ELSEVIER

Journal of Economic Behavior & Organization
Vol. 46 (2001) 357–377

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Dilemma games: game parameters and matching protocols

David Schmidt [a,b], Robert Shupp [c], James Walker [a,b,*],
T.K. Ahn [b,d], Elinor Ostrom [b,d]

[a] Department of Economics, Wylie Hall, Indiana University, Bloomington, IN 47405, USA
[b] Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN 47408, USA
[c] Department of Economics, Ball State University, Muncie, IN 47306, USA
[d] Department of Political Science, Indiana University, Bloomington, IN 47405, USA

Received 9 June 1999; received in revised form 16 November 2000; accepted 18 November 2000

Abstract

This study examines the impact of changes in pecuniary payoffs and the linkages between players in the game environment on strategy choice in repeated PD games. Rates of cooperation are found to be correlated with variations in game parameters related to concepts of Fear, Greed, and Cooperators' Gain. In addition, rates of cooperation are affected by the information that subjects have regarding the history of play of their counterparts, as well as the protocol used for matching subjects across decision rounds. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: C92; H41

Keywords: Public goods; Prisoner's dilemma; Experimental; Reputation

---

1. Introduction

Experimental studies of social dilemma games typically find that the level of cooperation observed in such games is not consistent with non-cooperative game-theoretic equilibrium predictions when all players are assumed to be purely self-interested, in the sense that they are taken to care only about their own pecuniary rewards. The level of cooperative choices, however, varies substantially from one experimental design to another.

To explain this substantial variation, the theoretical literature focuses on factors within a game and those in the environment surrounding a particular play of a game that are

---

[*] Corresponding author. Tel.: +1-812-855-2760.
E-mail address: walkerj@indiana.edu (J. Walker).

PII: S0167-2681(01)00158-5

posited to affect individual motivation and behavior. For instance, considerable debate exists regarding the relative weight of pecuniary payoffs; subjects' orientations to fairness, trust, or reciprocity; and how players in a game are linked to one another (Bolton and Ockenfels, 2000; Camerer, 1990; Kahneman et al., 1986; Fehr and Schmidt, 1999; Ochs and Roth, 1989; Gunnthorsdottir et al., 2000). Strong evidence exists that *all* of these factors affect, to some extent, the level of cooperation observed in social dilemma situations. These studies suggest that the assumption that subjects who participate in experimental studies are narrowly self-interested is suspect. Thus, equilibrium predictions based on this assumption are also suspect. If uncertainty about how subjects rank the possible outcomes exists, the experiment is best viewed as a Bayesian game.

In a particular experimental setting, where the number of players and the sets of admissible actions are held constant, we see four basic elements related to the structure of the experimental setting that affect behavior. These are:

- pecuniary benefits,
- player types,
- information about player types, and
- linkages between players that occur in repeated game situations.

These four elements can be utilized to explain observed behavior through their impact on player motivations and beliefs regarding the play of others. In the next section, we present a general discussion of these four elements. Following this discussion, we present results from a series of experimental designs that allow us to investigate how variations in game parameters, the information subjects have regarding the history of play of their counterparts, and the protocol used for matching subjects across decision rounds impact cooperation rates.

2. Structure and linkages

2.1. *Pecuniary benefits*

In non-cooperative game theory, predicted game equilibria depend on the utilities that players receive from alternative outcomes. When a game is operationalized in a laboratory, the traditional approach has been to assume that a player's pecuniary payoff is an appropriate proxy for the player's utility level. Much of the current debate within experimental social sciences is related to understanding when this approach yields predictions consistent with behavior. Once it is suggested that pecuniary benefits may not be a good proxy for utility, one has to be careful about describing experimental environments using standard game-theoretic terminology. Throughout this paper, when we describe an experimental setting as a PD game, for instance, we mean that the game matrix obtained by using pecuniary payoffs as proxies for utilities has the structure of a PD.

