Available online at www.sciencedirect.com
ScienceDirect
Games and Economic Behavior 60 (2007) 287–306
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Endogenous transfers in the Prisoner's Dilemma game: An experimental test of cooperation and coordination

Gary Charness[a,*], Guillaume R. Fréchette[b], Cheng-Zhong Qin[a]

[a] *Department of Economics, University of California at Santa Barbara, Santa Barbara, CA 93106-9210, USA*
[b] *New York University, Department of Economics, 269 Mercer Street, 7th floor, New York, NY 10003, USA*

Received 21 February 2005

Available online 29 January 2007

Abstract

We test a two-stage compensation mechanism for promoting cooperation in Prisoner's Dilemma games. Players first simultaneously choose binding non-negative amounts to pay their counterparts for cooperating, and then play the induced game knowing these amounts. In our games, all payment pairs *consistent* with mutual cooperation in subgame-perfect equilibrium transform these games into *coordination games*, with *both* mutual cooperation and mutual defection as Nash equilibria in the second stage. When endogenous transfer payments are not permitted, cooperation is much less likely. Mutual cooperation is most likely when the (sufficient) payments are identical, and it is also substantially more likely with payment pairs that bring the mutual-cooperation payoffs closer together. Both the Fehr–Schmidt and Charness–Rabin models predict that transfers that make final payoffs closer are preferred; however, they do not explain why equal transfers are particularly effective. Transfers are also effective in sustaining cooperation even when they are imposed and not chosen.

JEL classification: A13; B49; C72; C78; C91; K12

Keywords: Prisoner's dilemma; Endogenous transfer payments; Compensation mechanism; Coase theorem; Coordination games; Equilibrium selection

* Corresponding author. Fax: +1 805 893 8830.
E-mail addresses: charness@econ.ucsb.edu (G. Charness), frechette@nyu.edu (G.R. Fréchette), qin@econ.ucsb.edu (C.-Z. Qin).
URLs: http://www.econ.ucsb.edu/~charness (G. Charness), homepages.nyu.edu/~gf35 (G.R. Fréchette), http://www.econ.ucsb.edu/faculty/~qin/ (C.-Z. Qin).

doi:10.1016/j.geb.2006.10.010

1. Introduction

The prisoner's dilemma is by far the most famous example of a game with a unique Pareto-inefficient Nash equilibrium. The chief characteristic of this game is that while there are substantial gains that could be attained through cooperation, non-cooperation (*defection*) is dominant for each player. The theoretical result is that all players defect, even though joint defection leaves each player with less than he or she could have obtained through mutual cooperation. A multitude of experiments have been conducted on the Prisoners' Dilemma (see Rapoport and Chammah, 1965; Dawes, 1980, and Roth, 1988 for surveys of these experiments). The central finding in these studies is that mutual cooperation is indeed rather rare in the Prisoner's Dilemma. Since players always do better with respect to their individual payoffs by defecting, few people elect to cooperate in this environment, leading to poor social outcomes. It is thus desirable to design mechanisms that will implement the efficient outcome.

Coase (1960) presents an example involving a rancher and a farmer, in which the rancher's cattle stray onto the farmer's property and damage the crops beyond the benefit to the rancher. Coase argues that even if the rancher's cattle are legally allowed to trespass, the efficient outcome, as in the case where the rancher's cattle are legally prohibited from trespassing, will still result because the farmer would then have an incentive to pay the rancher to cooperate (reducing the number of straying cattle). That is, with well-defined property rights, no transaction costs, and fully symmetric information, efficiency is neutral to the assignment of responsibilities for damages; this result has come to be called the *Coase theorem.*

Varian (1994) presents a general two-stage compensation mechanism that can be seen as being complementary to the Coasian approach.[1,2] It implements efficient outcomes through subgame-perfect equilibria in a wide range of environments with externalities, including Prisoner's Dilemma games with certain specifications of the payoffs.[3] The mechanism provides a formalization of bargaining involved in the Coase theorem, and it does not involve a regulator or central planner mandating taxes or transfer payments; instead it relies upon the parties to design transfer payments that leads to the efficient outcome. In essence, the Prisoner's Dilemma can be seen as an environment with a two-sided externality.

Applying this mechanism to a Prisoner's Dilemma game, each party would make a binding pre-play offer to pay the other for cooperating in stage 1; upon observing these offers, each party then chooses to cooperate or to defect in the Prisoner's Dilemma game in stage 2. A natural solution concept is subgame-perfect equilibrium (henceforth SPE); while one wishes to offer enough to induce the other to cooperate, it is best to offer the minimum amount that is required to achieve this goal.[4] Qin (2002) characterizes the conditions on payment pairs that are necessary and sufficient to "induce the players to cooperate" (to be defined shortly).[5]

---

1 In fact, Varian presents two mechanisms, one of which is general and one of which only works for certain Prisoner's Dilemma games.

2 Varian (1995) explicitly spells out the connection with Coase.

3 Moore and Repullo (1988) show that, given certain assumptions, almost all choice rules can be implemented by subgame-perfect equilibria. The compensation mechanism seems to be considerably simpler than the examples provided in Moore and Repullo (1988).

4 See also Ziss (1997) where it is shown that the efficient outcome is not among the set of possible SPE outcomes for certain Prisoner's Dilemma games.

5 Jackson and Wilkie (2003) consider more general strategy-dependent transfer payments that players may offer to each other before playing a game in strategic form. For example, with a Prisoner's Dilemma game, players can offer among permissible transfer payments those that will be carried out only when both players defect or only when Player 1

We test the compensation mechanism experimentally, using three parameterizations of the Prisoner's Dilemma, and find substantial success: We observe cooperation rates of 43–68% when transfer payments are permitted, compared to 11–18% when transfer payments are not feasible. We also find distinct patterns in cooperation rates; factors include whether mutual cooperation is a (strict) Nash equilibrium in the subgame induced by the chosen transfer payments and whether the net effect of the transfer payments is to make the payoffs from mutual cooperation closer together or further apart. Cooperation rates are highest when qualifying transfer payment pairs are identical, while transfer payments that expand the gap in the mutual-cooperation payoffs are less effective than transfer payments that narrow this gap.

Our study has bearing on issues such as contractual performance and breach, where each party posts a reward for the other party's performance (or deposits a bond) in an escrow held by a neutral third party.[6] In the field, this is observed in real estate and construction matters, where performance bonds and escrows are the rule. Side payments can be seen in international fishing and international pollution 'contracts.' Alternatively, this could also be relevant for the provision of public goods, if the parties make pledges conditional on completion of the project. Agreements to make contributions contingent upon other contributions are seen in many forms of fund-raising, including public television and radio.

In addition, our results have clear legal implications for both positive and negative interpretations of the Prisoner's Dilemma. By positive interpretations, we mean situations where the Prisoner's Dilemma is a reduced form for cooperation being socially-beneficial, such as in public-goods games. The success of the mechanism suggests that it may not be necessary for the legislative authority to attempt to directly implement mutual cooperation. By negative interpretations, we mean situations where the Prisoner's Dilemma is a reduced form for a case where cooperation hurts society, such as with collusion in Cournot quantity competition. In this case, our results suggest that the legislative authority should be quite careful to ensure that side contracts are illegal.

2. Background and theory

There have been several laboratory tests of the Coase theorem, beginning with Hoffman and Spitzer (1982) and Harrison and McKee (1985). In these studies, there is an optimal choice of lotteries, in terms of total (expected) social payoffs. However, one agent, who has an individual incentive not to choose the optimal lottery, controls the lottery chosen after the parties can contract over side payments. These studies generally find that the parties are able to contract effectively.[7] Nevertheless, as the contracting problem in our case is much more difficult, given its two-sided nature, in some sense the Prisoner's Dilemma is a more challenging test.

