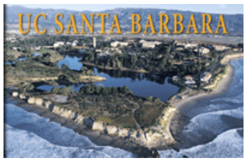


Title:
Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test Of Cooperation And Coordination

Author:
Charness, Gary B, Economics Department, University of California, Santa Barbara
Qin, Cheng-Zhong, University of California, Santa Barbara

Publication Date:
02-06-2005

Series:
Departmental Working Papers

Publication Info:
Departmental Working Papers, Department of Economics, UCSB, UC Santa Barbara

Permalink:
http://escholarship.org/uc/item/9cm846c4

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

Abstract:
We study experimentally a two-stage compensation mechanism for promoting cooperation in prisoner's dilemma games. In stage 1, players simultaneously choose binding non-negative amounts to pay their counterparts for cooperating in a given prisoner's dilemma game, and then play the prisoner's dilemma game in stage 2 with knowledge of these amounts. For the asymmetric prisoner's dilemma games we consider, all payment pairs consistent with mutual cooperation in subgame-perfect equilibrium transform these prisoner's dilemma games into coordination games, with both mutual cooperation and mutual defection as Nash equilibria in the stage-2 game. We find considerable empirical support for the mechanism, as cooperation is much more common when these endogenous transfer payments are feasible. We identify patterns among transfer pairs that affect the likelihood of cooperation. Mutual cooperation is most likely when the payments are identical; it is also substantially more likely with payment pairs that bring the payoffs from mutual cooperation closer together than with payment pairs that cause them to diverge. There is substantial scope for this compensation mechanism to achieve beneficial social outcomes in commerce and in international affairs, and reason to be concerned about the ability of firms to design collusive agreements.

Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test Of Cooperation And Coordination

Gary Charness, Guillaume Fréchette & Cheng-Zhong Qin*

February 6, 2005

Abstract: We study experimentally a two-stage compensation mechanism for promoting cooperation in prisoner's dilemma games. In stage 1, players simultaneously choose binding non-negative amounts to pay their counterparts for cooperating in a given prisoner's dilemma game, and then play the prisoner's dilemma game in stage 2 with knowledge of these amounts. For the asymmetric prisoner's dilemma games we consider, all payment pairs *consistent* with mutual cooperation in subgame-perfect equilibrium transform these prisoner's dilemma games into *coordination games,* with *both* mutual cooperation and mutual defection as Nash equilibria in the stage-2 game. We find considerable empirical support for the mechanism, as cooperation is much more common when these endogenous transfer payments are feasible. We identify patterns among transfer pairs that affect the likelihood of cooperation. Mutual cooperation is most likely when the payments are identical; it is also substantially more likely with payment pairs that bring the payoffs from mutual cooperation closer together than with payment pairs that cause them to diverge. There is substantial scope for this compensation mechanism to achieve beneficial social outcomes in commerce and in international affairs, and reason to be concerned about the ability of firms to design collusive agreements.

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

JEL Classifications: A13, B49, C72, C78, C91, K12

---

* Charness is at the Department of Economics, University of California Santa Barbara, Santa Barbara, CA 93106-9210, http://www.econ.ucsb.edu/~charness/, email: charness@econ.ucsb.edu, fax: 805-893-8830; Fréchette is at New York University, Department of Economics, 269 Mercer Street, 7th floor, New York, NY 10003, http://www.people.hbs.edu/gfrechette/html/econ.htm, e-mail: gfrechette@hbs.edu, and Qin is at the Department of Economics, University of California Santa Barbara, Santa Barbara, CA 93106-9210, http://www.econ.ucsb.edu/faculty/~qin/, email: qin@econ.ucsb.edu. Financial support from the UCSB Academic Senate is gratefully acknowledged.

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 Rapaport 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. 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.[1] 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.[2] Qin (2002) characterizes the conditions on payment pairs that are necessary and sufficient to "induce the players to cooperate" (to be defined shortly).[3]

To illustrate the mechanism, consider this example:

Figure 1: A prisoner's dilemma game

		Player 2	
		C	D
Player 1	C	10, 13	2, 15
	D	13, 2	7, 6

[1] 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).

[2] 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.

[3] 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 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.