In the one-shot PD game, both players have a dominant strategy called defection (D). When both players choose their dominant strategy, the outcome is Pareto inferior to the outcome resulting when both players play their dominated strategy called cooperation (C). Experimental research across numerous designs based on games with PD structure in pecuniary benefits, however, does not support the prediction that everyone will defect in such games. Even when double blind test procedures are used in one-shot experiments, a significant number of individuals choose the dominated strategy of cooperation (see Hayashi et al., 1999, for relevant cites). Among the many possible explanations for this result, one that must be considered is that the pecuniary benefits do not fully account for the subjects' rankings of the outcomes.

Since we will only consider symmetric games, the payoffs can be described by defining a payoff function, $\pi(\cdot, \cdot)$, where the first argument is the player's own action and the second is the action of the other player. PD games are constructed so that $\pi(\mathrm{D}, \mathrm{C}) > \pi(\mathrm{C}, \mathrm{C})$ and $\pi(\mathrm{D}, \mathrm{D}) > \pi(\mathrm{C}, \mathrm{D})$. In comparing the relative payoffs of these outcomes, we adopt three terms used in Rapoport and Chammah (1965) to describe the players' incentives in the game. First, if player 2 plays C, player 1's payoff is higher if he/she plays D rather than C. This gain, $\pi(\mathrm{D}, \mathrm{C}) - \pi(\mathrm{C}, \mathrm{C})$, is referred to as *Greed*. Similarly, if player 2 plays D, player 1's payoff is again higher if he/she plays D instead of C, by an amount of $\pi(\mathrm{D}, \mathrm{D}) - \pi(\mathrm{C}, \mathrm{D})$. This difference in payoffs is referred to as *Fear*. On the other hand, an incentive at the level of the group exists to play C, because both players receive higher monetary payoffs by playing (C, C) than when they play (D, D). The magnitude of this incentive, $\pi(\mathrm{C}, \mathrm{C}) - \pi(\mathrm{D}, \mathrm{D})$, is referred to as Cooperators' Gain.

In an experimental setting, if players' utilities are well represented by their monetary payoffs, subjects should defect in a one-shot PD no matter the objective level of Fear, Greed, or Cooperators' Gain. Prior studies have shown, however, that increases in Fear or Greed lower the rate of cooperation in PD games (Rapoport and Chammah, 1965; Rapoport, 1988; Saijo and Nakamura, 1995) and that increases in Cooperators' Gain increase the rate of cooperation (Komorita et al., 1980).[1]

2.2. Player types

Scholars from diverse traditions have put forward the proposition that some subjects rank outcomes based on things other than their own monetary payoffs (Sen, 1974; Raub, 1990). Psychologists have long argued that the social values held by an individual affect how they transform a pecuniary payoff matrix into an effective matrix (Kelley and Thibaut, 1978; Kramer et al., 1986; Kuhlman et al., 1986). Instead of assuming that only one type of player exists who maximizes pecuniary payoffs — ignoring the outcomes received by others or the morality of actions leading to outcomes — many scholars assume that players hold diverse orientations to games based on their own internal valuation systems and their linkage to others. Internal values in turn affect how pecuniary payoffs are transformed into utility. Rabin (1993), for example, recommends that one should derive "psychological games" from basic "material games" and include in the transformation how players perceive the fairness of particular strategies taken by themselves and others (see also Binmore, 1998).

Recent evidence suggests that at least some subjects facing a PD game in pecuniary benefits do not perceive the game as a PD in utilities.[2] For example, in an experimental

---

[1] In a related study, Clark and Sefton (2001) found that increasing Greed significantly lowered levels of cooperation in one-shot PD games where one subject moves first, followed by the second subject who has knowledge of the choice of the first subject.

[2] See Hayashi et al. (1999) for a discussion of the literature related to this view.

investigation of a one-shot PD game, based on questionnaire data, Ahn et al. (1999) report that approximately 40% of a pool of 136 subjects revealed a preference ranking in which the outcome (D, C) was ranked below the outcome (C, C) and the outcome (C, D) was ranked below (D, D).[3] If two players with these reported preferences played the game, and it was common knowledge that they both had these preferences, the game would actually be an assurance game with two equilibria in pure strategies; (C, C) and (D, D). We will refer to a player with this type of preferences as being an "assurance-type". On the other hand, most other players report preference orderings consistent with D being a dominant strategy; we refer to these players as being "PD types".