Andreoni and Varian (1999) were the first to experimentally test the performance of the compensation mechanism in the Prisoner's Dilemma. The cooperation rate nearly doubles with feasible transfers in their game, from 26 to 50%, showing considerable effectiveness for the mechanism. To some degree, our study follows in their footsteps; nevertheless our design per-

defects while Player 2 cooperates, and so on. The main result of their paper is a complete characterization of supportable equilibrium payoffs rather than transfer payments.

[6] Williamson (1983) discusses the merits of crafting ex ante incentive structures for the Prisoner's Dilemma.

[7] However, Kahneman et al. (1990) find substantially less trading of consumption goods than the level predicted by the Coase theorem, attributing the gap to the *endowment effect*.

mits us to go beyond their study in at least two important respects. First, we consider games where there is a substantial range of payment pairs that induce the players to cooperate in SPE. While mutual cooperation is predicted with all qualifying payment pairs, it may be that we can identify factors that behaviorally enhance or inhibit mutual cooperation. In contrast, with integer payments there is a unique SPE in the game considered in Andreoni and Varian (1999), making it difficult to discover any such patterns.

Second, the SPE payments in Andreoni and Varian lead to a solution in dominant strategies, facilitating cooperation given that sufficient transfer values have been chosen. In comparison, payment pairs required for inducing cooperation transform the experimental Prisoner's Dilemma games in the present paper into *coordination games* between the players, in which there are two distinct Nash equilibria, (C, C) and (D, D). In our games, mutual defection after choosing these payment pairs can typically be ruled out as part of an equilibrium strategy, as is explained later in the paper. Nevertheless, this analysis requires fairly sophisticated reasoning unlikely to manifest in an experimental game. This would seem to be a substantially more difficult task than selecting mutual cooperation when it is the unique Nash equilibrium in the second stage of the game, as in Andreoni and Varian (1999). Thus, ceteris paribus, one might expect cooperation rates to be lower in our case.

We are aware of only a few other experimental papers involving the compensation mechanism.[8] Hamaguchi et al. (2003) test the penalty version of the Varian mechanism in an emissions-control game with a large strategy space (15 choices of prices for each player, then one player has 15 choices of quantities), and find only 20% Nash equilibrium play in this difficult environment. Chen and Gazzale (2004) test the role of supermodularity in achieving convergence through learning, using a generalized version of the compensation mechanism; they find that supermodular games converge significantly better than those well below the threshold for supermodularity. Finally, Bracht et al. (2004) find that the mechanism proposed in Falkinger (1996) and tested in Falkinger et al. (2000) outperforms the Varian mechanism in a public-goods setting.

The compensation mechanism converts the Prisoner's Dilemma into a two-stage game. In the first stage, each player chooses (simultaneously) how much to pay his or her counterpart for cooperating. After learning the payments offered in the first stage, the players then choose between action C and action D in the Prisoner's Dilemma game. Using the standard notation for the original PD payoffs ($S_k < P_k < R_k < T_k$, $k = 1, 2$),[9] a payment pair (H_1, H_2) changes players' payoffs associated with the various action pairs to the ones shown in Fig. 1.

		Player 2: C	Player 2: D
Player 1	C	$R_1 - H_1 + H_2, R_2 + H_1 - H_2$	$S_1 + H_2, T_2 - H_2$
	D	$T_1 - H_1, S_2 + H_1$	P_1, P_2

Fig. 1. The transformed generic game.

[8] We thank an anonymous referee for comments in this regard.

[9] Mnemonically, one might think of these letters as representing T(emptation), R(eward), P(unishment), and S(ucker) payoffs, respectively.

We now provide a formal definition regarding when a payment pair induces the players to cooperate:

Definition. A payment pair H^* induces the players to cooperate if there is an SPE that involves players offering payments in H^* in stage 1 and cooperating in stage 2 conditional on payment pair H^*.

In Qin (2002) it is shown that a payment pair H^* induces the players to cooperate if and only if, for $i \neq j$,

$$T_i - R_i \leqslant H_j^* \leqslant R_j - S_j, \tag{1}$$

$$H_j^* - H_i^* \leqslant R_j - P_j, \tag{2}$$

$$H_j^* \leqslant T_i - R_i \quad \text{whenever} \quad P_i - S_i \leqslant T_i - R_i, \tag{3}$$

and

$$H_i^* \leqslant P_j - S_j \text{ and } H_j^* \leqslant P_i - S_i \quad \text{whenever} \quad P_i - S_i > T_i - R_i. \tag{4}$$

The set containing all such payment pairs H^* often describes a rectangle.[10]

The specific games we implement in the lab are the ones shown in Fig. 2.

For example, consider our Game 1, where $(S_1, P_1, R_1, T_1) = (8, 28, 40, 52)$ and $(S_2, P_2, R_2, T_2) = (8, 24, 52, 60)$. Applying conditions (1)–(4), we see that $8 \leqslant H_1^* \leqslant 16$ and $12 \leqslant H_2^* \leqslant 20$. Consider the transfer pair $(H_1^*, H_2^*) = (12, 16)$.[11] Figure 3 illustrates how this payment pair transforms Game 1.

Game 1

		Player 2	
		C	D
Player 1	C	40, 52	8, 60
	D	52, 8	28, 24

Game 2

		Player 2	
		C	D
Player 1	C	32, 52	4, 60
	D	40, 8	20, 24

Game 3

		Player 2	
		C	D
Player 1	C	44, 36	8, 44
	D	52, 0	32, 28

Fig. 2. Our experimental games.

[10] However, when the cost of cooperation $(P - S)$ is greater than the gain from defecting $(T - R)$ for each player, the set of such pairs is determined by conditions (1), (2), and (4) only. In this case, the set may be a hexagon, as with our Game 3.

[11] One might ask why the SPE of $(H_1^*, H_2^*) = (12, 16)$ is not undermined for Player 1 by $(11, 16)$. The answer is that Player 1 could (correctly) believe that Player 2 would choose to defect if $H_1 < 12$. This can be explained as follows: The payment pair H^* in SPE can reflect players' *demands* for payments to cooperate (as embodied in their contingent actions in the second stage), so that if a player's demand is not fulfilled, he can credibly refuse to cooperate. So, for

		Player 2	
		C	D
Player 1	C	44, 48	24, 44
	D	40, 20	28, 24

Fig. 3. Game 1 after transformation by $(H_1, H_2) = (12, 16)$.

		Player 2	
		C	D
Player 1	C	40, 52	20, 48
	D	40, 20	28, 24

Fig. 4. Game 1 transformed by $(H_1, H_2) = (12, 12)$.

Both (C, C) and (D, D) are Nash equilibria in the subgame in the transformed game. In fact, simple calculations shows that (C, C) and (D, D) are Nash equilibria in any subgame resulting from a payment pair satisfying (1)–(4) for Game 1. Nevertheless, in SPE, mutual defection cannot be an action pair in stage 2 conditional on the payment pair satisfying (1)–(4).

