Patience or Fairness?
Analyzing Social Preferences in Repeated Games

John Duffy
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260
E-mail: jduffy@pitt.edu

Félix Muñoz-García
School of Economic Sciences
Washington State University
Pullman, WA 99164
E-mail: fmunoz@wsu.edu

May 2009

Abstract

This paper investigates how the introduction of social preferences affects players' equilibrium behavior in both one-shot and infinitely repeated versions of the Prisoner's Dilemma game. We first show that defection survives as the unique equilibrium of the stage game if at least one player is not too concerned about inequity aversion. Second, we demonstrate that in the infinitely repeated version of the game, fairness concerns operate as a "substitute" for time discounting, as fairness helps sustain cooperation for lower discount factors. We then extend our results to more general simultaneous-move games, and more general preferences. Finally, we point out the implications of our findings for the design and analysis of experiments involving repeated games. In particular, repeated game equilibria which are thought to be supported by sufficiently large discount factors, may in fact be sustained by a combination of discounting and social preference parameters, an observation that may help rationalize recent experimental findings.

KEYWORDS: Prisoner's dilemma; Repeated games; Inequity aversion; Time discounting.

JEL CLASSIFICATION: C72, C73, H43, D91.

1 Introduction

Inspired by a large volume of experimental evidence, there has been much recent work on *social*, as opposed to *individual*, preferences reflecting individuals' concern for fairness in the income distribution; see, e.g., Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Charness and Rabin (2002) among others. Much of this literature has examined how social preferences might facilitate cooperation among individuals who interact in sequential move, strategic environments, such as worker-employee, principal-agent or investor-trustee relationships. However, there has been comparatively little application of social preferences to *simultaneous*-move games, and in particular, to *infinitely repeated* versions of those games.[1]

This paper contributes to the literature by investigating how social preferences might facilitate cooperation among players interacting in the canonical simultaneous-move game –the Prisoner's Dilemma– which is appropriate for the study of strategic environments with extreme competitive incentives.[2] Surprisingly, we have found no literature exploring the effects of social preferences in infinitely repeated games.[3] We seek to understand whether social preferences (alternatively, "inequity aversion" or "altruism") facilitate cooperation in both the one-shot and infinitely repeated versions of the game, and, in the latter case, how players' time and social preferences might interact with one other.

While altruism in the social preference literature typically refers to inequity averse individuals (i.e., individuals who experience disutility from unequal payoff distributions), in the literature on infinite horizon, intertemporal decision-making, "altruism" has a different meaning. In that literature, altruism typically refers to the tendency of individuals to reduce their consumption today in order to increase the consumption of future generations or future selves, and is thus associated with greater "patience," or equivalently, lower rates of time preference (higher discount factors). We may therefore distinguish between two forms of altruism: *intertemporal* altruism (arising from players' time preference), and *intratemporal* altruism (arising from players' aversion to unequal payoffs, within a given time period). Indeed, Fehr and Leibbrandt (2008) have considered both of these types of altruism in a recent study and report that strong intra- and intertemporal altruism in laboratory experiments are highly correlated with greater cooperative behavior in the field. Our equilibrium predictions confirm their findings, and importantly, we provide the missing link between these two forms of altruism.

---

[1] Nevertheless, there is experimental evidence for social preferences in repeated simultaneous-move games. For instance, Fischbacher and Gächter (in press) use a strategy method to find direct evidence of social preferences in a linear voluntary contribution game experiment that involves simultaneous decisions by groups of four players interacting repeatedly for a finite number of periods (no discounting).

[2] Note that the Prisoner's Dilemma game is strategically equivalent to a voluntary contribution or "public good" game with a finite set of actions. Thus, all our analysis is also applicable to the public good game as well. In the last section of the paper, we extend our results to a larger class of games.

[3] Montero (2007) introduces inequity aversion in the Baron and Ferejohn (1989) legislative bargaining game, showing that individuals' inequity aversion might lead to *more* inequality. Intuitively, during the bargaining process the responder experiences a greater disutility from being left outside the winning coalition when he is envious than when he is not, which induces him to accept lower offers thereby increasing payoff inequality. In our model there is no such risk, which eliminates the possibility of this kind of result.

We first show that introducing social preferences can lead to cooperative outcomes as the equilibrium of a one-shot Prisoner's Dilemma game, but only when *both* players assign a sufficiently high value to inequity aversion. Otherwise, the strategy profile predicting defection by both players remains the unique equilibrium of the game under social preferences. In the infinitely repeated version of the Prisoner's Dilemma game, we demonstrate that subgame perfect equilibria can be supported for *lower* discount factors when players have social preferences (i.e., they care about the fairness of the payoff distribution) than when they do not. Furthermore, we investigate the relationship between an individual's time preference (discounting) and his fairness concerns (social preferences). Interestingly, we show that such fairness concerns work as a "substitute" for time preferences, since higher concerns about fairness reduce the minimum discount factor necessary to support cooperative outcomes in the repeated game. Our results help rationalize experimental observations where players cooperate under relatively low discount factors –values for which the "Folk theorem" for repeated games with discounting would *not* predict cooperation.

Our conclusions are also related to those of Kreps *et al.* (1982), who consider the role of informational asymmetries about players' types in the finitely repeated Prisoner's Dilemma game. Specifically, in their model a "rational" player may assign some probability to the possibility that his opponent "irrationally" plays a conditionally cooperative, tit-for-tat strategy. They show that there is a sequential equilibrium of the finitely repeated game in which the "rational" player imitates the "irrational" player by also playing tit-for-tat. Similarly, in this paper, we demonstrate that the existence of social preferences (particularly, concerns for fairness) may lead to cooperation among players in situations where cooperation would not exist among purely self-interested players. However, we develop our result from a simpler, behavioral primitive –social preferences, specifically inequity aversion (or fairness concerns)– which is supported by strong empirical evidence (see, e.g., Fehr and Fischbacher (2002) or Camerer (2003)). Furthermore we also develop our result in the infinitely repeated Prisoner's Dilemma game (Kreps et al. (1982) only study the finitely repeated version), and we relate fairness concerns to time preferences. The study of cooperation in infinitely repeated Prisoner's Dilemma games has recently become the subject of much study by experimentalists (see, among others, Dal Bo (2005), Normann and Wallace (2006), Aoyagi and Fréchette (in press), Duffy and Ochs (2009), Blonski et al. (2007), Dal Bo and Fréchette (2007)) and so an understanding of the mechanisms by which cooperation can be sustained in such environments is both important and timely.

Finally, we analyze the set of feasible, individually rational payoffs that can be achieved when playing the infinitely repeated game and how this set changes with changes in the parameterization of players' social preferences. In particular, we find that this set *shrinks* as individuals become more concerned about fairness. This result provides an effect opposite to that shown by Abreu *et al.* (1990), wherein the set of equilibrium payoffs in the infinitely repeated game *weakly increases* with increases in the discount factor.[4] Interestingly, this implies a potential confusion in the experimental

---

[4] In this sense, our paper is also related to Rabin (1997), who analyzes the introduction of concerns about fairness in *finitely* repeated games. Similarly to Rabin (1997), we find that players' preferences for fairness facilitate their coordination to play equilibrium outcomes with Pareto superior payoffs. However, Rabin's (1997) results can only be

literature on the source of observed cooperation on repeated games. Indeed, it suggests that such cooperation may not be due to players' high discount factors alone, but could instead arise from a *combination* of individuals' time and social preferences. We suggest a method to disentangle when cooperative behavior in the repeated Prisoner's Dilemma game can be explained using time preferences alone, and when reliance must be placed on both time and social preferences in order to rationalize cooperative play.

The paper is organized as follows. In the next section we introduce the model and derive the players' best response function. Section three then analyzes equilibrium strategy profiles, both for the stage game and for the infinitely repeated version. Section four elaborates on the set of feasible payoffs, how they shrink as players become more concerned about social preferences, and the potential confound we might observe under certain parameter values. Section five extends our results to a more general class of simultaneous-move games, and to a more general class of social preferences. Finally, section six concludes.