Evolving theories posit different internal mechanisms to transform monetary payoffs into utilities that are consistent with the two types of preferences introduced above. In a model of inequity aversion developed by Fehr and Schmidt, individuals' utility functions incorporate negative components based on relative payoffs in a game in addition to objective earnings. In their model, individuals are assumed to differ with respect to an aversion to payoff outcomes in which their own payoff outcome is below or above the payoffs of other players in the game.[4] If an individual's inequity aversion is of sufficient magnitude, defect may no longer be a dominant strategy, and players could have what we refer to as assurance-type preferences. Crawford and Ostrom (1995) examine the role of norms represented by internal parameters added to or subtracted from objective payoffs on behavior. Güth and Yaari (1992) propose an indirect evolutionary approach whereby selection operates on the objective payoffs obtained from interactions, but positive or negative internal weights are placed on objective outcomes as a result of the norms adopted by individuals. The distribution of particular norms in a population is thus indirectly affected by how well the carriers of these norms do in repeated interactions with others carrying different or similar norms (see also Güth and Kliemt, 1998). This suggests a more dynamic model of the relationship between outcomes and utilities than the static models discussed above.

Theories of reciprocity also provide an explanation for diverse internal transformations of pecuniary payoffs into expected utilities. Instead of presuming that individuals cooperate out of a concern for others alone, or only because they feel an action is the right thing to do, reciprocity as an explanation implies that players make choices contingent on their beliefs concerning the likely actions of others and what these actions reveal about the intentions of others (see Falk and Fischbacher, 1998; Wagner, 1998; Hoffman et al., 1998; Ostrom, 1998; Croson, 1999). In a study closely related to the work presented here, Gunnthorsdottir et al. (2000) examine the consequences of sorting players by type in a multi-round public goods game. In their design, 12 subjects are sorted into groups of size four in each decision round. Two treatment conditions are investigated. In the first, subjects are randomly sorted into one of the three groups. In the second, subjects are sorted according to their contribution to the public good in the previous round. One group is composed of the four subjects making

---

[3] The questionnaire was administered, double blind, after subjects made their own game decision, but before subjects learned of the decision of the person with whom they were matched in the PD game. (See also Watabe et al. (1996), for similar findings from Japan, and Hayashi et al. (1999) for a comparison of responses from Japan and the US.)

[4] An alternative explanation might be found in the work of Palfrey and Prisbrey (1997). They specify a utility function, in the context of voluntary contribution experiments, which includes the warm-glow term and other subjects' payoffs as well as one's own monetary payoff. Also see Bolton and Ockenfels (2000).

the highest contributions, one is composed of the four subjects making the next highest contributions, and the last is composed of the four subjects making the lowest contributions. In both designs, subjects know only that they will be regrouped each round. They are not informed of the sorting procedure. Strong evidence is found supporting the notion of subject types. Based on first-period decisions, subjects are identified as free riders or cooperators. In both treatment conditions, cooperators tend to have higher contributions than free riders across decision rounds. A significant sorting effect exists. Contributions are lower when subjects are sorted randomly, as compared to when they are sorted based on contributions in the previous round. When free riders and cooperators interact more frequently, the rate of decay in contributions is greater than in the case where subjects are sorted according to types.

2.3. Information about player types

Once one accepts the possibility that more than one type of player exists, it is reasonable to assume that players also assume there are multiple types of players among those they encounter. Consequently, subjects' decisions may be influenced by the kind of information that they obtain about those with whom they are paired.