To see this, suppose on the contrary that a payment pair satisfying (1)–(4), or equivalently satisfying $8 \leqslant H_1^* \leqslant 16$ and $12 \leqslant H_2^* \leqslant 20$, induces (D, D). Then, Player 1 receives 28 and Player 2 receives 24 from the corresponding SPE. However, from Game 1 together with Fig. 4, it follows that C is strictly dominant for Player 1 in the subgame led by payment pair (H_1^*, H_2) with $H_2 > 20$. Given that Player 1 plays his or her SPE strategy, by offering to pay $H_2 > 20$ and by cooperating in the second stage conditional on payment pair (H_1^*, H_2), Player 2's payoff would become $52 - H_2 + H_1^* \geqslant 60 - H_2$. But this means that by choosing $20 < H_2 < 28$ and inducing Player 1 to choose C, Player 2 could have received a payoff bigger than 24, since $60 - H_2 > 24$. Consequently, given Player 1's strategy in that SPE with payment $8 \leqslant H_1^* \leqslant 16$, Player 2 would wish to change his transfer from $12 \leqslant H_2^* \leqslant 20$ to $20 < H_2 < 28$. This contradicts the supposition that a pair (H_1^*, H_2^*) satisfying (1)–(4) for Game 1 induces (D, D) in equilibrium.

3. Experimental design and hypotheses

We wished to not only test the general effectiveness of endogenous payments for cooperation in achieving cooperation and efficiency, but to also investigate the determinants of cooperation given that mutual cooperation is a Nash equilibrium in the induced subgame or is part of an SPE with the contingent payments chosen. In other words, are there particular patterns in (qualifying) payment pairs that are particularly effective in affecting cooperation? This may well affect equilibrium-selection in a coordination game.

We therefore chose three games where the payment pairs inducing cooperation transform the game into coordination games. Further, in order to test for the effect of possible determinants on cooperation, we chose games in which the region of payment pairs inducing the players to cooperate was substantial and included points completely in its interior. While in theory any payment is allowed, only integer values are permitted in the experiment; we chose larger nominal payoffs, in order to have many payment pairs that are completely inside the SPE-region.

Appendix B, on the *Games and Economic Behavior* supplementary materials website, illustrates, for the three games used in the lab, the bounds of the SPE-transfer pairs consistent with

instance, Player 1 would prefer to offer 11 instead of 12 to player 2 for cooperating. However, that is not acceptable to Player 2 because he can say no to Player 1 by planning to defect (in a credible way, since mutual defection is also a Nash equilibrium in the subgame). The SPE here implies a certain degree of bargaining between the two players.

mutual cooperation being a subgame-perfect action pair. We noted earlier that mutual cooperation is the unique action pair consistent with SPE in Game 1. Given the conditions mentioned above, we can see that this is also the case for Game 2. However, matters are more complex for Game 3, as the conditions ruling out mutual defection are only satisfied for some SPE payment pairs. It turns out that (C, C) is the only SPE action pair when either $H_1 > 16$ or $H_2 > 16$, but the action pair (D, D) is also consistent with subgame-perfection for other values in the SPE-region.[12,13]

For purposes of statistical analysis, one cannot assume that each observation is independent, since each person plays in 25 periods and interacts with other players during the session. While we account for this in regression analysis, we also perform non-parametric statistical tests across conditions. To facilitate these tests, we partitioned the 16 participants in each session into four separate groups, with the four people in each group interacting only with each other over the course of the session. In this way, we obtain four completely independent observations in each session.[14]

3.1. The experiment

We conducted a series of experiments in 15 separate sessions at the University of California at Santa Barbara. We had three sessions for each of three different prisoner's-dilemma games with endogenous transfer payments. As controls, we conducted one session for each game without the possibility of transfer payments, and one session for each game with transfer payments imposed by the experimenter (these were drawn from the realized transfers in the other treatments). There were 16 participants in each session, with average earnings of about \$15 (including a \$5 show-up payment) for a one-hour session. Participants were recruited by e-mail from the general student population.[15]

We provided instructions on paper, which were discussed at the beginning of the session; a sample of these instructions is presented in Appendix A, on the *Games and Economic Behavior* supplementary materials website. Our computerized experiment was programmed using the *z*-tree software (Fischbacher, 1999). After a practice period, participants played 25 periods; each person was a Row player in some periods and a Column player in others, with one's role being drawn at random from period to period, and the person with whom one was matched also being determined at random from the other members of the subgroup.

[12] The reasoning goes as follows: Assume (D, D) is played after a payment pair H^* in the SPE-region, in which case Player 1 gets 32. Notice that Player 1 can make C strictly dominant for Player 2 in the action subgame by choosing any $H_1 > 28$. If she does so, then Player 2 plays C and Player 1 can thus guarantee herself payoff $44 - H_1 + H_2^*$. This payoff is greater than 32 if and only if $H_1 < 12 + H_2^*$. Thus Player 1 would have an incentive to change her payment in the pair H^* if there exists a payment that simultaneously satisfy $28 < H_1 < 12 + H_2^*$. Such a payment exists if and only if $H_2^* > 16$. Thus $H_2^* > 16$ is incompatible with the assumption that (D, D) is induced by payment pair H^*. This shows that (D, D) cannot be induced by a payment pair H^* in the SPE-region if $H_2^* > 16$. Similar reasoning shows that (D, D) cannot be induced by a payment pair in the SPE-region if $H_1^* > 16$. In summary, (D, D) cannot be induced by a payment pair H^* in the SPE-region if either $H_1^* > 16$ or $H_2^* > 16$.

[13] In this case the mechanism has two SPEs; we are not aware of any study that investigates what might be expected as an outcome, depending on structural characteristics of the situation, when there are multiple equilibria for a mechanism.

[14] Subjects were told that they were randomly re-matched, but not that this was done in subgroups.

[15] Since part of what we wish to study are decisions conditional on transfers making cooperation an SPE, to improve our chances of observing SPE transfers potential subjects were told (in a mass e-mail) that we were particularly interested in students who either had high grade-point averages or who were majoring in mathematics or the sciences. We then screened the applicants using these criteria; participants typically were either graduate students or had GPAs above 3.70. As discussed in Section 5, our results in the standard Prisoner's Dilemma are similar to those in other experiments.

Players first learned their roles for the period and then (if cooperation-rewards were feasible) chose amounts to transfer to their counterparts in the event of their cooperation. After learning the amounts chosen, both players in a pair then simultaneously chose whether to cooperate or defect in the subgame, and were then informed of the outcome.

3.2. Hypotheses

In this section, we formulate several hypotheses based on the predictions of the theory. We also explore some of the tensions that may stop these predictions from being realized. First, given that cooperation is an SPE of the game with transfers, but not of the standard Prisoner's Dilemma, we have:

Hypothesis 1. There will be more cooperation in the sessions where players can choose transfer payments than in the sessions where no transfers are made.

Since cooperation is only a Nash equilibrium in the induced subgame for some transfer pairs, we have:

Hypothesis 2. There will be more cooperation when mutual cooperation is a Nash equilibrium in the subsequent subgame led by the chosen transfer payments.

We next consider whether, given transfer pairs consistent with mutual cooperation being an equilibrium, there are certain characteristics of transfer pairs that are particularly effective in leading to mutual cooperation. In principle, the theoretical arguments hold regardless of the location of a point within the mutual-cooperation Nash or SPE-regions. Thus, the hypothesis that emerges from the theory on this point is:

Hypothesis 3. Given that a transfer induces cooperation, the cooperation rate will not differ according to any characteristics of the transfer pair.