2 Model

Consider the stage game shown below. To make this a Prisoner's Dilemma game, both players' payoffs must satisfy the restriction $b > a > d > c$. In that case, both players' best response in the one-shot game is to choose D, "defect," either when the other player chooses C, "cooperate" (given that $b > a$), or when the other player defects as well (since $d > c$). Hence, the strategy profile (D,D) is the unique equilibrium of the one-shot stage game.

		Player 2	
		C	D
Player 1	C	a,a	c,b
	D	b,c	d,d

In this paper, however, we wish to analyze the game played by players who possess Fehr and Schmidt (1999)-type social preferences, a standard and tractable specification.[5] For the case of two players, Fehr and Schmidt's (1999) utility function reduces to:

$$U_i(x_i, x_j) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\},$$

where x_i is player i's payoff, and x_j is the other player j's payoff. The parameter α_i represents the disutility from allocations that are disadvantageously unequal for player i (i.e., due to envy about player j's higher payoff), while β_i captures the disutility from allocations that are advantageously unequal for player i (e.g., due to guilt over earning a higher payoff than player j). Additionally, Fehr

---

supported when the per-period payoffs are negligible, and he does not investigate the substitutability between social and temporal preferences.

[5] We consider more general forms of social preferences later in section 5. Until then, when we use the term "social preferences" we are referring to the Fehr and Schmidt (1999) specification.

and Schmidt (1999) assume that players' envy of the higher payoffs of others is always stronger than their guilt from earning higher payoffs than others. We capture this by assuming that $0 < \beta_i \leq \alpha_i \leq 1$.[6] We will contrast this case of "social preferences" (alternatively, "concerns for fairness" or "inequity aversion") with the more standard, self-regarding preferences where $\alpha_i = \beta_i = 0$ for all i.

Taking social preferences into account, the stage game can be reformulated as follows:

		Player 2	
		C	D
Player 1	C	a, a	$c - \alpha_1(b-c), b - \beta_2(b-c)$
	D	$b - \beta_1(b-c), c - \alpha_2(b-c)$	d,d

Notice in particular, that every player i's utility level decreases when he is either: the player with the highest payoff in the group (guilt), e.g., Player 1 under outcome (D,C), or when he is the player with the lowest payoff in the group (envy), e.g., Player 1 under outcome (C,D).

We next analyze a player's best response to the other player's actions under social preferences. Let $p_i(C)$ denote the probability that player i plays action C (cooperate). We have:

Lemma 1. *In the Prisoner's Dilemma game where players have social preferences, player i's best response to player j's actions is given by:*

$$p_i(C) = \begin{cases} 0 \textit{ for any } p_j(C) \in [0,1] & \textit{and } \beta_i < \frac{b-a}{b-c}, \\ 0 \textit{ if } p_j(C) < \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}, \\ (0,1) \textit{ if } p_j(C) = \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}, \\ 1 \textit{ if } p_j(C) > \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}, \end{cases}$$

where $\overline{p_j} = \frac{c-d+\alpha_j(c-b)}{b+c-a-d-(\alpha_j+\beta_j)(b-c)}$. *In addition,* $\overline{p_j}$ *is increasing in envy,* α_j*, and in guilt,* β_j.

The proof of Lemma 1 (as well as all other results that follow) may be found in the Appendix.

Lemma 1 reveals that player i's best response is to defect, for any action of the other player, $p_i(C) = 0$, if his guilt feeling, β_i, is sufficiently low. By contrast, if i's guilt feeling is sufficiently high, then he only defects if the probability with which player j cooperates is sufficiently low. Otherwise, if player j cooperates with high probability, player i cooperates as well, since the reduction in utility he would experience (given his high β_i) from receiving the highest payoff in the population (if he defected) would be too great.[7]

---

[6] Intuitively, $\alpha_i \geq \beta_i$ implies that players (weakly) suffer more from inequality directed at them than inequality directed at others. On the other hand, $\beta_i > 0$ means that players dislike being better off than others (this assumption rules out cases in which individuals are status seekers, but this serves to simplify the analysis). Finally, $\beta_i < 1$ suggests that when player i's payoff is higher than that of player j's by one unit (e.g. a dollar), player i is never willing to give up more than one unit in order to reduce this inequality. For a more detailed explanation of these assumptions, see Fehr and Schmidt (1999).

[7] Note that this best response function is similar to what Cooper et al. (1996) call "best response altruists," namely players for whom cooperate is their best response to cooperation, but defect is their best response to defection, as opposed to "dominant strategy altruists" for whom cooperation is always a best response, regardless of other players' strategies.

Importantly, note that only the guilt parameter (β) impacts on players' best responses, whereas the envy parameter (α) only raises the particular probability $\overline{p_j}$ that makes player i indifferent between cooperating and defecting. Indeed, guilt reduces players' incentives to defect when the other player is cooperating. In contrast, envy concerns only reinforce players' incentives to defect when the other player is defecting. These results suggest that the introduction of considerations about inequity aversion induce players to coordinate their actions. Certainly, when their concerns about inequity aversion are sufficiently high, players want to avoid miss-coordinated outcomes, such as (D,C) for Player 1 or (C,D) for Player 2. Indeed, these outcomes greatly reduce Player 1's (respectively Player 2) payoffs, because of the guilt feelings experienced from being the individual with the highest payoff in the population. We confirm this intuition in the following section, which examines the equilibrium strategy profiles for play of both the one-shot stage game and of the infinitely repeated game.

3 Equilibrium analysis

3.1 Stage game

In this section we analyze the equilibrium predictions for the one-shot Prisoner's Dilemma game under social preferences. In the next section we will examine players' equilibrium strategies in the infinitely repeated version of the same game.

Proposition 1. *In the one-shot Prisoner's Dilemma game where players have social preferences ($\beta_i > 0$), the following strategy profiles can be supported as Nash equilibria of the game:*

1. *(D,D), if $\beta_i < \frac{b-a}{b-c}$ for any player; and*
2. *(C,C) and (D,D) if $\beta_i > \frac{b-a}{b-c}$ for both players.*

Hence, for relatively low values of guilt aversion for *either* (or *both*) players, the unique Nash equilibrium of the game, (D,D), coincides with that in games where players have *no* concerns about the fairness of the payoff distribution (standard preferences)– see Figure 1. However, when *both* individuals are sufficiently concerned about fairness (shaded area), we can identify two different pure strategy Nash equilibria: one in which both players defect, and one in which both players cooperate. Importantly, note that the cutoff for guilt feelings supporting cooperative strategy profiles as the Nash equilibrium of this static game, $\beta_i > \frac{b-a}{b-c}$, depends on the particular payoff structure of the game. In particular, versions of the game where the incentive to deviate from the cooperative outcome, $b-a$, is high (strong competitive contexts), induce an increase in this cutoff, shrinking the set of parameter values (β_1, β_2) supporting any of the equilibria in the shaded area. Hence, in interactive environments with strong competitive pressures, the strategy profile (D,D) remains as the unique Nash equilibrium of the game for most parameter values.

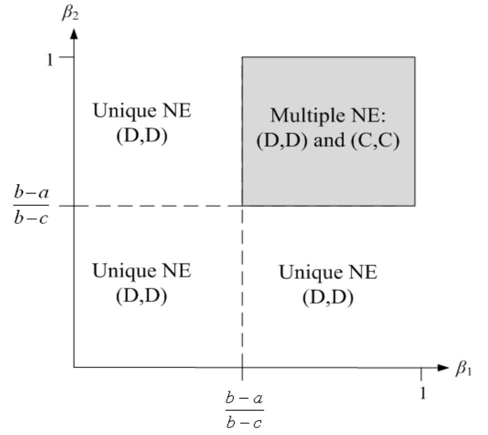

Figure 1. Equilibria of the stage game

This result is related to some other findings in the literature. For instance, Rabin (1993) shows that under social preferences, cooperative outcomes cannot be supported as equilibria when players' competitive incentives are too high. Rapoport and Chammah (1965) examine how variation in the parameterization of symmetric Prisoner's Dilemma games affects cooperation rates in a series of experiments. Analyzing their data we find that the correlation between the seven values of $\frac{b-a}{b-c}$ used in their experimental design and the corresponding frequency of cooperative play is negative (-0.55), i.e., consistent with Proposition 1, a higher value of $\frac{b-a}{b-c}$ is associated with a lower cooperation rate.[8] We can hence observe cooperation as an equilibrium strategy in the Prisoner's Dilemma if *both* players are sufficiently concerned about fairness. Such cooperation need not be a mistake in players' choices (tremble), nor a misunderstanding of the structure of the game (confusion), but could instead comprise an equilibrium strategy given players' social preferences.