Suppose player i ($i=1,2$) offers to pay an amount H_i to player j on the condition that player j cooperates. Then, if the payments H_1 and H_2 are large enough, mutual cooperation will follow in the subsequent play of the prisoner's dilemma game. For example, if $H_1 = H_2 = 10$, then players' payoffs in the subsequent play will be:

Figure 2: Transformation of the game in Figure 1 by payment pair (10,10)

		Player 2	
		C	D
Player 1	C	10, 13	12, 5
	D	3, 12	7, 6

In this case, (C,C) is the unique Nash equilibrium. However, the strategy of choosing a payment of 10 and then subsequently cooperating cannot survive subgame-perfection. To see this, suppose that player 2 chooses payment 10. Then, cooperating is dominant for each player in the subsequent play of the prisoner's dilemma game as long as player 1 chooses a payment greater than 4. Consequently, player 1 can increase her payoff by lowering her payment to, say 5.

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 (almost a fourfold decrease). 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 transfer payments make the payoffs from mutual cooperation closer together or further apart than without the transfers. Cooperation rates are highest when qualifying transfer payment pairs are identical, while transfer payments that cause the gap in the mutual-cooperation payoffs to grow are less effective than transfer payments that cause the gap to shrink.

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.[4] 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 fundraising, 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

[4] Williamson (1983) discusses the merits of crafting *ex ante* incentive structures for the prisoner's dilemma.

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.[5] 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 permits 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 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

[5] 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*.

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, *ceterus paribus,* one might expect cooperation rates to be lower in our case.

To describe the compensation mechanism more formally, consider a situation that is captured by a prisoner's dilemma game in Figure 3 below:

Figure 3: The generic prisoner's dilemma game

		Player 2	
		C	D
Player 1	C	R_1, R_2	S_1, T_2
	D	T_1, S_2	P_1, P_2

$S_k < P_k < R_k < T_k$, $k = 1, 2$.

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. A payment pair (H_1,H_2) changes players' payoffs associated with the various action pairs to:

Figure 4: The transformed generic game

		Player 2	
		C	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

How might one interpret this mechanism in a field context? Suppose that two firms agree to participate in a joint venture, and each firm can either contribute minimum effort (D) or full effort (C), which is more costly. If both firms choose minimum effort, the resulting payoffs are (P_1, P_2). However, each firm can pledge to pay a performance reward, perhaps by depositing money (to be paid upon observing full effort) in an escrow. With suitable endogenous reward payments, we should expect better performance in the joint venture.[6]

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 a 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 $i \neq j$,

$$T_i - R_i \leq H_j^* \leq R_j - S_j, \quad (1)$$

$$H_j^* - H_i^* \leq R_j - P_j, \quad (2)$$

$$H_j^* \leq T_i - R_i \text{ whenever } P_i - S_i \leq T_i - R_i, \quad (3)$$

and

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

[6] In real joint ventures, of course one may not know *ex ante* the precise values of the exogenous parameters. However, these values can be considered in expected terms, if we assume that the firms are risk neutral. Similarly, it may not be feasible to perfectly monitor effort, but we might well be able to observe factors that are correlated with this effort, such as specific investments or elements of the firm's performance.

The set containing all such payment pairs H^* often describes a rectangle.[7] 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 $8 \leq H_1^* \leq 16$ and $12 \leq H_2^* \leq 20$. Consider the transfer pair $(H_1^*, H_2^*) = (12, 16)$.[8] Figure 5 illustrates how one such payment pair transforms Game 1:

Figure 5: Game 1 transformation by (H_1,H_2) = (12,16)

Game 1 with no transfers

Player 2

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

Game 1 after $(H_1,H_2) = (12,16)$

Player 2

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

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).

[7] 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.

[8] One might ask why the SPE of (H_1^*,H_2^*) = (12, 16) isn't 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 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). The SPE here implies a certain degree of bargaining between the two players.

To see this, suppose on the contrary that a payment pair satisfying (1) – (4), or equivalently satisfying $8 \leq H_1^* \leq 16$ and $12 \leq H_2^* \leq 20$, induces (D, D). Then, player 1 receives 28 and player 2 receives 24 from the corresponding SPE. However, from Game 1 together with Figure 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 $52 - H_2 + H_1^* \geq 60 - H_2$. 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 $8 \leq H_1^* \leq 16$, player 2 would wish to change his transfer from $12 \leq H_2^* \leq 20$ 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.