While the standard arguments predict no differences in behavior for qualifying transfer pairs, the fact that there are multiple equilibria in the subgame suggests that secondary factors will influence the choice of play in the subgame, falsifying Hypothesis 3. For example, reward-pairs that are on the Nash 'border' seem less likely to lead to cooperation. Consider Game 1, with $8 \leqslant H_1^*$ and $12 \leqslant H_2^*$. Suppose the transfer pair is $(H_1^*, H_2^*) = (12,\ 12)$, on the border of the Nash or SPE-regions. The induced subgame, where both (C, C) and (D, D) are Nash equilibria, is shown in Fig. 4 (see).

If Player 1 thinks Player 2 is going to cooperate, he stands to get 40 with either C or D; however, C for Player 1 is weakly-dominated by D in the subgame. Furthermore, Player 2 stands to gain a lot (32) by Player 1 choosing C over D. In this case, Player 1 may feel unhappy that Player 2 has chosen to give no incremental reward for cooperative play, while hoping or expecting to reap large rewards from mutual cooperation. In this sense, a border reward is like a zero offer in the ultimatum game—a rejection does not really cost the rejector anything, but punishes the selfish party. Thus, border reward pairs may be less effective in achieving cooperation.

All else equal, we might also expect players to be more likely to cooperate when transfer payments (and thus the rewards for cooperation) are higher, even when all transfer pairs considered are within the Nash or SPE-regions. Here risk-dominance considerations might serve to help select the equilibrium in the induced coordination game. We also consider how the characteristics of transfer pairs that lead to cooperation dovetail with the Fehr and Schmidt (1999) ("F-S") and Charness and Rabin (2002) ("C-R") models of social preferences.

Our final hypothesis concerns whether behavior is sensitive to identifiable characteristics of the three different games we test. In principle, since it is possible in each of these games to choose reward-pairs leading to cooperation in equilibrium, we should see mutual cooperation in every case that such reward-pairs are chosen. Even if this is not the case, we should still see no difference in the effectiveness of transfer payments in achieving cooperation across these games:

Hypothesis 4. Given that a transfer pair is consistent with mutual cooperation either in the subgame or as part of an SPE, the cooperation rate will not differ across the three experimental games. Furthermore, the effectiveness of transfer payments in enhancing cooperation will not vary across games.[16]

On the other hand, we might intuitively expect to see a relationship between risk and reward. Rapoport and Chammah (1965) presents some ideas on how the relationships between the entries in the payoff matrix of the Prisoner's Dilemma might be expected to influence cooperation rates (without transfers); however, they only consider a 'symmetric' Prisoner's Dilemma, where players have identical payoffs in each cell of the 2×2 game. Nevertheless, the concept of risk and reward may well influence decisions. One simple idea is to compare the size of the joint (or individual) payoffs with mutual cooperation to those with mutual defection; bigger gains from mutual cooperation should translate into more cooperation.

4. Experimental results and hypothesis tests

In this section, we first present a summary of our experimental data. We then provide a regression analysis of the data. Throughout, we relate the data analysis to the hypotheses elaborated in Section 3.

4.1. Data summary

Figure 5 shows the average cooperation rates by game and treatment.

Cooperation rates are clearly higher in treatments where transfers are allowed; the comparisons are 15.8 vs. 53.9% for Game 1, 17.5 vs. 68.1% for Game 2, and 10.8 vs. 42.9% for Game 3.[17] All of these differences are highly statistically significant (p-value < 0.01, one-sided

[16] We might expect some differences between Game 3 and Games 1 and 2, since mutual defection in the induced subgame is still consistent with some of the SPE transfers in Game 3. We could restrict Hypothesis 4 to those transfer pairs for which only mutual cooperation is consistent with SPE; however, given doing so does not change the results stated in the analysis below.

[17] Our cooperation rates when transfers were not permitted ranged from 11 to 18%; this is in line with rates of cooperation observed in previous studies—Roth and Murnigham (1978), Cooper et al. (1996) and Andreoni and Miller (1993) observed 10, 25, and 18% cooperation rates, respectively.

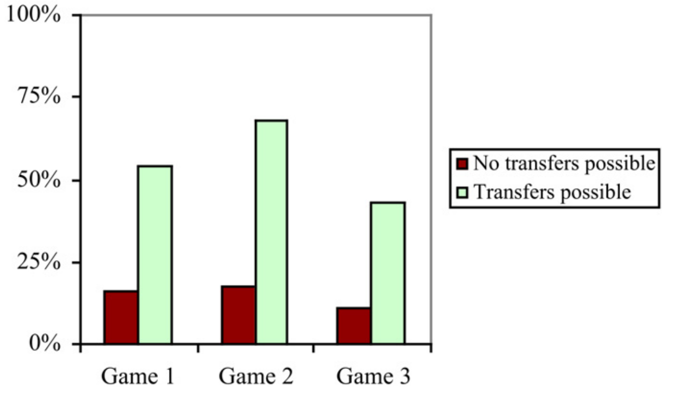

Fig. 5. Cooperation rates, by game.

Mann–Whitney test),[18] and are also of significant magnitude. Thus, there is clear support for Hypothesis 1.

Hypothesis 4 does not find support. When transfers are allowed, all pairwise comparisons of rates of cooperation are statistically different (p-values of two-sided Mann–Whitney test < 0.1 in all cases).[19,20] On the other hand, when transfers are not permitted, the rates of cooperation are not statistically different (p-values of two-sided Mann–Whitney test > 0.1 in all pairwise comparisons).[21] This is different from previous results (Rapoport and Chammah, 1965); perhaps the differences in the entries of the Prisoner's Dilemma game are not big enough, these effects are not stable, or these effects do not generalize to non-symmetric games.

In Fig. 6, we consider only those sessions in which transfers were possible, and display cooperation rates as a function of whether mutual cooperation was a Nash equilibrium in the subgame induced by the transfer pair chosen.

In games where transfers are allowed, cooperation rates are lowest if the reward-pairs chosen are not consistent with mutual cooperation being a Nash equilibrium in the subgame. These differences are statistically significant in all treatments (two-sided p-values of Sign test < 0.01 in all treatments), lending support to Hypothesis 2. Cooperation rates are substantially (10 to 25 percentage points) higher for reward pairs on the 'border' of the NE-region, with a further substantial (18 to 38 percentage points) increase for reward pairs in the interior of the NE-region. This difference between the border and the interior is statistically significant (one-sided p-values of Sign test < 0.05 in all treatments). This is the first observation against Hypothesis 3. Namely, some SPE-consistent reward-pairs are less likely than others to lead to cooperation.

In Fig. 7 we show the proportions of the reward-pairs that were variously consistent with mutual cooperation in an SPE were such that mutual cooperation was a Nash equilibrium in the

[18] When performing hypothesis tests of this sort, we will do it for both subject averages and group averages to eliminate correlation across time. If results are sensitive to the unit of observation, it will be noted. Otherwise, as in this particular case, the result holds in both cases.

[19] The p-value of the Kruskal–Wallis test is less than 0.01.

[20] Note that this result is robust to whether we restrict attention to SPE transfer pairs.

[21] The p-value of the Kruskal–Wallis test which examines the hypothesis that the samples are from the same population is also greater than 0.1.

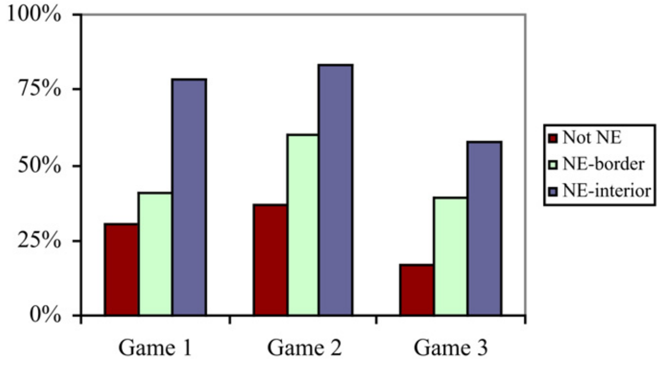

Fig. 6. Cooperation rates, by reward-pair consistency with equilibrium.