3.2 Infinitely repeated game

We now turn to the main focus of the paper, namely equilibrium strategies in the infinitely repeated version of the Prisoner's Dilemma game under social preferences.

Proposition 2. *In the infinitely repeated Prisoner's Dilemma game where players have social preferences ($\beta_i > 0$), mutual cooperation can be sustained as the subgame perfect Nash equilibrium (SPNE) of the infinitely repeated game by use of the following grim-trigger strategy by player i in period t, $\sigma_i(t)$:*

---

[8] Rapoport and Chammah's (1965) experiment involved the play of seven, symmetric but differently parameterized, finitely repeated Prisoner's Dilemma games under random matching protocols. Our analysis is based on the game parameterizations and cooperation frequencies reported on pp. 37-39 of Rapoport and Chammah (1965). The correlation coefficient is significantly negative according to a 1-tailed t-test (p=.10).

$$\begin{aligned}\sigma_i(1) &= C \\ \sigma_i(\tau) &= \begin{cases} C \text{ if } \sigma_j(\tau-1) = C \text{ for all } j, \text{ and for all } t = 2,3,...,\tau-1, \\ D \text{ otherwise,} \end{cases}\end{aligned}$$

for any discount factor $1 > \delta_i > \delta_i^F$ *for every player* i*, where*

$$\delta_i^F(\beta_i) = \delta_i^{NF} - \frac{(b-c)(a-d)\beta_i}{(a+b-d)[a+b-d-\beta_i(b-c)]},$$

and $\delta_i^{NF} = \frac{b}{(a+b-d)}$ *denotes the minimal discount factor supporting cooperation in the infinitely repeated game in the case that individuals are not concerned about fairness. The discount factor when players have social preferences and are concerned about fairness,* $\delta_i^F(\beta_i)$*, is decreasing in* β_i*. Furthermore,* $\delta_i^F(\beta_i) \leq \delta_i^{NF}$*.*

The finding that the minimal discount factor supporting cooperation in the infinitely repeated Prisoner's Dilemma game where players care about fairness, $\delta_i^F(\beta_i)$, is *decreasing* in the parameter measuring guilt aversion, β_i, implies that mutual cooperation can be sustained under a broader set of parameter values when players possess social preferences than when they do not. We provide an example illustrating how the minimum discount factor necessary to sustain mutual cooperation varies with β_i in Figure 2 (under parameter values $a = 2$, $b = 3$, $c = 0$, $d = 1$).

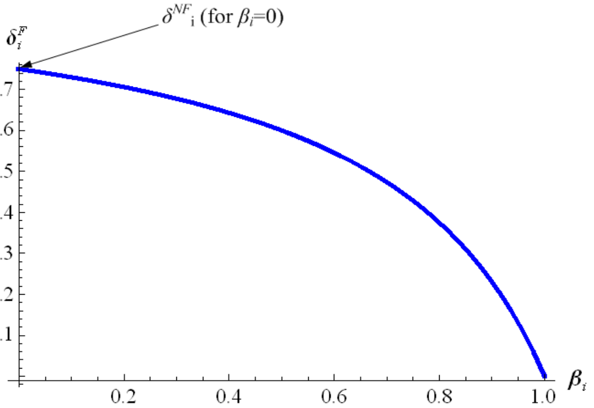

Figure 2. Discount factors $\delta_i^F(\beta_i)$ and δ_i^{NF}.

Therefore, we can interpret fairness considerations as a "substitute" for patience in sustaining mutual cooperation in the infinitely repeated Prisoner's Dilemma game. That is, cooperation can be supported if players either: (1) do not discount future payoffs too much –they have a sufficiently high $\delta_i^F(\beta_i)$– but they do not assign any value to fairness ($\beta_i \to 0$, in the vertical intercept in Figure 2); or (2) they discount future payoffs heavily –$\delta_i^F(\beta_i)$ is close to zero– but are strongly concerned about fairness (in the horizontal intercept in Figure 2); or (3) any combination of these two extremes. Hence, for a given discount of future payoffs, players' concerns about fairness make the Folk theorem for repeated games with discounting (see, e.g., Fudenberg and Maskin (1986)) applicable under a broader range of parameter values.[9]

The intuition for this result is related to the effect of fairness considerations on the players' utility function. In particular, a player's incentives to deviate from cooperation (his incentives to "cheat") are now reduced by the guilt he experiences from obtaining a higher payoff than that of other players. This result can help to explain experimental findings such as those reported by Murnighan and Roth (1983, Table 4), Dal Bo (2005, Table 5) and Dal Bo and Fréchette (2007, Table 5) where experimental subjects playing an indefinitely repeated Prisoner's Dilemma game are observed to cooperate even when continuation probabilities (induced discount factors) do not support such cooperation as an equilibrium of the repeated game under standard, self-interested preferences. Our result showing that fairness concerns may "substitute" for patience can be used to rationalize these experimental observations.

Note that for extremely high fairness concerns, ($\beta_i = 1$), the above SPNE is supported for most discount factors.[10] This result confirms our earlier intuition: when a player is sufficiently concerned about guilt, his utility from cheating is so reduced (from obtaining higher payoffs than other players) that he would not find it profitable to deviate, for most discount factors. Similarly, if we recall that fairness works as a "substitute" for patience in terms of supporting cooperation in the repeated game, we can also interpret this result as suggesting that when players are highly concerned about fairness they replace patience for fairness concerns as their mechanism for sustaining cooperation.

Corollary 1. *The minimal discount factor, $\delta_i^F(\beta_i)$, supporting cooperation in the infinitely repeated Prisoner's Dilemma game where players have social preferences, is concave in the weight given to fairness concerns (guilt), β_i.*

In addition to the possibility that players can substitute for patience with fairness concerns, Corollary 1 specifies that the substitution between the minimal discount factor and fairness concerns occurs at an *increasing* rate: the higher the value players assign to fairness, β_i, the more weight on future payoffs, $\delta_i^F(\beta_i)$, they can give up and still sustain cooperation.

4 Feasible payoffs

Let us now examine how the previous result translates into the set of *feasible payoffs* for the infinitely repeated game. Figure 3(a) represents the set of feasible payoffs for the case where individuals do not assign any value to fairness considerations, $\beta_i = 0$, i.e., payoff pairs within the diamond-shaped figure. Let us denote this set of feasible payoffs as FP_{NF}, where the subscript NF denotes that

---

[9] Mutual cooperation can also be supported as the SPNE of the infinitely repeated game by the use of other type of strategies, such as those in which defection is punished only during a limited number of time periods, or other reciprocal strategies like "tit-for-tat." In this section, we focus for simplicity in one type of strategy in order to analyze how social preferences can work as substitute for temporal preferences.

[10] When $\beta_i \to 1$, the minimal discount factor becomes $\delta_i^F = \frac{c}{(a-d)+c} \geq 0$. In the numerical example illustrated in Figure 2, where $c = 0$, δ_i^F approaches 0 as $\beta_i \to 1$.

players are not concerned about fairness. Formally, the set of feasible payoffs is defined as the convex hull of all payoffs $v \in \mathbb{R}_+^2$ feasible under the set of available actions $a \in A$, i.e., $FP =$convex hull $\{v \mid \text{for all } a \in A \text{ such that } U(a) = v\}$.

