-\pi)l(a)$$

subject to the constraint

$$[d(a) - v(a)] - \frac{1-\pi}{\pi} l(a) \leqq \frac{\delta}{1-\delta}[V^F - V^S]. \tag{47}$$

Examine the left-hand side of (47). Note that $[d(a) - v(a)]$ is a continuous and strictly convex function, 0 at 0 and strictly positive otherwise. It follows that $[d(a) - v(a)]$ is strictly increasing in a. Hence, the entire left-hand side of the above inequality is increasing in a, while the right-hand side is independent of a. The feasible set is therefore an interval, as claimed.

Now, since $a^*(\pi)$ solves the unconstrained maximization problem, it must be the solution to the constrained problem if $a^*(\pi)$ satisfies the constraint, i.e. if $a^*(\pi) \leqq \bar{a}(V^F, V^S, \pi, \delta)$. If this condition fails, then (47) must bind, because the objective function is strictly concave. So in this case, the solution to the problem is given by $\bar{a}(V^F, V^S, \pi, \delta)$, and the proof is complete. ‖

Now we complete the proof of the proposition. Begin with some value of δ for which $a^S = \psi(V^F, V^S, \pi, \delta) = \bar{a}(V^F, V^S, \pi, \delta) \neq a^*(\pi)$ in equilibrium. So (47) is binding, and recalling the argument leading up to (28) (see proof of Proposition 4A),

$$V^S = \pi d(a^S) \tag{48}$$

Now consider a higher value of δ. By Lemma 3, either the new value of a^S is $a^*(\pi)$ which is independent of δ), in which case a^S has increased, or it is still less than $a^*(\pi)$, in which case (47) is still binding, so that (48) above still holds. By Proposition 5, the left-hand side (V^S) must go up due to a rise in δ, and hence it follows that $d(a^S)$ must go up too. Because $d(a^S)$ is increasing, we conclude that a^S must increase in this latter case as well.

Now consider an initial value of δ such that $a^S = a^*(\pi)$. Then (47) holds with (weak) inequality. By an argument analogous to that yielding (28),

$$V^S \geqq \pi d(a^*(\pi)) \tag{49}$$

which is necessary and sufficient for a^S to be equal to $a^*(\pi)$ in the social equilibrium. Since, by Proposition 5, V^S goes up with an increase in δ, it follows that the above inequality continues to hold at higher values of δ. This completes the proof. ‖

Acknowledgements. We wish to thank Tilman Börgers, Yeon-Koo Che, Douglas Gale, Karla Hoff, Deborah Minehart, Robert Rosenthal, Kunal Sengupta, Chun-Lei Yang and two anonymous referees for helpful comments on earlier drafts of the paper. Financial support under National Science Foundation Grant SBR-9414114 and Grant no. PB90-0172 from the Ministerio de Educación y Ciencia, Government of Spain, is gratefully acknowledged.

REFERENCES

ABREU, D. (1986), "Extremal equilibria of oligopolistic supergames", *Journal of Economic Theory*, **39**, 191–228.

ABREU, D. (1988), "Towards a theory of discounted repeated games", *Econometrica*, **56**, 383–396.

AKERLOF, G. (1983), "Loyalty Filters", *American Economic Review*, **73**, 54–63.

ALEEM, I (1993), "Imperfect information, screening, and the costs of informal lending: a study of a rural credit market in Pakistan", in K. Hoff, L. Braverman and J. Stiglitz (eds.), *The Economics of Rural Organization: Theory, Practice and Policy.*

AXELROD, R. (1984), *The Evolution of Cooperation* (New York: Basic Books).

BELL, C. (1993), "Interactions between institutional and informal credit agencies in rural India", in K. Hoff, L. Braverman and J. Stiglitz (eds.), *The Economics of Rural Organization: Theory, Practice and Policy.*

BERNHEIM, B. D. and RAY, D. (1989), "Collective dynamic consistency in repeated games", *Games and Economic Behavior*, **1**, 295–326.

CHO, I. K. and KREPS, D. (1987), "Signaling games and stable equilibria", *Quarterly Journal of Economics*, **102**, 179–222.

DATTA, S. (1993), "Building Trust" (Mimeo).

DAWKINS, R. (1993) *The Selfish Gene* (Oxford: Oxford University Press).

ELLISON, G. (1994), "Cooperation in the prisoner's dilemma with anonymous random matching", *Review of Economic Studies*, **61**, 567–588.

FARRELL, J. and MASKIN, E. (1989), "Renegotiation in repeated games", *Games and Economic Behavior*, **1**, 327–360.

FRANK, R. (1988) *Passions Within Reason* (New York: Norton).

FUDENBERG, D. and MASKIN, E. (1986), "The folk theorem in repeated games with discounting or with incomplete information", *Econometrica*, **54**, 533–556.

GHOSH, P. and RAY, D. (1995), "A theory of informal credit markets" (Mimeo).

KANDORI, M. (1992), "Social norms and community enforcement", *Review of Economic Studies*, **59**, 63–80.

KOHLBERG, E. and MERTENS, J. F. (1986), "On the strategic stability of equilibria", *Econometrica*, **54**, 1003–1038.

KRANTON, R. E. (1995), "The formation of cooperative relationships" (Mimeo).

MATSUSHIMA, H. (1990), "Long-term partnership in a repeated prisoner's dilemma with random matching", Economics Letters, **34**, 245–248.

OKUNO-FUJIWARA, M. and POSTLEWAITE, A. (1990), "Social norms and random matching games" (CARESS Working Paper No. 90-18, University of Pennsylvania).

ROSENTHAL, R. (1979), "Sequences of games with varying opponents", *Econometrica*, **47**, 1353–1366.

ROSENTHAL, R. and LANDAU, H. (1979), "A game theoretic analysis of bargaining with reputations", *Journal of Mathematical Psychology*, **20**, 235–255.

SAH, R. (1991), "Social osmosis and patterns of crime", *Journal of Political Economy*, **99**, 1272–1295.

SHAPIRO, C. and STIGLITZ, J. (1984), "Equilibrium unemployment as a worker disciplining device", *American Economic Review*, **74**, 433–444.

SOBEL, J. (1985), "A theory of credibility", *Review of Economic Studies*, **52**, 557–573.

TIROLE, J. (1996), "A theory of collective reputations", *Review of Economic Studies*, **63**, 1–22.

TULLOCK, G. (1985), "Adam Smith and the prisoner's dilemma", *Quarterly Journal of Economics*, **100**, 1073–1081.

VANBERG, V. J. and CONGLETON, R. D. (1992), "Rationality, morality and exit", *American Political Science Review*, **86**, 418–431.

WATSON, J. (1995), "Building a relationship" (Mimeo).