For some parameterizations, there is a unique transfer pair that, in combination with mutual cooperation, is consistent with subgame perfection (i.e., on the equilibrium path). Andreoni and Varian (1999) test a prisoner's dilemma with $(S_1,P_1,R_1,T_1) = (0,3,6,9)$ and $(S_2,P_2,R_2,T_2) = (0,4,7,11)$. Conditions (1) – (4) imply that $H_1^* = 4$ and $H_2^* = 3$; Andreoni and Varian point out that this is the unique SPE when the side payments can be any real number, but that there is an additional SPE when the side payments are restricted to be integers. In this equilibrium, supported by pessimistic expectations when the other player is indifferent between cooperation and defection, both players add one unit to the side payments, leading to the transfer pair (5,4) and this induced subgame:

Figure 6: The transformed Andreoni and Varian (1999) game

		Player 2 C	Player 2 D
Player 1	C	5, 8	4, 7
	D	4, 5	3, 4

Here cooperation is the strictly-dominant strategy for each player and so (C,C) is the only Nash equilibrium in the subgame.[9]

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 a 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 is not likely to be an issue where these payment pairs make cooperation the dominant strategy for each player, but 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

[9] We note that the transfer pair (4,3) induces the subgame:

		Player 2 C	Player 2 D
Player 1	C	5, 8	3, 8
	D	5, 4	3, 4

Here each player receives the same payoff regardless of his or her own choice and so each and every combination of pure or mixed strategies is a Nash equilibrium in this subgame.

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.

Figure 7: Our experimental games

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

Appendix B illustrates, for each of these games, the bounds of the SPE-transfer pairs consistent with 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.[10,11]

For purposes of statistical analysis, there is a multiple-observation problem, 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.[12]

The Experiment

We conducted a series of experiments in nine separate sessions at the University of California at Santa Barbara. We had three sessions for each of three different prisoner's-dilemma games. For each game, endogenous transfers were permitted in two of these sessions, while the third session served as a control. There were 16 participants in each session, with

---

[10] 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 $H_2^* > 16$.

[11] In this case the mechanism has two subgame-perfect equilibria; 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.

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

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.[13]

We provided instructions on paper, which were discussed at the beginning of the session; a sample of these instructions is presented in Appendix A. 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.

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.

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 a 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.*

---

[13] Since part of what we wish to study are decisions conditional on transfers making cooperation a 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.

Since cooperation is a Nash equilibrium in the induced subgame for only 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 leads us to suspect that secondary factors will influence the choice of play in the subgame, thereby falsifying Hypothesis 3. For example, reward pairs that are on the Nash 'border' seem less likely to lead to cooperation. Consider Game 1, with $8 \leq H_1^*$ and $12 \leq H_2^*$. Suppose the transfer pair is $(H_1^*,H_2^*) = (12,12)$, on the southern border of the Nash or SPE regions. The induced subgame, where both (C,C) and (D,D) are Nash equilibria, is:

Figure 8: Game 1 transformed by (H_1,H_2) = (12,12)

Player 2

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

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 doesn't 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 (which choice does better if the other person randomly chooses whether or not to cooperate) might serve to help select the equilibrium in the induced coordination game.

Finally, it is possible that social preferences come into play, even for 'qualifying' reward pairs. Since rewards are simply transfers and do not change the total social payoff contingent upon the players' actions in the subgame, there is no role for 'efficiency' *per se*. However, the reward pairs chosen affect the difference between the players' payoffs (or alternatively, the minimum payoff) that result from mutual cooperation. We might expect the likelihood of cooperation to increase as the difference in the net payoffs from mutual cooperation decreases, in line with the Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002) utility models.

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 a 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.*

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 2x2 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.

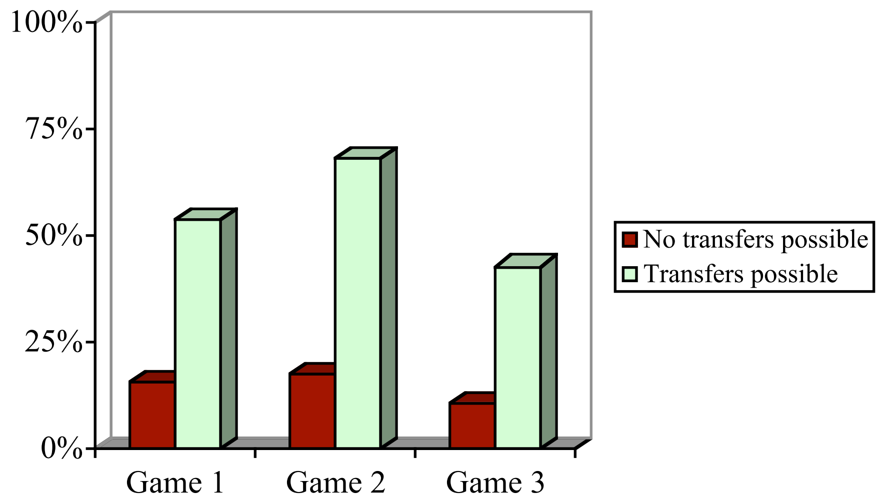
4. Experimental Results and Hypothesis Tests

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.