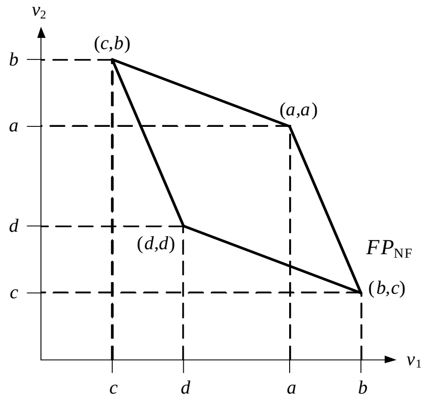

Figure 3(a). Set of feasible payoffs.

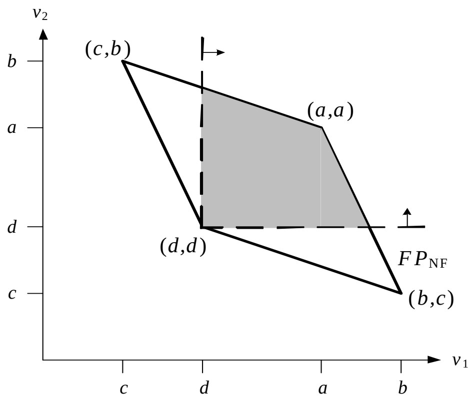

Figure 3(b). Individually rational payoffs.

Figure 3(b), additionally, shades the set of feasible payoffs that are *individually rational*, i.e., all those payoffs such that $v_i > \widehat{v_i}$ for both players, where $\widehat{v_i}$ is the reservation utility (or minmax value) $\widehat{v_i} = \min_{a_j}\left[\max_{a_i} U_i\left(a_i, a_j\right)\right]$. Hence, the shaded area in Figure 3(b) depicts the set of feasible, individually rational payoffs, where the minmax payoff is (d, d).

Let us now analyze how the feasible set is affected as players' concerns about fairness increase. In particular, Figure 4 illustrates sets of feasible payoffs for players with positive concerns about fairness, $\beta_i > 0$, and compares those with the set of feasible payoffs for a player who assigns no value to fairness, FP_{NF}.

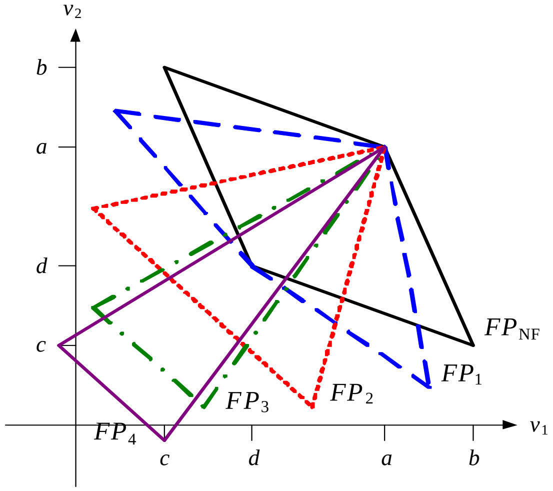

Figure 4. Set of feasible payoffs for $\beta_i > 0$.

Specifically, the blue set (long-dash) of feasible payoffs, FP_1, illustrates players with low concerns about fairness, i.e., $\beta_1, \beta_2 \in \left(0, \frac{b-a}{b-c}\right]$. Further increases in fairness concerns are represented by the red set (short-dash) of feasible payoffs, FP_2, where $\beta_1, \beta_2 \in \left(\frac{b-a}{b-c}, \frac{b-d}{b-c}\right]$, the green set (dash-dot) of feasible payoffs, FP_3, where fairness concerns become $\beta_1, \beta_2 \in \left(\frac{b-d}{b-c}, 1\right)$, and finally the violet set (solid) of feasible payoffs, FP_4, which corresponds to the case where $\beta_1 = \beta_2 = 1$. Graphically, note that the set of feasible payoffs shrinks as a fan closing its arms along the main diagonal, with its end at the pair of payoffs resulting from mutual cooperation, (a, a).

Next, for the FP_2 set of feasible payoffs illustrated in Figure 4, Figure 5 below shades the portion of that set representing feasible *and* individually rational -FIR- payoffs. (Other FIR sets given FP sets are constructed similarly). Notice first that the FIR payoffs when players are concerned about fairness are *not* simply the payoffs to the northeast of the payoff vector (d, d), as in the case where individuals are not concerned about fairness. Instead, when players are concerned about fairness, they now experience a disutility from all payoffs that lie *far away from* the main diagonal (unequal payoff vectors), which results in the set of FIR payoffs becoming more compressed around more egalitarian payoffs. We identify the set of FIR payoffs in the case where players are concerned with fairness, in particular the constraining payoff vectors $\overline{v_i}$ that are illustrated in Figure 5, in the following proposition.

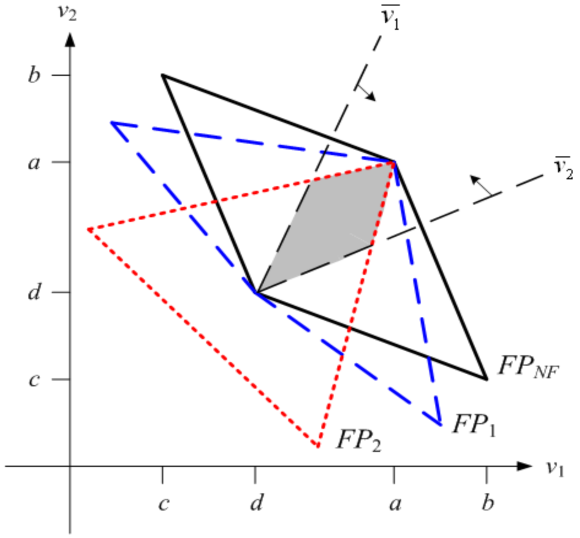

Figure 5. Effects of higher α_i on the FIR set.

Proposition 3. *In the infinitely repeated Prisoner's Dilemma game where individuals have social preferences and are concerned with fairness, every player i's individually rational payoffs (within the set of feasible payoffs) must satisfy* $v_i > \overline{v_i}$*, where*

$$\overline{v_i} = d + \frac{\alpha_i}{1+\alpha_i} v_j \text{ for players } i \text{ and } j$$

Additionally, $\overline{v_i}$ *is increasing in the envy parameter,* α_i*.*

Hence, as individuals become more envious (i.e., as α_i increases), the lower bound on the set of FIR payoffs, $\overline{v_i}$, shifts (downwards for player 1, $\overline{v_1}$, and upwards for player 2, $\overline{v_2}$) shrinking this set from above and below, respectively– resulting in the shaded area, as illustrated in Figure 5. Furthermore, increases in players' guilt aversion β_i must satisfy the preference assumption that $1 \geq \alpha_i \geq \beta_i > 0$. Due to this constraint, higher β_i will necessarily imply higher α_i. Thus, higher β_i will serve to shrink the size of the FIR set, as illustrated in Figure 5. In particular, in the limit, as $\beta_i, \alpha_i \to 1$, Proposition 3 reveals that the FIR set collapses to the main diagonal line segment that starts at (d, d) and ends at (a, a) as illustrated by the thick dashed line in Figure 6.

Note the different roles played by the envy and guilt dimensions of fairness concerns in the repeated game. On the one hand, Proposition 2 shows that the minimal discount factor necessary to support cooperation in the infinitely repeated game decreases with players' guilt concerns (β_i). On the other hand, increases in individual's envy concerns (α_i) affect how egalitarian the payoff distribution in the repeated game must provided that players choose strategy profiles that improves on their per-period payoff relative to that of the Nash equilibrium of the stage game, i.e., as Proposition 3 describes, an increase in α_i shrinks the set of FIR payoffs. Therefore, guilt works as a "tool" to support cooperation under a larger set of parameter values, whereas envy allows players to reach more equitable payoffs, provided that cooperative behavior can be sustained.