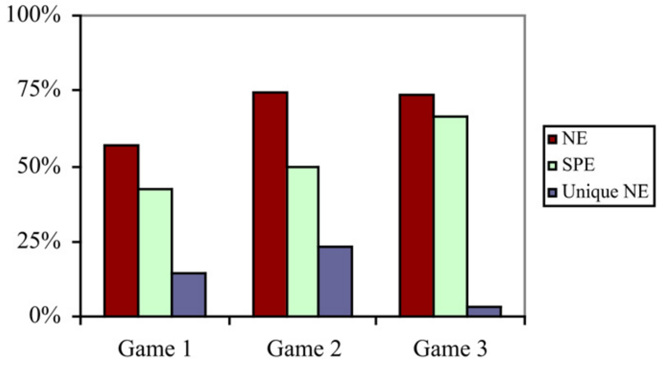

Fig. 7. Proportion of reward-pairs such that cooperation is consistent with NE and SPE.

subgame or were in the region where the transfer pairs make mutual cooperation the unique Nash equilibrium.

The proportion of joint transfers that make mutual cooperation a Nash equilibrium is rather high, about 68% across the three games, with a lower proportion in Game 1. Further, most endogenous reward-pairs were consistent with an SPE involving mutual cooperation.

We now consider how the likelihood of mutual cooperation is affected by how a reward-pair, consistent with an SPE involving mutual cooperation, affects the difference in net payoffs with mutual cooperation: Table 1 reports the rates of mutual cooperation as a function of how transfers affect the final payoffs. This is done for SPE transfers and SPE transfers excluding the NE border (by border we mean the cases where at least one of the two subjects is indifferent between cooperation and defection given that the other person cooperates).

We can see the rate of mutual cooperation for SPE-consistent reward-pairs is highest (or tied for highest) in all cases when these transfers are exactly equal. In four of the six cases, mutual cooperation is nearly as likely when the reward-pair brings the players' mutual-cooperation payoffs closer together, while in two cases it is substantially less likely. We also see that mutual cooperation is always least likely when the qualifying reward-pair makes the mutual-cooperation payoffs further apart; this is a second observation against Hypothesis 3. Overall, there were many more

Table 1
Percentages of mutual cooperation, by net transfer category

Transfers make MC payoffs:	Diverge	Equal	Closer	Total
All SPE Transfers				
Game 1	40% (20)	77% (22)	52% (127)	54% (169)
Game 2	47% (30)	67% (45)	61% (122)	60% (197)
Game 3	23% (120)	35% (37)	32% (109)	29% (266)
SPE Transfers, excluding NE Border				
Game 1	60% (10)	94% (17)	65% (101)	69% (128)
Game 2	55% (22)	73% (41)	71% (86)	64% (159)
Game 3	25% (81)	40% (25)	40% (86)	33% (192)

Number of observations is in parentheses.

qualifying reward-pairs chosen that reduce the difference in mutual-cooperation payoffs than the opposite direction (631 to 283).[22]

One natural question is how the data evolve over time. For instance, Chen and Gazzale (2004) find, in a related environment, that the fraction of equilibrium play evolves substantially over the course of their experiment. Hence, they estimate three different learning models to determine which one fits the data best; this allows them to simulate the behavior predicted by that model. However, there does not appear to be any systematic change in behavior over the course of our experiments. For example, the fraction of subgame-perfect equilibrium play (both transfer choices and choice of play) is not statistically different for the first and last five periods.

This is also confirmed if we perform a regression analysis similar to that of Chen and Gazzale and estimate a probit regression where the dependant variable takes value 1 if the play was SPE and 0 otherwise.[23] The resulting coefficient estimates for the interacted treatment dummy-ln(period) regressors are not statistically significant either individually or jointly (at the 10% level). Note also that, as we shall see, the fact that the treatments with exogenous transfers resulted in similar levels of cooperation suggests that changes over time are not a crucial part of what we observe. That is, in the treatments with exogenous transfers, there is no endogenous evolution of transfers, yet cooperation levels were similar to when transfers are determined by the subjects. It is also possible that his lack of evidence of learning is an artifact of having a shorter horizon than in Chen and Gazzale, who used 60 periods in comparison to our 25 periods.

4.2. Regression analysis

Table 2 reports random-effects probit estimates of the determinants of cooperation where the regressors are what the subject offers to pay (*Would pay*), what the subject is offered (*Would receive*) again interacted with a dummy variable for the case where cooperation should result in equilibrium, a dummy variable taking value 1 if the transfers are such that they are on the border of the NE region and 0 otherwise, and an indicator variable for when transfers are equal and one for when they make final payoffs closer. This is estimated on all transfers such that cooperation is a NE. The data are separated in two cases: closer is better (CB), meaning all the cases where having transfers that make final payoffs closer imply that the subjects' own payoffs are higher

[22] There are only 187 cases where the sufficient transfers are equal, since this is the most stringent of the conditions.

[23] The regressors are treatment dummies (excluding one dummy) as well as treatment dummies interacted with ln(period) using clustered standard errors at the level of the group.

Table 2
Determinants of cooperation random-effects probit and marginal-effects estimates in NE region

	Closer is better		Closer is worse	
	RE probit	Marginal effects	RE probit	Marginal effects
Would pay	0.002	0.001	−0.018	−0.006
	(0.011)	(0.004)	(0.019)	(0.006)
Would receive	0.093***	0.031***	0.128***	0.041***
	(0.018)	(0.006)	(0.023)	(0.008)
NE border[a]	−0.603***	−0.216***	−0.746***	−0.263***
	(0.166)	(0.063)	(0.179)	(0.067)
Equal transfers[a]	0.475**	0.140***	0.398*	0.114**
	(0.194)	(0.050)	(0.219)	(0.056)
Final payments are closer[a]	0.536***	0.180***	−0.035	−0.011
	(0.155)	(0.053)	(0.173)	(0.055)
Game 2	0.450*	0.143*	0.347	0.107
	(0.273)	(0.082)	(0.306)	(0.055)
Game 3	−0.406	−0.138	−0.733**	−0.246**
	(0.267)	(0.094)	(0.309)	(0.108)
Constant	−1.012**		−0.411	
	(0.399)		(0.539)	
Observations	820	820	820	820
Number of subjects	94	94	96	96

Standard errors in parentheses.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.
[a] Marginal effects report the change in probability when the regressor goes from 0 to 1.

(Type 1 in Games 1 and 2, and Type 2 in Game 3) and closer is worse (CW), which is the opposite (Type 2 in Games 1 and 2, and Type 1 in Game 3).[24] The table also reports the marginal effects for the average subject at the sample mean of the regressors, except for dichotomous ones where it gives the difference in probabilities when the variable equals 1 or 0.

Clearly the amount one is offered matters. That regressor is always positive and statistically significant. This is a third observation that is not in line with Hypothesis 3. However, how much one would pay never has a statistically-significant impact. If the transfers are exactly on the NE border, then cooperation is less likely, confirming what we had already noticed from the summary statistics. If the transfers are exactly the same, then cooperation is more likely. Finally, cooperation is more likely if transfers narrow the payoff difference from mutual cooperation, but only when the difference of transfers is in the player's favor. These last observations also

[24] Appendix C, on the *Games and Economic Behavior* supplementary materials website, provides results by game and role (row or column). Results are similar but two observations should be made. First, the NE border effect is not as strong, meaning that although it has a negative impact in all games for both roles, it is statistically significant in only half of them. Second, the effect of equal transfers is statistically significant in half the cases, and it has the opposite sign in one of the cases where it is not.