Data Summary

Figure 9 shows the average cooperation rates by game and treatment:

Figure 9 - Cooperation rates, by game



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. All of these differences are highly statistically significant (p-value < 0.01, one-sided Mann-Whitney test),[14] and are also of significant magnitude. Thus, there is clear support for Hypothesis 1.

Hypothesis 4 on the other hand finds mixed support. In the conditions without transfer, the rates of cooperation are not statistically different (p-values of two-sided Mann-Whitney test > 0.1 in all pair-wise comparisons).[15] This is different from previous results (Rapoport and Chammah 1965) and might indicate that the differences in the entries of the prisoner's dilemma

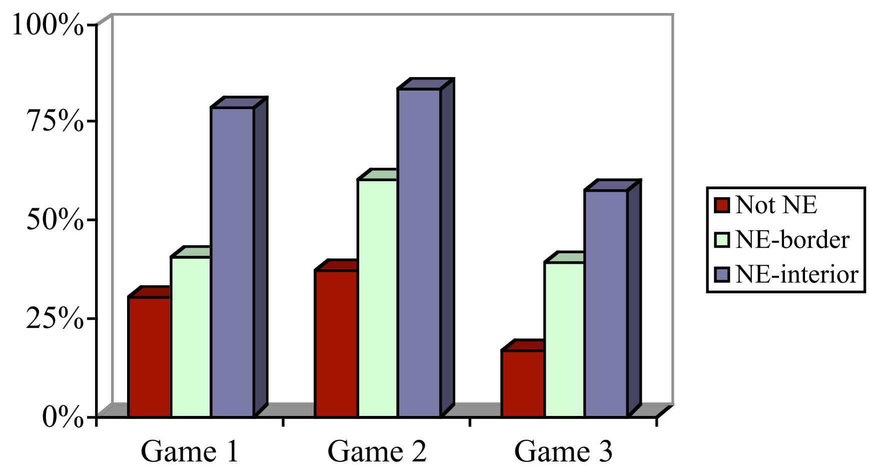
[14] When performing hypothesis tests of this sort, we will do it for both subject averages and group averages to eliminate correlation across time. If results are sensitive to the unit of observation, it will be noted. Otherwise, as in this particular case, the result holds in both cases.

[15] 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.

game are not big enough, that these effects are not stable, or that these effects do not generalize to non-symmetric games. But when transfers are allowed, we do observe differences. All pair-wise comparisons of rates of cooperation are statistically different (p-values of two-sided Mann-Whitney test < 0.1 in all cases).[16]

In Figure 10, 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:

Figure 10 - Cooperation rates, by reward-pair consistency with equilibrium



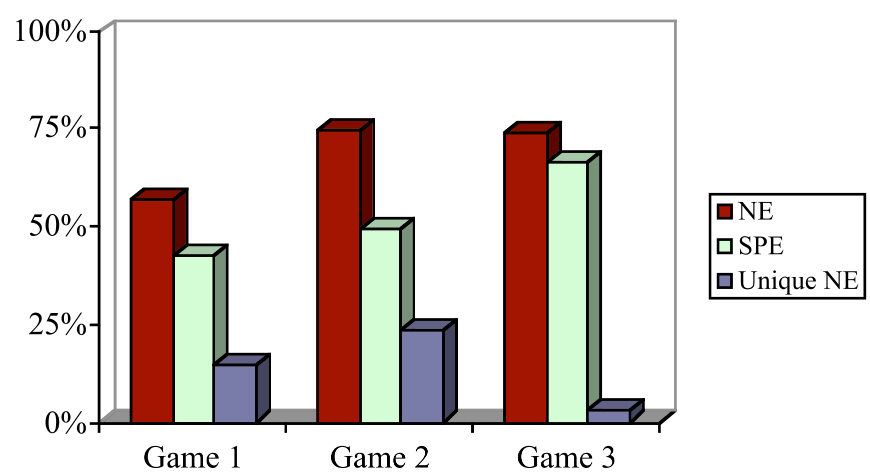
In games where transfers are allowed, cooperation rates are lowest if the reward pairs chosen are not consistent with mutual cooperation being a Nash equilibrium in the subgame. These differences are statistically significant in all treatments (two-sided p-values of Sign test < 0.01 in all treatments) thus 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