Hence, we find that the set of FIR payoffs in the infinitely repeated game weakly shrinks as players become more concerned about fairness. This finding may be contrasted with that of Abreu *et al.* (1990), who show, in the context of infinitely repeated games in which players are not concerned about fairness, that the set of FIR payoffs weakly *expands* with increases in the players' discount factor (i.e., as players assign a higher value to future payoffs, the set of FIR payoffs that can be supported as equilibria of the repeated game expands). Thus our result complements that of Abreu *et al* (1990) by suggesting the existence of an *opposing force* affecting the size of the set of FIR payoffs: higher discount factors weakly expand this set, while higher concerns about fairness serve to shrink the same set. Let us finally add a corollary to the previous results.

Corollary 2. *In the infinitely repeated Prisoner's Dilemma game with social preferences and concerns for fairness, unequal average equilibrium payoffs ($v_i \neq v_j$) can be supported as SPNE only if either players' envy is not extreme ($\alpha_i \neq 1$), or their guilt feelings are not extreme ($\beta_i \neq 1$), or both.*

We represent this last result in Figure 6, where players' guilt feelings are assumed to be extreme, $\beta_i = \beta_j = 1$. In this case, the set of FIR payoffs shrinks as much as possible, becoming in the limit the diagonal (thick long-dashed) line where $v_i = v_j$. Hence, as the previous corollary states, unequal payoffs (payoffs away from the main diagonal) can only be supported in the perfect equilibrium of the infinitely repeated game if players are not extremely concerned about inequity aversion.

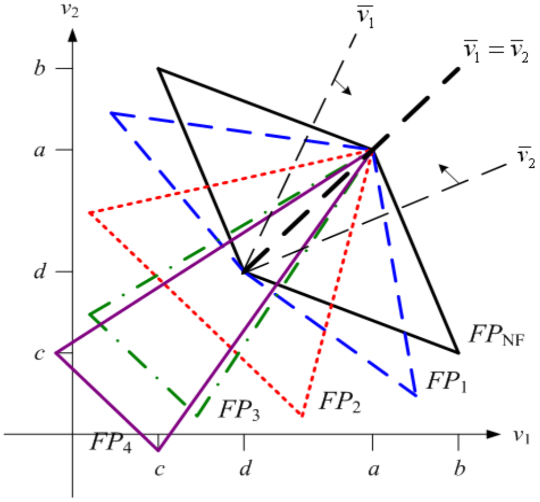

Figure 6. Extreme concerns about fairness.

4.1 Patience or Fairness? Experimental evidence

The above results suggest that a certain pair of payoffs can be sustained with a continuum of discount factors and concerns about fairness (different combinations of δ and β). Importantly, this implies that observed cooperation between players in infinitely repeated games could be due to a mix of these two factors. Hence, our results suggest the possibility of some confusion as to which concern it is that leads players to sustain cooperation over time in repeated game experiments: is it patience alone (high δ values), is it fairness alone (high β values), or is it a combination of the two?

Our previous results provide a partial answer to this question. In the area in which no overlap occurs (the unshaded area of FP_{NF} in Figure 5 which does not overlap with FP_2, for instance), players' cooperation is sustained because of individuals' time preferences alone. However, in the overlapping regions (the shaded area of FP_{NF} coinciding with that of FP_2 in Figures 5), players' cooperation in repeated games could be supported by combinations of discount factors and/or concerns about fairness.

Let us relate this theoretical prediction to some experimental data from an indefinitely repeated prisoner's dilemma game experiment reported on in Duffy and Ochs (2009). In Figures 7(a)-7(b) we show FIR payoffs for the parameterization of the indefinitely repeated Prisoner's Dilemma game that Duffy and Ochs implemented in the laboratory.[11] We then add a representative sample of realized average payoff data (from Duffy and Ochs's fixed pairings, indefinitely repeated game treatment) so as to compare these realized payoffs with our equilibrium predictions. In these figures, we use black dots to represent the average payoffs accruing to fixed pairs of subjects over all rounds played in an indefinitely repeated Prisoner's Dilemma game.

---

[11] Duffy and Ochs (2009) parameterize the stage game using $a = 20$, $b = 30$, $c = 0$ and $d = 10$. These numbers correspond to payoffs in US$ cents per round played. They use a continuation probability, $\delta = .90$. For details see Duffy and Ochs (2009).

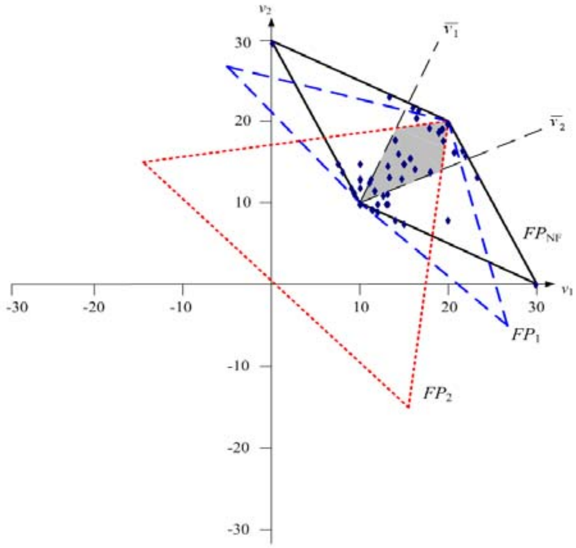

Figure 7(a). $\beta_i = \frac{1}{3}$

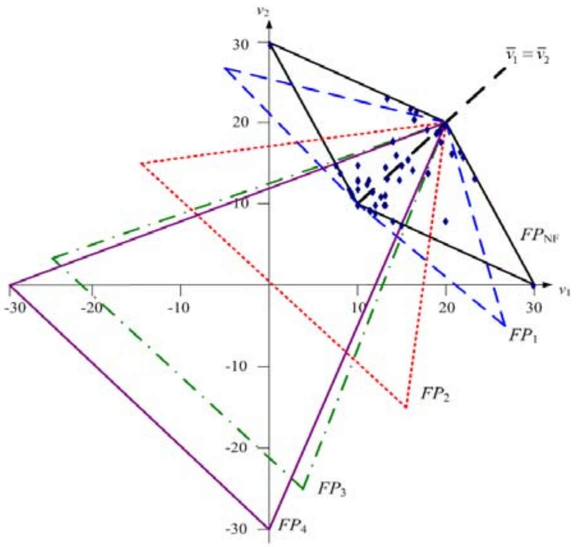

Figure 7(b). $\beta_i = 1$

In particular, Figure 7(a) compares our predictions with respect to observed behavior for the case where players are moderately concerned about inequity aversion, i.e., $\beta_1, \beta_2 \in \left(\frac{b-a}{b-c}, \frac{b-d}{b-c}\right]$, labeled FP_2,while Figure 7(b) focuses on the case of highly concerned players, $\beta_1 = \beta_2 = 1$. As these figures suggest, individual subject behavior in the experiment can be explained: (1) by relying on individuals' time preferences *alone* (see the average payoff values lying outside the set of shaded FIR payoffs in Figure 7a, or apart from the diagonal line in Figure 7b); or (2) relying both on individuals' time *and* social preferences (payoffs lying within the shaded set of FIR payoffs in Figure 7a and on the diagonal line in Figure 7b). In particular, we see that when players' concerns about fairness are moderate, $\beta_i = 0.33$, as in Figure 7a, the shaded set of FIR payoffs is relatively large and we observe that most of the experimental observations on payoffs (64%) lie within this FIR set. Such payoff observations can be supported using either social or time preferences or some combination of both. However, as concerns about fairness become more extreme, $\beta_i = 1$, as in Figure 7b, the set of FIR payoffs collapses to the main diagonal -dashed line in Figure 7b. We find that only 26% of the experimental observations lie on this diagonal and can therefore be rationalized by either social or time preferences, or both. Most payoffs do not rely on the diagonal and could only be rationalized by appeal an appeal to concerns about future payoffs (time preferences). Nevertheless, the percentage of average payoff observations found lying on the diagonal is surprisingly high suggesting that some subjects playing the infinitely repeated prisoner's dilemma might be cooperating as the result of extremely high concerns about guilt.