We have also tested to see if controlling for periods affected the results. We have done this by including the period and the period squared as regressors or by including indicator variables for blocks of five periods. In neither case did it have any qualitative impact on the results. Furthermore, the majority of coefficient estimates of the effect of period (for both specifications) were statistically insignificant. In the interest of space these are not included but are available from the authors on request.

do not square with Hypothesis 3, nor with what one might expect given the current models of social preferences. That is, since the only condition where equality seems to matter is when it is self-advantageous, people do not seem willing to forego pecuniary benefits in favor of greater equality of distributions. On the other hand, they do not necessarily prefer to increase inequality even if it favors them.

The marginal effects inform us of each factor's relative importance.[25] Would receive averages 14 and 13 in CB and CW respectively (with standard deviations of 5 and 7). Thus, in the case of CB, increasing the amount a subject is offered by one standard deviation has the same impact as making the transfers equal, increasing the probability of cooperation by 15 percentage points. Making the transfers such that final payoffs are closer increases cooperation by 21 percentage points while being on the NE border reduces cooperation by 23 percentage points. The effects are similar in CW except that the effect of what one is offered is greater, while the effect of having equal transfers is slightly less; having closer final payoffs has no impact.

Thus far we have analyzed determinants of individual behavior. Such behavior has implications for the groups, and we now turn to a more detailed analysis of the factors affecting mutual cooperation. We estimate a random-effects probit model of the determinants of mutual cooperation, reported in Table 3.[26] Besides indicator variables for equal transfers and transfers that make

Table 3
Determinants of mutual cooperation random-effects probit and marginal-effects estimates in SPNE region

	Game 1	Game 2	Game 3	All games	All games: Marginal
NE border[a]	−2.224***	−1.294***	−0.080	−0.943***	−0.321***
	(0.527)	(0.350)	(0.353)	(0.210)	(0.060)
Sum of transfers	0.053	−0.001	0.102***	0.067***	0.026***
	(0.053)	(0.041)	(0.030)	(0.021)	(0.008)
Equal transfers[a]	0.895	0.170	0.682*	0.543**	0.213**
	(0.637)	(0.390)	(0.356)	(0.236)	(0.092)
Final payments are closer[a]	−0.296	0.401	0.609**	0.350**	0.132**
	(0.465)	(0.316)	(0.297)	(0.188)	(0.070)
Game 1[a]				0.666***	0.259***
				(0.237)	(0.090)
Game 2[a]				1.268***	0.474***
				(0.247)	(0.081)
Constant	−0.782	0.350	−3.808***	−2.526***	
	(1.598)	(0.995)	(0.928)	(0.634)	
Observations	169	197	266	632	632
Number of group	69	66	85	220	220

Standard errors in parentheses.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.
[a] Marginal effects report the change in probability when the regressor goes from 0 to 1.

[25] For binary regressors this reports the change in probability when the regressor goes from 0 to 1 (keeping all other regressors at their sample mean).

[26] Of course these possibly imply complicated correlation structures across observations since they are a single observation for a pair of subjects. The estimates reported use the pair as the random effects. Estimates where individuals are taken as the random effect are provided in Appendix D, on the *Games and Economic Behavior* supplementary materials website. All results are qualitatively unchanged.

final payoffs closer, these include dummies for the NE border and for the sum of transfers. The same specification is estimated pooling all games, with indicator variables for Games 1 and 2 included and marginal effects reported. The analysis will focus on the pooled results.

Being on the NE border statistically hurts mutual cooperation, decreasing its probability by 32 percentage points and making it one of the most important effects in magnitude. The sum of the transfers also has a statistically-significant effect on the probability of mutual cooperation.[27] On average, the transfers sum to 24, with a range of 16 to 42. Hence going from lowest to largest would affect cooperation by 68 percentage points and a movement of one standard deviation (which is 5) would affect the probability by 13 percentage points. Having equal transfers make cooperation more likely than transfers that make final payoffs further apart by 21 percentage points while making them closer only increase by 13 percentage points. Game 2 has a higher probability of mutual cooperation, other things equal than all other games and Game 1 has a higher probability of mutual cooperation than Game 3. The significant coefficients for the Game 1 and Game 2 dummies in the rightmost columns are evidence against Hypothesis 4, as cooperation rates do vary across our games when transfers are permitted.

4.3. Exogenous transfers

Thus far, it has been implicitly assumed that the increased cooperation observed in the treatments with transfers occurs for the reasons posited by the model. However, in the treatments with endogenous transfers, players may make inferences from the compensation amounts chosen about the intended play of other players; perhaps affecting behavior.

To investigate this possibility, we conducted three additional sessions—one for each game—where the transfers were randomly selected from the previous experiments (200 for each game). Participants were shown the game without transfers, and in each period they were shown the transfers that would be in effect; they were not told how those transfers were determined. We selected transfer pairs at random from the previous experiments. To make the data analysis comparable between the exogenous and endogenous cases, we recomputed the results for the exogenous treatments on the subsample of randomly selected transfers. Consequently, results for the endogenous transfer case are slightly different from those in the preceding section.

Table 4 reports cooperation rates and statistical tests comparing the exogenous and endogenous conditions for different subsets of the data. We conduct tests in the aggregate for each game, as well as tests when transfers make mutual cooperation an equilibrium outcome in the stage game, or are transfers in the SPE-region in the game with endogenous transfers.

Table 4
Percentages of cooperation by transfer category

	Game 1		Game 2		Game 3	
	Endo	Exo	Endo	Exo	Endo	Exo
All	0.53	0.66	0.68	0.66	0.46	0.41
NE transfers	0.71	0.77	0.80	0.73	0.54	0.46
SPE transfers	0.72	0.74	*0.78*	*0.59*	0.53	0.44

Entries in italics indicate pairwise difference significant at the 10% level (two-tailed tests) using each subject as the unit of observation.

[27] We note that, holding the difference between transfers constant, an increase in the sum of transfers improves the payoff for a person who cooperates when the other person does not, thus making cooperation more attractive.

We do not see large differences overall across exogenous and endogenous transfers. The differences are modest for NE transfers and SPE transfers, with the cooperation rate slightly higher with endogenous transfers in Game 2 and Game 3, but not for Game 1. Only one comparison is statistically significant, and only at the 10% significance level. Rank-sum tests do not reject the null hypothesis that the cooperation rates are the same in the treatments with endogenous and exogenous transfers; this is true at the 10% level independently of whether averages over subjects or groups are used as the unit of observation.[28] Hence, intention does not seem to be the central force at play in the increased cooperation observed with transfers.

5. Discussion

We find a much higher rate of cooperation when players can choose contingent rewards for cooperation than when they cannot, in three different asymmetric variants of the Prisoner's Dilemma. The increased cooperation rate occurs not simply because transfers are feasible, but also depends substantially on the values of the reward-pairs. In all of our transfer games, cooperation rates are at least double for reward pairs in the interior of the NE transfer region than for reward pairs that lead to mutual cooperation not being a Nash equilibrium. Furthermore, reward-pairs on the border of the NE-region lead to intermediate levels of cooperation that are significantly different from both the low levels with no transfers permitted or the higher levels in the interior of the transfer region.