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

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 Figure 11 we show the proportions of the reward pairs that were variously consistent with mutual cooperation in a SPE, were such that mutual cooperation was a Nash equilibrium in the subgame, or were in the region where the transfer pairs make mutual cooperation the unique Nash equilibrium:

Figure 11 - Proportion of reward-pairs such that cooperation is consistent with NE and SPE



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, more than half of all endogenous reward pairs were consistent with a SPE involving mutual cooperation.

We now consider how the likelihood of mutual cooperation is affected by how a reward pair, consistent with a 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 whether transfers make final payoffs closer, further or if they remain the same. 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):

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 in Parentheses)

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, and also doesn't exactly square with what one might have expected; that is, given the evidence on social preferences (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, and Charness and Rabin 2002) one might have expected that cooperation is highest when transfers make final payoffs closer.

However, this is not the case. Overall, there were many more qualifying reward pairs chosen that reduce the difference in mutual-cooperation payoffs than the opposite direction (631 to 283).

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 (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).[17] 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 and when it equals 0.

[17] Appendix C 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.

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	0.001	0.000	-0.013	-0.004
	(0.011)	(0.004)	(0.019)	(0.006)
Would Receive	0.093***	0.031***	0.136***	0.042***
	(0.018)	(0.006)	(0.023)	(0.008)
NE Border§	-0.637***	-0.231***	-0.746***	-0.255***
	(0.166)	(0.063)	(0.186)	(0.067)
Equal Transfers§	0.535***	0.157***	0.496**	0.131***
	(0.194)	(0.050)	(0.219)	(0.051)
Final Payments are Closer§	0.603***	0.206***	0.077	0.023
	(0.156)	(0.055)	(0.172)	(0.053)
Constant	-1.054***		-0.729	
	(0.346)		(0.485)	
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%
§ Marginal effects report the change in probability when the regressor goes from 0 to 1.

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 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.[18] 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.[19] Besides indicator variables for equal transfers and transfers that make final payoffs closer, these include a dummy for the NE border, and a dummy for the sum of transfers. The same specification is estimated pooling all games, with indicator variables for game 1 and game 2 included and marginal effects reported. The analysis will focus on the pooled results.

[18] 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).

[19] 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. All results are qualitatively unchanged.

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§	-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§	0.895	0.170	0.682*	0.543**	0.213**
	(0.637)	(0.390)	(0.356)	(0.236)	(0.092)
Final Payments are Closer§	-0.296	0.401	0.609**	0.350**	0.132**
	(0.465)	(0.316)	(0.297)	(0.188)	(0.070)
Game 1§				0.666***	0.259***
				(0.237)	(0.090)
Game 2§				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%
§ Marginal effects report the change in probability when the regressor goes from 0 to 1.

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. 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.

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.

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, DeJong, Forsythe, and Ross (1996) and Andreoni and Miller (1993) observed 10%, 25%, and 18% cooperation rates, respectively. As the individual cooperation rate in the no-transfer game in Andreoni and Varian (1999) was 26% and so higher than our base rates, it appears that endogenous transfers were at least as effective, relative to the no-transfer case, in enhancing cooperation via induced coordination games as when mutual cooperation is the unique Nash equilibrium in the subgame.

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.[20]

A more standard form of social preference appears to also influence behavior, as we see 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 can perhaps 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.

Looking across our three experimental games, we see some substantial differences in rates when contingent rewards are feasible. The size of the gains from mutual cooperation

[20] "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).

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. In our experiments, players are frequently changing roles; perhaps the symmetry makes it easier to use the same transfer choice or to identify it.[21]

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.