5 Extensions to more general games and preferences

Let us now analyze equilibrium strategies in more general infinitely repeated games. In particular, we consider simultaneous-move games with complete information, and a finite number of players and actions. For simplicity, we restrict our attention to games where players can improve via cooperative action choices, their per-period payoffs, relative to the Nash equilibrium stage game payoff, which we denote by, $\widetilde{v_i}$. That is, we consider games where there exists an action profile $a = (a_i, a_{-i})$ with payoff $U_i(a_i, a_{-i}) = v_i$, where $v_i > \widetilde{v_i}$, for every player i.[12] When players do not assign a value to fairness, i.e., $\beta_i = \alpha_i = 0$, mutual cooperation can be sustained as a SPNE of the infinitely repeated game for any discount factor δ_i such that $\delta_i > \delta_i^{NF}$ for all i; see, e.g., Friedman (1971). Similarly, as we have shown, when players with social preferences assign a value to fairness, mutual cooperation can be supported for any discount factor δ_i such that $\delta_i > \delta_i^F(\beta_i)$. This section examines under what conditions $\delta_i^F(\beta_i) \leq \delta_i^{NF}$, i.e., that cooperation can be supported under a larger set of conditions when players are concerned about fairness than when they are not.

We begin by defining a "weak symmetry" condition that we will make use of in Proposition 4 below. Specifically, a game satisfies "weak symmetry" if and only if all players' payoffs coincide when they play the Nash equilibrium of the stage game[13], i.e., $\widetilde{v_i} = \widetilde{v_j}$, as well as when they play the subgame perfect equilibrium of the repeated game, $v_i = v_j$. For instance, in the Prisoner's Dilemma game, this weak symmetry assumption implies that players 1 and 2 earn the same payoff when they both choose to defect (in the Nash equilibrium of the stage game), and when they both choose to cooperate (in the SPNE of the infinitely repeated game). Hence, if both individual's payoff coincides under these two strategy profiles, utility levels when players are concerned about fairness will not be diminished (since there is no inequality in the payoff distribution). Note that weak symmetry is not as restrictive as stronger forms of symmetry, whereby players' payoffs coincide under *all* strategy profiles.

Proposition 4. *If the stage game's payoff structure satisfies the "weak symmetry" condition, then $\delta_i^F(\beta_i) \leq \delta_i^{NF}$ holds for all parameter values, $\alpha_i, \beta_i > 0$. Otherwise, $\delta_i^F(\beta_i) \leq \delta_i^{NF}$ if and only if*

$$\frac{\max_a U_i(a) - v_i}{\max_a U_i(a) - \widetilde{v_i}} \geq \frac{\max_a U_i^F(a) - v_i^F}{\max_a U_i^F(a) - \widetilde{v_i}^F}.$$

This result confirms our previous intuition from the prisoner's dilemma game: cooperation can be supported for broader conditions when players are concerned about fairness than when they are not. In particular, when the game is weakly symmetric in players' payoffs, the main factor

---

[12] For simplicity, we assume that the payoff $U_i(a_i, a_{-i}) = v_i$ can be achieved using pure strategies. Otherwise, one can suppose that any randomization producing payoff v_i is publicly observed by all players, thus allowing deviations to be detected by every player.

[13] Note that we are implicitly assuming the existence of a unique Nash equilibrium in the stage game. Our results can be extended to stage games with multiple Nash equilibria, and use $\widetilde{v_i}$ to denote the payoff that individuals obtain in the Nash equilibrium providing the highest payoff.

driving players towards cooperation is the fact that deviations are less profitable when individuals are concerned about fairness than when they are not. Indeed, when players are concerned about fairness, their utility from deviating is reduced by the guilt they experience from being the player with the higher payoff. Importantly, this result can be applied to many simultaneous-move games besides the Prisoner's Dilemma game, including voluntary contribution (public good) games, and coordination games. In contrast, when the game does not satisfy the weak symmetry condition, cooperative outcomes can be supported with a lower discount factor only if the above condition is satisfied. Intuitively, this condition holds if a player's incentives to deviate from the cooperative outcome are relatively stronger when he is not concerned about fairness than when he is concerned about fairness.

Notice that, under "weak symmetry," the payoff inequality in the Nash equilibrium outcome of the stage game coincides with that in the subgame perfect equilibrium of the infinitely repeated game (for sufficiently high discount factors). If we do not assume "weak symmetry" in payoffs, however, income inequality might vary when the stage game is repeated infinitely many times: income inequality increases (decreases, or remains unchanged, respectively) if the distribution of payoffs across players is more unequal (more equal, or remains unchanged) in the Nash equilibrium of the stage game than in the perfect equilibrium of the repeated game. By contrast, if the game satisfies "strong symmetry" (i.e., players' payoffs coincide for every strategy profile), then an inequity averse player does not experience any guilt when he deviates from cooperation, e.g., his payoff in the (D,C) outcome of the Prisoner's Dilemma coincides with that of other players. Hence, under strong symmetry, the minimal discount factor supporting cooperation in the infinitely repeated game when players are concerned about fairness, $\delta_i^F(\beta_i)$, is equal to that when players are not concerned about fairness, δ_i^{NF}.

In the following corollary we show that the result of Proposition 3 can be extended to inequity averse players with preferences different from those in Fehr and Schmidt (1999).

Corollary 3 *The result from Proposition 3, $\delta_i^F(\beta_i) \leq \delta_i^{NF}$, holds both under linear and non-linear social preferences.*

Proposition 4 holds for the linear, Fehr and Schmidt (1999) specification of social preferences, where every unit of payoff inequality induces the same disutility (either in the form of envy or guilt). Corollary 3 shows that such result extends to more general (possibly *non-linear*) social preferences, for instance, social preferences of the variety suggested by Neilson (2006):

$$U_i(x) = x_i - \alpha_i \sum_{i \neq j} u(x_i - x_j),$$

where u is any continuous function of the level of payoff inequality, $x_i - x_j$, and α_i is player i's sensitivity to such payoff inequality. Specifically, we can assume that u is increasing in $x_i - x_j$ whenever $x_i > x_j$, i.e., individuals experience a disutility (guilt) from receiving a higher payoff relative to other players in the population. Further, the disutility from guilt can be either increasing in payoff inequality (if u is a convex function), or decreasing in payoff inequality (if u is a concave function). Our results are thus applicable to players with relatively general social preferences.[14]

6 Conclusions

In this paper we have investigated how the introduction of social preferences and fairness concerns may affect players' equilibrium behavior in both one-shot and infinitely repeated versions of the Prisoner's Dilemma game. In particular, we analyze how fairness concerns modify players' incentives to cooperate in both versions of the game.

We first show that introducing players who are concerned about fairness might lead to cooperative outcomes as the equilibrium of the one-shot stage game, but only if *both* players assign a sufficiently high value to inequity aversion. Otherwise, the strategy profile predicting defection by both players remains the unique equilibrium of the one-shot game.

We next show that the set of feasible, individually rational payoffs supported as subgame perfect equilibrium of the infinitely repeated game can be sustained for lower discount factors when players are concerned about fairness than when they are not. We further show that the latter result generalizes to a larger class of infinitely repeated games. Intuitively, fairness concerns (alternatively "inequity aversion" or "intratemporal altruism") operate as a *substitute* for time discounting in terms of helping individuals sustain cooperation in repeated games: individuals with strong concerns about inequity aversion can be more impatient (assigning lower value to future payoffs) and still sustain cooperation in repeated games, as compared with individuals who are more patient and unconcerned about fairness.

Finally, we show that the set of feasible, individually rational payoffs when players are concerned about fairness (are inequity averse) is a (potentially large) subset of the set of feasible individually rational payoffs when players are unconcerned about fairness (have standard preferences). We further observe that as fairness concerns increase, the set of feasible, individually rational equilibrium payoffs shrinks relative to set of feasible individually rational payoffs under standard preferences.