Cain (1998) found that expectations of play in a PD game were affected by information regarding prior play in a dictator game. Stingy players — those who retained at least 70% of their endowment in a prior dictator game — tended to predict that all players would defect. Nice players — those who gave away at least 30% of their endowment — tended to predict that nice players would cooperate and stingy players would defect (Cain, 1998; p. 151). Information about prior actions also affected behavior — specifically, nice players chose cooperation 69% of the time when they were paired with other nice players and 39% of the time when they were paired with stingy players.

Expectations of others' play may be derived from repeated play in a setting, or from experiences outside of the immediate setting. Evidence suggests that individuals may perceive different levels of obligation to respond reciprocally depending on the person with whom they are paired and whether there are pre-existing social linkages. Kollock (1998) found that the cooperation rates of UCLA fraternity brothers were strongly influenced by whether they thought they were paired with other members of their own fraternity, members of other fraternities, UCLA students at-large, USC students, or UCLA police officers.

Not only is information about the presence or absence of pre-existing social connectedness or other player types an important source of different behavior, knowing how much information about one's actions in a game will be conveyed to others is also a factor potentially affecting decisions. Knowing that your identity will be revealed to others after or during an experiment may increase the weight of internal norms triggered by external observation (Coleman, 1988; Crawford and Ostrom, 1995; Hoffman et al., 1996; Bohnet and Frey, 1998).

2.4. Linkages among players in repeated games

Within the literature on non-cooperative game theory, cooperation in the PD game is deemed possible if the game is repeated infinitely or if there are certain types of asymmetric information in a finitely repeated game. PD experiments typically last no more than a few hours, so the time horizon is clearly not infinite. The sequential equilibria described by Kreps et al. (1982) depend on the possibility of a player type who may respond to previous cooperative plays by playing cooperatively in the future. This result depends on players' abilities to form reputations, which is something that varies over different experimental designs.

If subjects have beliefs that some subjects have assurance-type preferences in PD experiments, PD-type players may mimic assurance-type players in the early rounds of a sequential equilibrium. Since most players would be acting as if the game was an assurance game, they would face an equilibrium selection problem between (C, C) and (D, D), where (C, C) is presumably the payoff dominant equilibrium. Experimental research on assurance games has shown that information about prior play, as well as matching protocols, affect players' abilities to resolve the equilibrium selection problem (Schmidt et al., 2000).

Further, the exact form of the linkages between players' actions and resulting outcomes may affect the level of cooperation. An example is whether the game is sequential or simultaneous (Clark and Sefton, 2001; Yamagishi and Cook, 1993). Finally, if players believe that the persons with whom they will be matched in future decisions depend on current decisions, such linkages could affect behavior. This form of linkage is one of the issues we address in the research reported here. [5]

3. Game parameters and experimental design

In this section, we introduce the specific game configurations used in our study and summary details related to the overall experimental protocol and design.

3.1. PD games: payoff parameters

The six PD games shown in Fig. 1 are used as the bases for examining variations in game parameters. With the intention of keeping the stakes comparable across the six games, the games were designed so that the maximum payoff was always 110 and the minimum payoff was always 10. The games were also designed to systematically vary values of Fear, Greed, and Cooperators' Gain. Several attributes of the games are worth noting. Due to our choice to keep the maximum and minimum payoffs constant, the sum of Fear, Greed, and Cooperators' Gain must always sum to 100. In games 1, 3, and 5, each change in Fear (Greed) is accompanied by a change in Greed (Fear), holding Cooperators' Gain constant. The same is true for games 2, 4, and 6. Further, in games 1, 3, and 5, the level of Cooperators' Gain is twice that of games 2, 4, and 6.

---

[5] In related research, Andreoni and Miller (1993) investigate random matching and fixed matching in two-person PD games. Subjects are paired with human subjects or computerized decision makers. Palfrey and Rosenthal (1994) investigate the impact of random and fixed matching in a repeated, step-level public goods game with groups of size three and four. Both studies find more cooperation in fixed match settings relative to those with random matching.