The success of the reward mechanism in our induced coordination games is achieved with contingent rewards that are much smaller than in games where there is a unique SPE reward pair that leads to mutual cooperation being the only Nash equilibrium in the induced subgame. For example, in Andreoni and Varian (1999), the transfer payments required for SPE in integers is nine, compared to total payoffs of seven with mutual defection (78%). In contrast, in Game 1

[21] In fact, there were 27 instances where the reward pair in Game 2 was (9,9), the minimum point inside the SPE region (of which 23 resulted in mutual cooperation); the corresponding minimum point in Game 1, (9,13), is less 'focal' and is never observed.

the equivalent total transfer required is 22, compared to the mutual-defection total payoffs of 52 (42%). In Game 2, the comparison is 18 to 44 (41%); in Game 3, the comparison is 18 to 60 (30%). In this sense, perhaps the players could be said to achieving success with less exposure. Nevertheless, we hesitate to draw any strong conclusion in this regard, as the populations are different and there may well be other uncontrolled sources of variation as well.

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 2004, 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 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 very difficult environment with a two-sided externality.

Our games have a substantial range of integer transfer pairs consistent with mutual cooperation being part of a 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; avoid reward pairs that make these payoffs diverge.

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. There is substantial scope for this compensation mechanism to achieve beneficial social outcomes in commerce and in international affairs, and reason to be concerned about the ability of firms to design collusive agreements.

References

Andreoni, J. and J. Miller (1993), "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal*, **103**, 570-585.

Andreoni, J. and H. Varian (1999), "Pre-play contracting in the prisoner's dilemma," *Proceedings of the National Academy of Sciences*, **96**, 10933-10938.

Aoyagi, M. and G. Fréchette (2004), "Cooperation in Repeated Games with Imperfect Monitoring," mimeo.

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

Bolton, G., and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, **90**, 166-193.

Charness, G., and M. Rabin (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, **117**, 817-869.

Coase, R. (1960), "The problem of social cost," *Journal of Law and Economics*, **3**, 1-44.

Cooper, R., D. DeJong, R. Forsythe, and T. Ross (1996), "Cooperation without reputation: Experimental evidence from prisoner's dilemma games," *Games and Economic Behavior*, **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," *Annual Review of Psychology*, **31**, 169-193.

Duffy, J. and J. Ochs (2003), "Cooperative Behavior and the Frequency of Social Interaction," mimeo.

Fehr, E., and K. Schmidt (1999). “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics*, **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.

Harrison, G. and M. McKee (1985), “Experimental evaluation of the Coase theorem,” *Journal of Law and Economics*, **28**, 653-670.

Hoffman, E. and M. Spitzer (1982), “The Coase theorem: Some experimental tests,” *Journal of Law and Economics*, **25**, 73-98.

Jackson, M. and S. Wilkie (2003), “Endogenous games and mechanisms: Side payments among Players,” mimeo.

Kahneman, D., J. Knetsch, and R. Thaler (1990), “Experimental tests of the endowment effect and the Coase theorem,” *Journal of Political Economy*, **98**, 1325-1348.

Moore, J. and R. Repullo (1988), “Subgame Perfect Implementation,” *Econometrica*, **56**, 1191-1220.

Porter, M. (1983), “A Study of Cartel Stability: the Joint Executive Committee, 1880-1886,” *The Bell Journal of Economics*, **14**, 301-314.

Qin, C.-Z. (2002), “Penalties and rewards as inducements to cooperate,” UCSB Working Paper in Economics #13-02 (available on line at www.econ.ucsb.edu/~qin).

Rapoport, A and A. Chammah (1965), *Prisoner's Dilemma*, Ann Arbor: University of Michigan Press.

Roth, A. (1988), “Laboratory experimentation in economics: A methodological overview, *Economic Journal*, **98**, 974-1031.

Roth, A. and J. Murnigham (1978), “Equilibrium behavior and repeated play of the prisoner's dilemma, *Journal of Mathematical Psychology*, **17**, 189-198.

Varian, H. (1994), “A solution to the problem of externalities when agents are well-informed,” *American Economic Review*, **84**, 1278-1293.

Ziss, S. (1997), “A solution to the problem of externalities when agents are well-informed: Comment,” *American Economic Review*, **87**, 231-235.

APPENDIX A

INSTRUCTIONS

Thank you for participating in our experiment. You will receive $5 for showing up on time, plus you will receive your earnings from the choices made in the session.

There will be 25 periods. In each period, each person will be matched with one other person. The person with whom you are matched will be randomly re-drawn after every period. You are paired anonymously, which means that you will never learn the identity of the other person in any of the periods.