Our findings suggest that there is a potential confound in the interpretation of experimental results showing high levels of cooperative behavior in infinitely (indefinitely) repeated games. First, our findings can be used to rationalize observed cooperative behavior in experimental settings with low induced discount factors where the Folk theorem for repeated games with discounting (under standard preferences) would predict an absence of cooperative behavior. Second, even in settings where this Folk theorem applies, we have shown how observed cooperation frequencies may be explained by time preferences alone or by a combination of time and social preferences. As a first step toward disentangling these two effects, we provide payoff vectors for which cooperation in the repeated game may only be rationalized using time discounting. Nonetheless, more experimental

---

[14] Note that Bolton and Ockenfels (2000) consider that an individual's utility level increases in his own share in total payoffs. Hence, a player's benefit from defecting in the repeated game is in fact augmented, since his share in total payoffs increases from defecting, thus increasing the minimal discount factor inducing him to cooperate, relative to the case of selfish preferences.

research on this topic is clearly needed, in order to clarify this potential confound.

7 Appendix

7.1 *Proof of Lemma 1*

Let us first analyze player 1's best response function. If player 2 chooses C, then player 1 selects D if and only if $a < b - \beta_1(b-c)$, i.e., if $\beta_1 < \frac{b-a}{b-c}$. If player 2 chooses D, then player 1 selects D if and only if $c - \alpha_1(b-c) < d$, i.e., if $\frac{c-d}{b-c} < \alpha_1$. Since $c - d < 0$, then $\frac{c-d}{b-c} < \alpha_1$ is satisfied for any $\alpha_1 > 0$. Hence, player 1 responds to D by also choosing D, for any parameter values. By symmetry, player 2's best response function is analogous. Finally, the probability, $p_j(C)$, with which player j chooses C such that player $i \neq j$ is made indifferent between selecting C and D must satisfy:

$$p_j(C)a + (1 - p_j(C))[c - \alpha_1(b-c)] = p_j(C)[b - \beta_1(b-c)] + (1 - p_j(C))d.$$

Solving for $p_j(C)$, we have $\overline{p_j} = \frac{c-d+\alpha_j(c-b)}{b+c-a-d-(\alpha_j+\beta_j)(b-c)}$. Summarizing,

$$p_i(C) = \begin{cases} 0 \textit{ for any } p_j(C) \in [0,1] & \textit{and } \beta_i < \frac{b-a}{b-c}, \\ 0 \textit{ if } p_j(C) < \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}, \\ (0,1) \textit{ if } p_j(C) = \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}, \\ 1 \textit{ if } p_j(C) > \overline{p_j} & \textit{and } \beta_i \geq \frac{b-a}{b-c}. \end{cases}$$

In addition, $\frac{\partial \overline{p_j}}{\partial \alpha_j} = \frac{(b-c)\left[a-b+\beta_j(b-c)\right]}{\left[a-b-c+d+(\alpha_j+\beta_j)(b-c)\right]^2}$, and $\frac{\partial \overline{p_j}}{\partial \alpha_j} = -\frac{(c-b)[c-d+\alpha_j(c-b)]}{\left[a-b-c+d+(\alpha_j+\beta_j)(b-c)\right]^2}$, are both positive if and only if $\beta_i \geq \frac{b-a}{b-c}$, which is satisfied by definition for any $\overline{p_j}$.

7.2 *Proof of Proposition 1*

Given players' best response function from Lemma 1, if *either* $\beta_1 < \frac{b-a}{b-c}$ or $\beta_2 < \frac{b-a}{b-c}$, then the unique Nash equilibrium of the game is (D,D). Otherwise (if *both* $\beta_1 > \frac{b-a}{b-c}$ and $\beta_2 > \frac{b-a}{b-c}$), then for both players we have that choosing C is a best response to C, and choosing D is a best response to D. Hence, when $\beta_1 > \frac{b-a}{b-c}$ and $\beta_2 > \frac{b-a}{b-c}$, (C,C) and (D,D) are Nash equilibria of the game in pure strategies. We now must check for the existence of mixed strategy Nash equilibria. We know that if $p_j(C) \in (0,1)$ and $\beta_i > \frac{b-a}{b-c}$, then player i randomizes with probability $p_i(C) = \frac{c-d+\alpha_j(c-b)}{b+c-a-d-(\alpha_j+\beta_j)(b-c)}$. However, $p_i(C) \geq 1$ for all $\beta_i \geq \frac{b-a}{b-c}$, which eliminates the possibility of equilibria in which either player uses non-degenerate mixed strategies. ■

7.3 *Proof of Proposition 2*

Consider a representative period, τ, and suppose that both players have cooperated in all prior periods $t = 1, 2, ..., \tau - 1$. If player i deviates to $\sigma_i(\tau) = D$ (the best response to player j choosing C when $\beta_1 < \frac{b-a}{b-c}$), player j's trigger strategy specifies $\sigma_j(t) = D$ for all future periods after the deviation, i.e., for all $t \geq \tau + 1$. Thus in this case, the deviation by player i in period τ, yields him the discount payoff of $[b - \beta_1(b-c)] + \frac{\delta}{1-\delta}d$.

By contrast, if player i does not deviate in period τ, so that both individuals continue cooperating, he obtains a discounted payoff of $\frac{1}{1-\delta}a$. Comparing these two payoffs, we find that the deviation by player i is unprofitable if and only if $\delta_i > \frac{b-b\beta_i+c\beta_i}{a+b-d-b\beta_i+c\beta_i}$ for every player i. Note that $\frac{b-b\beta_i+c\beta_i}{a+b-d-b\beta_i+c\beta_i} = \frac{b}{a+b-d} - \frac{(b-c)(a-d)\beta_i}{(a+b-d)[a+b-d-\beta_i(b-c)]}$, where $\delta_i^{NF} \equiv \frac{b}{a+b-d}$.

Finally, we still need to show that a player would choose D forever, once either individual deviated in an earlier period, i.e., we must check that the trigger strategy specified in Proposition 2 is also a best response off-the-equilibrium path. In order to prove this, note that if player j deviates, then he would be required to play D in all future periods. Further, player i's best response to individual j's playing D is to play D itself (we showed that in proposition 1). Therefore, the trigger strategies defined above comprise a subgame perfect equilibrium of this infinitely repeated prisoner's dilemma game. Note that $\frac{b}{a+b-d} - \frac{(b-c)(a-d)\beta_i}{(a+b-d)[a+b-d-\beta_i(b-c)]} \leq \frac{b}{a+b-d}$, since $\frac{(b-c)(a-d)\beta_i}{(a+b-d)[a+b-d-\beta_i(b-c)]} \geq 0$ for all parameter values given that $b > a > d > c$. Hence, $\delta_i^F(\beta_i) \leq \delta_i^{NF}$.

Differentiating $\delta_i^F(\beta_i)$ with respect to β_i, we obtain $\frac{\partial \delta_i^F(\beta_i)}{\partial \beta_i} = -\frac{(b-c)(a-d)}{(a+b-d-b\beta_i+c\beta_i)^2}$,which is negative for any parameter values since $b > c$ and $a > d$. Hence, $\delta_i^F(\beta_i)$ is strictly decreasing in β_i. ■

7.4 *Proof of Corollary 1*

Double differentiating the minimal discount factor that supports cooperation when player are concerned with fairness, (defined in proposition 2), $\delta_i^F(\beta_i)$, with respect to β_i, we obtain $\frac{\partial^2 \delta_i^F(\beta_i)}{\partial \beta_i^2} = -\frac{2(b-c)^2(a-d)}{(a+b-d-b\beta_i+c\beta_i)^3}$, which is negative for any parameter values since $b > c$ and $a > d$. Hence, $\delta_i^F(\beta_i)$ is concave in β_i. ■

7.5 *Proof of Proposition 3*

Payoffs pairs (v_i, v_j) above the reservation utility for player i imply that $v_i - \beta_i(v_i - v_j) > d$ when $v_i \geq v_j$; and $v_i - \alpha_i(v_j - v_i) \geq d$ when $v_i < v_j$. More compactly, we have:

$$\begin{aligned} v_i &\geq \frac{d}{1+\beta_i} + \frac{\beta_i}{1+\beta_i} v_j \text{ for all } i \text{ and } j, \text{ if } v_i > v_j; \text{ and} \\ v_i &\geq \frac{d}{1+\alpha_i} + \frac{\alpha_i}{1+\alpha_i} v_j \text{ for all } i \text{ and } j, \text{ if } v_i < v_j. \end{aligned}$$