An innovative feature of our design is that the endogenous SPE reward-pairs induce *coordination games* in which mutual cooperation and mutual defection are both Nash equilibria. While mutual defection can be ruled out as a SPE action pair for all SPE reward-pairs in Games 1 and 2 (and nearly half of the feasible SPE reward-pairs in Game 3), doing so requires somewhat sophisticated arguments. Nevertheless, we see fairly high cooperation rates, with a range of between 42.9 and 68.1% in the three games; this compares to the 50.5% cooperation rate for the game in Andreoni and Varian (1999), which features cooperation being a dominant strategy for the unique (in integers) SPE-consistent reward-pair.

We find that cooperation is substantially more likely when it happens that the sufficient rewards chosen are equal. In a sense, this effect of the equality of payoffs on cooperation seems intuitive and focal, but it is nevertheless completely outside the current economic models of social preferences. This effect could possibly be seen as a mutual recognition of the problem, or as a simplification in the process that makes it seem more likely that the other player will cooperate. Perhaps there is something attractive about reaching the original targets of the payoffs from mutual cooperation. This could be similar to firms that collude or have mutually beneficial arrangements in which they try to control market shares.[29]

We also see that a player is more likely to cooperate when the reward-pair decreases the difference between the players' payoffs with mutual cooperation. However, this only has a significant effect when equality favors the chooser, so that one might view this as a form of self-serving bias. Reward-pairs that reduce the difference in the payoffs from mutual cooperation are more than twice as common as reward-pairs that cause this difference to increase.

---

[28] Pairwise comparisons using the rank-sum test also fail to reject the null for any of the three games (using subjects or groups).

[29] "In the case of the JEC [Joint Executive Committee], the cartel took the form of market share allotments rather than absolute amounts of quantities shipped" (Porter, 1983).

Looking across our three experimental games, we see some substantial differences in cooperation rates when contingent rewards are feasible. The size of the gains from mutual cooperation relative to the size of the mutual defection payoffs does seem to correspond to the cooperation rates in our games. These gains are 40 in both Games 1 and 2, compared to 20 in Game 3, and the ratio of these gains to the sum of the payoffs from mutual defection is 0.77, 0.91, and 0.33, respectively; the corresponding cooperation rates without contingent rewards are 15.8, 17.5, and 10.8%. There is also considerable differentiation among these games when transfers are allowed, with overall cooperation rates of 53.9, 68.1, and 42.9% in Games 1, 2, and 3, respectively. Perhaps one explanation for this stems from the symmetry of the SPE-region in Game 2, as seen in Appendix B, on the *Games and Economic Behavior* supplementary materials website. In our experiments, players are frequently changing roles; perhaps the symmetry makes it easier to use the same transfer choice or to identify it.

There is an additional dimension that differentiates Game 3 from the other two games. Recall that mutual cooperation is the only action pair consistent with SPE transfers in Games 1 and 2, but that mutual defection is also consistent with SPE transfers in Game 3 whenever neither H_1 nor H_2 is larger than 16 (as is true for more than 75% of the chosen transfer pairs in the SPE-region). In any event, it seems plausible that the increased uncertainty with lower transfers acts as a damper on the attraction toward mutual cooperation in Game 3, helping to explain the lower rate of mutual cooperation seen in Table 1.

5.1. Social preferences and cooperation

Given that in our games, subjects can affect the relative payoff positions by the specific choices of proposed transfers, we might wonder if, to the extent that we see deviations from the theory, models of social preferences can offer some useful insights. Thus far, we have assumed that people care only about own-money-maximization. We now examine the effects of social preferences on the possibility and likelihood of cooperation in our games, using the model presented in Fehr and Schmidt (1999) and the reciprocity-free portion of the Charness and Rabin (2002) model, and their adaptation to our two-player games.[30]

The basic idea behind the F-S model is that people care about the differences in material payoffs between players, as well as their own material payoffs. People dislike having less than the other person more than they dislike having less than the other person. The C-R model is more complex, and the full model involves reciprocity considerations. Nevertheless, the heart of the model is its distributional framework, which is directly applicable to our games. It consists of a weighted average of the total payoff of both players that itself embeds concerns for total payoffs and the lower of the two payoffs.

In an intuitive sense, one might expect these models to predict that mutual cooperation will be more attractive to a player when the payoffs from mutual cooperation are closer together, holding everything else constant. The intuition is as follows: In the Fehr and Schmidt (1999) model, people dislike differences in payoffs, ceteris paribus. In the Charness and Rabin (2002) model, if we hold total payoffs constant (as with a transfer payment), there is a preference for increasing the minimum payoff in the reference group, which is equivalent to having net transfers bring payoffs closer together. We provide a formal analysis in Appendix E, on the *Games and Economic Behavior* supplementary materials website, of the effect of these social preferences on

[30] Another prominent model is Bolton and Ockenfels (2000). As this model does not have a specific functional form, we have omitted this alternative model, in spirit having some similarities to the F-S model.

mutual cooperation in our games, as well as the effect of more egalitarian transfer pairs on the ‘attractiveness’ of cooperation. We also provide an illustrative example applying F-S and C-R preferences with different transfer pairs.

With respect to C-R preferences, mutual cooperation is always an equilibrium in the second stage in our games for *all* permissible parameter values for transfer pairs within the SPE-region. Each player's utility from mutual cooperation increases with more egalitarian transfer pairs when the C-R parameter values are relatively large.[31] However, when the parameter values are smaller the player who would receive a higher material payoff from mutual cooperation prefers less egalitarian transfer pairs to more egalitarian transfer pairs, while the opposite preferences hold for the other player. Nevertheless, there is a substantial range of sufficient parameter values; as previous experimental papers have demonstrated considerable heterogeneity for subjects' social-preference parameter values, it seems reasonable to presume that more egalitarian transfer pairs are more socially rewarding for some participants, thus leading to the observed result that cooperation is more likely with more egalitarian transfer pairs.[32]

With respect to F-S preferences, it is possible that the player who would receive (net of transfers) the smaller material payoff from mutual cooperation would prefer to defect rather than to cooperate when the other player cooperates. However, for Game 1 or Game 3, the parameter value range that drives such a preference shrinks as the transfer pairs become more egalitarian, and it is empty in the limit when transfers are completely egalitarian (each player's material payoff from mutual cooperation is identical). Although the range is not empty with completely egalitarian transfers for Game 2, it does become very small. Our F-S example in Appendix E, on the *Games and Economic Behavior* supplementary materials website, illustrates how even a small parameter value can make defection by the disadvantaged player a best response to cooperation when transfer pairs cause the material payoffs from mutual cooperation to diverge, while the parameter value must be much higher for this to occur with a more egalitarian transfer pair. We thus conclude that, with F-S preferences, mutual cooperation is more likely with more egalitarian transfer pairs.

Thus, both C-R and F-S preferences are consistent with some of the results. For instance, they both indicate that for many parameter values, transfers that make final payoffs closer are preferred, something for which we find some indications in the data. However, they do not shed light on why equal transfers have such a strong effect on cooperation. As such, it is not clear that either model does much better than the analysis with standard atomistic preferences.

6. Conclusion

Achieving cooperation in the Prisoner's Dilemma has challenged theorists for decades. Most previous studies have focused on repetitions for solutions: either finite (see Bereby-Meyer and Roth, 2003) or infinite (see Aoyagi and Fréchette, 2006; Dal Bo, 2002, and Duffy and Ochs, 2003). We test experimentally an endogenous reward mechanism described in Varian (1994) and further elaborated in Qin (2002), where ‘gifts’ are contingent upon cooperation in a one-shot

[31] By this we mean that $\lambda(1+\delta) > 1$, where $\lambda, \delta \in [0, 1]$. For more details, see Appendix E, on the *Games and Economic Behavior* supplementary materials website.