One person will have the role of ROW and the other person will have the role of COLUMN. Your role will also be randomly re-drawn in each period, so that sometimes you will have the role ROW and sometimes you will have the role of COLUMN.

Here is the basic game:

		COLUMN Left	**COLUMN** Right
ROW	Up	40, 52	8, 60
	Down	52, 8	28, 24

The ROW and COLUMN players make choices simultaneously. The ROW player chooses Up or Down; the COLUMN player chooses Left or Right.

The **1st number in each cell** refers to the **payoff** (in cents) **for the** ROW **player**, while the **2nd number in each cell** refers to the **payoff** (in cents) **for** the COLUMN **player**. Thus, for example, if ROW chooses Up and COLUMN chooses Left, the ROW player would receive 40 and the COLUMN player would receive 52.

However, before these game choices are made, ROW may choose a binding amount to be paid (*transferred*) by him or her to COLUMN if and only if COLUMN chooses Left. COLUMN (at the same time) may offer a binding amount to be transferred to ROW if and only if ROW chooses Up. These amounts must be non-negative integers.

The amounts that you each choose will be communicated to each of you **prior** to your choices in the game above. You will then make your game choice (Up or Down if you are ROW, or Left or Right if you are COLUMN). You will then learn your payoff for the period, from which you can infer the game choice made by the person with whom you are paired.

This completes one period of play. We'll do 25 periods and pay people individually and privately.

FURTHER EXPLANATION

Offers to pay money contingent on the other person choosing Up (or Left, if the other person is a COLUMN player) have the effect of changing the payoff matrix. Note that whatever amount you state will be transferred to the other person if he or she plays Up as a ROW player or Left as a COLUMN player; **this money will be transferred regardless of your game choice.**

Suppose, for example, that ROW offers to pay $\$x$ to COLUMN if COLUMN plays Left and COLUMN offers (independently and simultaneously) to pay $\$y$ to ROW if ROW plays Up. Then the payoff matrix becomes:

		COLUMN Left	Right
ROW	Up	$40 + y - x$, $52 + x - y$	$8 + y$, $60 - y$
	Down	$52 - x$, $8 + x$	28, 24

We explain the 4 possible outcomes below. Remember, the values of x and y are always determined by the ROW and COLUMN players, respectively, before making game choices.

1) If ROW chooses Up and COLUMN chooses Left, then ROW must pay x units to COLUMN and COLUMN must pay y units to ROW. Thus, ROW would receive $40 + y - x$ and COLUMN would receive $52 + x - y$.

2) If ROW chooses Up and COLUMN chooses Right, then COLUMN must pay y units to ROW, but ROW pays nothing to COLUMN (because COLUMN did not choose Left). Thus, ROW would receive $8 + y$ and COLUMN would receive $60 - y$.

3) If ROW chooses Down and COLUMN chooses Left, then ROW must pay x units to ROW, but COLUMN pays nothing to ROW (because ROW did not choose Up). Thus, ROW would receive $52 - x$ and COLUMN would receive $8 + x$.

4) If ROW chooses Down and COLUMN chooses Right, then neither player pays the other anything. Thus, ROW would receive 28 and COLUMN would receive 24.

We don't wish to illustrate this with an example with realistic numbers, as this could bias your behavior. However, we can use an example where $x = 999$ and $y = 1000$. (We don't expect anyone to choose these values for x and y.) In this case, the payoff matrix becomes:

		COLUMN Left	Right
ROW	Up	41, 51	1008, -940
	Down	-947, 1007	28, 24

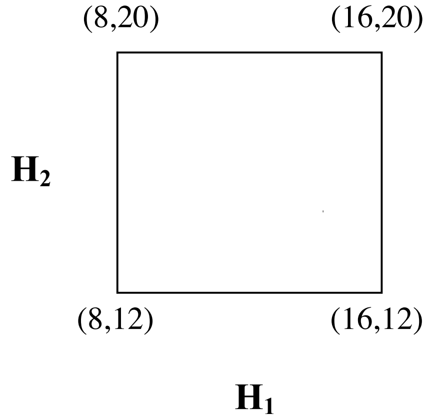
We encourage people to work out scenarios on paper, drawing a game matrix for each possibility.

Are there any questions? Please feel free to ask, by raising your hand.