These two lower bounds cross at (d, d). Similarly for player j. Considering player i and j's inequalities simultaneously, only the second inequality is binding for each player. Therefore, the set of individually rational payoffs can be defined by $\overline{v_i} = v_i \geq d + \frac{\alpha_i}{1+\alpha_i} v_j$ for all i and j. Differentiating with respect to α_i, we obtain $\frac{\partial \overline{v_i}}{\partial \alpha_j} = \frac{v_j}{(1+\alpha_i)^2}$, which is positive for any parameter values. ■

7.6 *Proof of Proposition 4*

Assume that there is an action profile $a = (a_i, a_{-i})$ with payoff $U(a) = v$, where $v \in V$ and $v_i > \widetilde{v_i}$ for every player i, and consider the following strategy profile: in period zero each player i plays a_i. Each player i continues to play a_i so long as a was played in all previous periods. If at least one player did not play according to a, then every player i reverts to the minmax action for the rest of the game, with associated payoff $\widetilde{v_i}$. This strategy profile is a Nash equilibrium for discount factors, δ_i, such that

$$\frac{1}{1-\delta_i} v_i \geq \max_a U_i(a) + \frac{\delta_i}{1-\delta_i} \widetilde{v_i} \iff \delta_i \geq \frac{\max_a U_i(a) - v_i}{\max_a U_i(a) - \widetilde{v_i}} = \delta_i^{NF}$$

This strategy profile is subgame perfect, given that, in every subgame off the equilibrium path, the strategies are to play $\widetilde{v_i}$ forever, the Nash equilibrium of the stage game. Finally, note that when player i is concerned about fairness, his maximal benefit to deviating from cooperation, $\max_a U_i^F(a)$, is weakly lower than when he is not concerned about fairness, $\max_a U_i(a)$, because of the guilt he feels from being the player with the highest payoff, i.e., $\max_a U_i^F(a) \leq \max_a U_i(a)$, Hence, $\delta_i^{NF} \geq \delta_i^F(\beta_i)$ is satisfied if and only if

$$\frac{\max_a U_i(a) - v_i}{\max_a U_i(a) - \widetilde{v_i}} \geq \frac{\max_a U_i^F(a) - v_i^F}{\max_a U_i^F(a) - \widetilde{v_i}^F},$$

where we do not impose any assumption on the symmetry of payoffs, i.e., $v_i \neq v_i^F$ and $\widehat{v_i} \neq \widehat{v_i}^F$ for any $\alpha_i, \beta_i > 0$. Otherwise, when the payoff structure satisfies weak symmetry, so that both $v_i = v_i^F$ and $\widetilde{v_i} = \widetilde{v_i}^F$ hold, this implies that

$$\frac{\max_a U_i(a) - v_i}{\max_a U_i(a) - \widetilde{v_i}} \geq \frac{\max_a U_i^F(a) - v_i}{\max_a U_i^F(a) - \widetilde{v_i}^F}$$

is satisfied for any parameter values, since $\max_a U_i(a) \geq \max_a U_i^F(a)$ and $v_i > \widetilde{v_i}$, and thus, $\delta_i^{NF} \geq \delta_i^F(\beta_i)$. ■

7.7 *Proof of Corollary 3*

Note that social preferences are introduced in Proposition 3 by assuming that a player's maximal benefit to deviating from cooperation when he is concerned about fairness, $\max_a U_i^F(a)$, is weakly lower than that when he is not concerned, $\max_a U_i(a)$, i.e., $\max_a U_i^F(a) \leq \max_a U_i(a)$. No conditions are assumed about the players' payoffs v_i and v_i^F, or about $\widetilde{v_i}$ and $\widetilde{v_i}^F$. These assumptions embody both linear and non-linear social preferences.■

References

[1] Abreu, Dilip, David G. Pearce and Ennio Stacdhetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring," Econometrica 58, 1041-1063.

[2] Aoyagi, Masaki and Guillaume Fréchette (in press), "Collusion in repeated games with imperfect public monitoring," forthcoming in Journal of Economic Theory.

[3] Baron, D. P. and J. A. Ferejohn (1989), "Bargaining in Legislatures," American Political Science Review 94, 73-88.

[4] Benoit, Jean-Pierre and Vijay Krishna (1985), "Finitely repeated games," Econometrica 53, 890-904.

[5] Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A theory of equity, reciprocity, and competition," American Economic Review 90, 166-93.

[6] Blonski, Matthias, Peter Ockenfels and Giancarlo Spagnolo (2007), "Co-operation in infinitely repeated games: Extending theory and experimental evidence," working paper.

[7] Camerer, Colin F. (2003), *Behavioral game theory*, Princeton: Princeton University Press.

[8] Charness, Gary and Matthew Rabin (2002), "Understanding social preferences with simple tests," Quarterly Journal of Economics 117, 817-869.

[9] Cooper, Russell, Douglas V. DeJong, Robert Forsythe and Thomas W. Ross (1996), "Cooperation without reputation: Experimental evidence from prisoner's dilemma games," Games and Economic Behavior 12, 187-218.

[10] Dal Bo, Pedro (2005), "Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games," American Economic Review 95, 1591-1604.

[11] Dal Bo, Pedro and Guillaume R. Fréchette (2007), "The evolution of cooperation in infinitely repeated games: Experimental evidence" working paper, Brown University and NYU.

[12] Duffy, John and Jack Ochs (2009), "Cooperative behavior and the frequency of social interaction," Games and Economic Behavior 66, 785–812.

[13] Fehr, Ernst and Urs Fischbacher (2002), "Why social preferences matter: The impact of non-selfish motives on competition, cooperation and incentives," Economic Journal 112, C1-C33.

[14] Fehr, Ernst and Andreas Leibbrandt (2008), "Cooperativeness and impatience in the tragedy of the commons," IZA Discussion paper no. 3625.

[15] Fehr, Ernst and Klaus Schmidt (1999), "A theory of fairness, competition and cooperation," Quarterly Journal of Economics 114, 817-68.

[16] Fischbacher, Urs and Simon Gächter (in press), "Heterogeneous social preferences and the dynamics of free riding in public good experiments," forthcoming in American Economic Review.

[17] Friedman, James (1971), "A noncooperative equilibrium for supergames," Review of Economic Studies 38, 1-12.

[18] Fudenberg, Drew and Eric Maskin (1986), "The folk theorem in repeated games with discounting or with incomplete information," Econometrica 54, 533-556.

[19] Kreps, David, Paul Milgrom, John Roberts and Robert Wilson (1982), "Rational cooperation in the finitely repeated prisoner's dilemma," Journal of Economic Theory 27, 245-52.

[20] Montero, Maria (2007), "Inequity aversion may increase inequity," Economic Journal 117, 192-204.

[21] Murnighan, J. Keith and Alvin E. Roth (1983), "Expecting continued play in prisoner's dilemma games: a test of several models," Journal of Conflict Resolution 27, 279-300.

[22] Neilson, William S. (2006), "Axiomatic reference-dependence in behavior towards others and towards risk," Economic Theory 28, 681-92.

[23] Normann, Hans-Theo and Brian Wallace (2006), "The impact of the termination rule on cooperation in a Prisoner's Dilemma experiment," working paper.

[24] Rabin, Matthew (1993), "Incorporating fairness into game theory and economics," American Economic Review 83, 1281-1302.

[25] Rabin, Matthew (1997), "Fairness in repeated games," University of California at Berkeley working paper no. 97-252.

[26] Rapoport, Anatol and Albert M. Chammah (1965), *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Ann Arbor: University of Michigan Press).

[27] Selten, Reinhard and Rolf Stoecker (1986), "End behavior in sequences of finite prisoner's dilemma supergames. A learning theory approach," Journal of Economic Behavior and Organization 7, 47-70.