[32] For example, consider a person who can unilaterally allocate either nothing or 25% of his or her endowment to a recipient who has no endowment. If he or she chooses to give away 25% (as many subjects would), calculations show (see Charness and Rabin, 2002, p. 852) that $(1-\lambda\delta) < 0.75(1-\lambda\delta) + 0.25\lambda\delta$, or $\lambda\delta > 0.5$. All combinations of λ and δ that satisfy this condition also satisfy $\lambda(1+\delta) > 1$ for $\lambda, \delta \in [0, 1]$.

environment. Typically, the contracting parties bind themselves in a way that usually leads to the efficient outcome in Games 1 and 2, with cooperation rates roughly quadrupled in all three experimental games. This increase in efficiency occurs despite the fact that the reward-pairs consistent with mutual cooperation being a subgame-perfect action pair induce a coordination game rather than a subgame in which mutual cooperation is the unique Nash equilibrium, as with the game in Andreoni and Varian (1999). Our results provide support for the Coase theorem in a difficult environment with a two-sided externality.

Our games have a substantial range of integer transfer pairs consistent with mutual cooperation being part of an SPE, and so we can examine patterns within this region to see which ones tend to be beneficial in fostering mutual cooperation and efficiency. Our analysis suggests at least two main prescriptions: First, sufficient transfers that make cooperation a strict best response to expected cooperation are considerably more effective than transfers that lead to indifference with expected cooperation. Second, sufficient transfers that are identical are particularly effective, followed by transfers that narrow the gap between the players' payoffs in the event of mutual cooperation; least effective are reward-pairs that make these payoffs diverge. We also find that transfer pairs that lead to mutual cooperation being an equilibrium outcome in the resulting subgame are effective at achieving cooperation even when they are not chosen by the players themselves. This may perhaps be seen as support for the importance of payoff dominance in equilibrium selection in coordination games.

Despite the difficulties inherent with asymmetric payoffs and sophisticated inferences, we observe a reasonably high degree of cooperation in our games, achieved with rewards for cooperation that are modest in size relative to the payoffs in the game. This seems a hopeful sign for efficiency in contracting, as the choice of play in a coordination game is not as obvious as when one has a dominant strategy. We suspect that in practice it would be difficult for multiple parties to fashion such suitable transfers, given the substantial coordination problem, and we also suspect that large strategy spaces could pose problems for successful implementation. Nevertheless in less complex environments in commerce and in international affairs, there may be substantial scope for this compensation mechanism to achieve beneficial social outcomes, and also reason to be concerned about the ability of firms to design collusive agreements.

Acknowledgments

We thank Eric Maskin, Tom Palfrey, Catherine Weinberger, and seminar participants at the Institute for Advanced Study in Princeton and at Ben-Gurion University for helpful comments. All errors are our own. Charness and Qin gratefully acknowledge the financial support from the UCSB Academic Senate. Fréchette's research was partially supported by the Center for Experimental Social Science, the C.V. Starr Center and the National Science Foundation (Grant SES-0519045).

Supplementary Appendices

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.geb.2006.10.010.

References

Andreoni, J., Miller, J., 1993. Rational cooperation in the finitely repeated Prisoner's Dilemma: Experimental evidence. Econ. J. 103, 570–585.

Andreoni, J., Varian, H., 1999. Pre-play contracting in the Prisoner's Dilemma. Proc. Nat. Acad. Sci. 96, 10933–10938.

Aoyagi, M., Fréchette, G.R., 2006. Collusion as public monitoring becomes noisy: Experimental evidence. Mimeo.

Bereby-Meyer, Y., Roth, A., 2003. Learning in noisy games: Partial reinforcement and the sustainability of cooperation. Mimeo.

Bolton, G., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. Amer. Econ. Rev. 90, 166–193.

Bracht, J., Figuières, C., Ratto, M., 2004. Relative performance of two simple incentive mechanisms in a public good experiment. Mimeo.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quart. J. Econ. 117, 817–869.

Chen, Y., Gazzale, R., 2004. The role of supermodularity in an experimental setting. Amer. Econ. Rev. 94, 1505–1535.

Coase, R., 1960. The problem of social cost. J. Law Econ. 3, 1–44.

Cooper, R., DeJong, D., Forsythe, R., Ross, T., 1996. Cooperation without reputation: Experimental evidence from Prisoner's Dilemma games. Games Econ. Behav. 12, 187–218.

Dal Bo, P., 2002. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. Mimeo.

Dawes, R., 1980. Social dilemmas. Ann. Rev. Psychol. 31, 169–193.

Duffy, J., Ochs, J., 2003. Cooperative behavior and the frequency of social interaction. Mimeo.

Falkinger, J., 1996. Efficient private provision of public goods by rewarding deviations from average. J. Public Econ. 62, 413–422.

Falkinger, J., Fehr, E., Gachter, S., Winter-Ebmer, R., 2000. A simple mechanism for the efficient provision of public goods: Experimental evidence. Amer. Econ. Rev. 90, 247–264.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quart. J. Econ. 114, 817–868.

Fischbacher, U., 1999. z-Tree—Zurich toolbox for readymade economic experiments—Experimenter's manual. Institute for Empirical Research in Economics, University of Zurich.

Hamaguchi, Y., Mitani, S., Saijo, T., 2003. Does the Varian mechanism work? Emissions trading as an example. Int. J. Bus. Econ. 2, 85–96.

Harrison, G., McKee, M., 1985. Experimental evaluation of the Coase theorem. J. Law Econ. 28, 653–670.

Hoffman, E., Spitzer, M., 1982. The Coase theorem: Some experimental tests. J. Law Econ. 25, 73–98.

Jackson, M., Wilkie, S., 2003. Endogenous games and mechanisms: Side payments among players. Mimeo.

Kahneman, D., Knetsch, J., Thaler, R., 1990. Experimental tests of the endowment effect and the Coase theorem. J. Polit. Economy 98, 1325–1348.

Moore, J., Repullo, R., 1988. Subgame perfect implementation. Econometrica 56, 1191–1220.

Porter, M., 1983. A study of cartel stability: The joint executive committee, 1880–1886. Bell J. Econ. 14, 301–314.

Qin, C.-Z., 2002. Penalties and rewards as inducements to cooperate. Working paper in economics #13-02. UCSB. Available on-line at http://www.econ.ucsb.edu/~qin.

Rapoport, A., Chammah, A., 1965. Prisoner's Dilemma. Univ. of Michigan Press, Ann Arbor.

Roth, A., 1988. Laboratory experimentation in economics: A methodological overview. Econ. J. 98, 974–1031.

Roth, A., Murnigham, J., 1978. Equilibrium behavior and repeated play of the Prisoner's Dilemma. J. Math. Psychol. 17, 189–198.

Varian, H., 1994. A solution to the problem of externalities when agents are well-informed. Amer. Econ. Rev. 84, 1278–1293.

Varian, H., 1995. Coase, competition, and compensation. Japan World Economy 7, 13–27.

Williamson, O.E., 1983. Credible commitments: Using hostages to support exchange. Amer. Econ. Rev. 73 (4), 519–540. September.

Ziss, S., 1997. A solution to the problem of externalities when agents are well-informed: Comment. Amer. Econ. Rev. 87, 231–235.