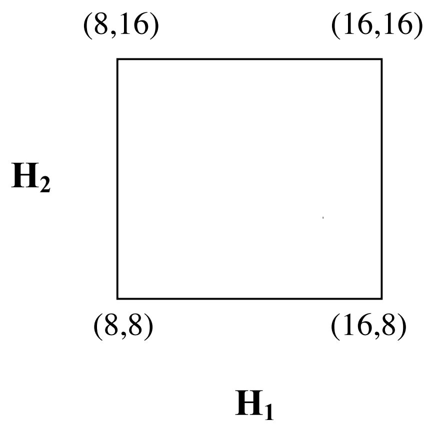
APPENDIX B

Transfer-pair regions consistent with (C,C) being a subgame-perfect action pair

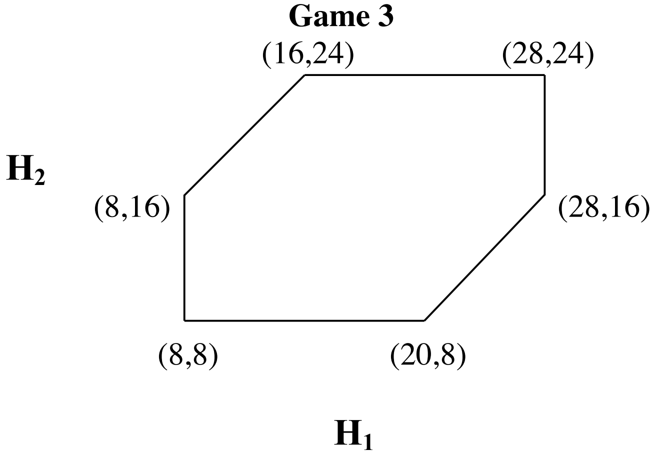
Game 1



Game 2



Game 3



APPENDIX C

Determinants of cooperation
Random-effects probit estimates in NE region

	Game 1: Row	Game 1: Column	Game 2: Row	Game 2: Column	Game 3: Row	Game 3: Column
Would Pay	-0.03	-0.013	0.056	-0.011	-0.013	0.029
	(0.019)	(0.049)	(0.056)	(0.036)	(0.029)	(0.026)
Would Receive	0.102*	0.151***	0.113**	0.012	0.172***	0.076***
	(0.054)	(0.055)	(0.050)	(0.051)	(0.033)	(0.022)
NE Border	-0.562	-1.359***	-0.990***	-1.087***	-0.067	-0.45
	(0.371)	(0.386)	(0.344)	(0.326)	(0.280)	(0.288)
Equal Transfers	0.935*	0.054	0.432	-0.435	0.711**	0.522*
	(0.535)	(0.556)	(0.362)	(0.486)	(0.315)	(0.301)
Final Payments are Closer	-0.262	-0.368	1.129***	-1.060**	0.425*	0.725***
	(0.353)	(0.372)	(0.393)	(0.438)	(0.249)	(0.247)
Constant	-0.185	-0.349	-1.702*	2.055**	-2.085***	-1.671***
	(0.939)	(1.304)	(0.946)	(0.986)	(0.636)	(0.578)
Observations	228	228	298	298	294	294
Number of Subjects	31	32	31	32	32	32

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

APPENDIX D

Determinants of mutual cooperation
Random-effects probit with one way subject error terms
and marginal-effects estimates in SPE region

	Game 1	Game 2	Game 3	All Games	All Games: Marginal
NE Border	-2.138***	-1.275***	-0.113	-0.936***	-0.321***
	(0.460)	(0.328)	(0.286)	(0.183)	(0.053)
Sum of Transfers	0.035	0.035	0.076***	0.057***	0.022***
	(0.045)	(0.050)	(0.024)	(0.018)	(0.007)
Equal Transfers	0.875	0.183	0.512*	0.487**	0.191**
	(0.561)	(0.363)	(0.281)	(0.199)	(0.078)
Final Payments are Closer	-0.179	0.357	0.468**	0.340**	0.129**
	(0.411)	(0.297)	(0.219)	(0.157)	(0.058)
Game 1				0.608***	0.237***
				(0.214)	(0.082)
Game 2				1.008***	0.385***
				(0.212)	(0.075)
Constant	-0.401	-0.535	-2.858***	-2.146***	
	(1.387)	(1.198)	(0.721)	(0.543)	
Observations	169	197	266	632	632
Number of Groups	28	28	32	88	88

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%