		Player 2	
		D	C
Player 1	D	40 40	110 10
	C	10 110	80 80

Game 1
Greed=30, Fear = 30, Cooperators' Gain=40

		Player 2	
		D	C
Player 1	D	50 50	110 10
	C	10 110	70 70

Game 2
Greed=40, Fear = 40, Cooperators' Gain=20

		Player 2	
		D	C
Player 1	D	50 50	110 10
	C	10 110	90 90

Game 3
Greed=20, Fear = 40, Cooperators' Gain=40

		Player 2	
		D	C
Player 1	D	60 60	110 10
	C	10 110	80 80

Game 4
Greed=30, Fear = 50, Cooperators' Gain=20

		Player 2	
		D	C
Player 1	D	30 30	110 10
	C	10 110	70 70

Game 5
Greed=40, Fear = 20, Cooperators' Gain=40

		Player 2	
		D	C
Player 1	D	40 40	110 10
	C	10 110	60 60

Game 6
Greed=50, Fear = 30, Cooperators' Gain=20

Fig. 1. Game parameters.

3.2. Overview: experimental protocol and design

Subjects were volunteers from introductory undergraduate courses in economics at Indiana University.[6] Prior to volunteering, subjects were informed that they would participate in a decision-making exercise and be paid in cash an amount dependent upon their decisions and the decisions of others in the experiment. Subjects were recruited to the experimental laboratory in cohorts of size eight. Upon arriving at the laboratory, subjects were randomly seated at computer monitors. Subjects were anonymously assigned an identification number. Subjects knew their own identification number, but could not associate any other subject in the room with that subject's identification number. After receiving an identification number, the subjects were informed of the conditions for the experiment.

---

[6] Less than 2% of the students in these classes are economics majors, although many are business majors.

Table 1
Experimental designs

Experimental design	Matching protocol across decision rounds	Method for grouping subjects in Phase 2	Number of experimental sessions
Design I	Random	Random grouping	3
Design II	Random	Endogenous grouping	3
Design III	Fixed	Endogenous grouping	3

Subjects privately read through a set of computerized instructions that explained the game situation. After reviewing the instructions for the game, the experimenter presented further instructions that explained the decision environment publicly on an overhead projector. The subjects did not know at any time with whom they were paired in any of the decision trials, and they were told that their decisions would remain anonymous throughout and after the experiment.

The games played were described as board games with a row and column player. Each subject made a choice between option “C” and option “D”. In the experiments, the strategy choices “D” and “C” were actually labeled as “A” and “B”. Subjects were given complete information about the payoff structure. Each subject always saw himself/herself as a row player. Subjects were informed that at the end of the experimental session they would privately receive their earnings in cash, plus US$ 5 for keeping their experiment appointment. Subjects were instructed to think of the payoffs in the individual games as “computer pesos”, where the conversion rate was 100 pesos equal US$ 1. Subjects, on average, earned US$ 19.60 and participated for approximately 1 hour.

The experimental results relate to outcomes from three experimental designs: Designs I, II, and III. Table 1 summarizes key aspects of the three designs. All experiments consisted of two parts: Phase 1 and Phase 2. Each phase included twelve decision rounds. In each phase, each of the six games discussed above was played twice. The order of play was based on game number, where the games were played in the following order: 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6. At the beginning of each experimental session, subjects were publicly informed of all aspects of the experimental design including: the number of rounds to be played, the structure of the experiment with regards to the two phases, the matching protocol that would be used to pair them each round, and the information conditions in each phase.

The designs differ in the way Phase 2 was conducted and/or the protocol used for matching subjects between decision rounds. Details related to Phases 1 and 2 are presented in the subsequent sections of the paper. Three experimental sessions were conducted in each design condition. Given that there are eight subjects per session and 24 decision rounds per session, this yields a total of 576 decisions in each design condition. Because each of the six games was played four times in each session, there are 96 decisions for each game specification in each design condition.

The presentation of results is organized around the order of investigation we used in this study. Designs I and II differ in only one respect and were investigated first. Results from these two designs then led to conducting Design III, which is very similar to Design II, but fairly different from Design I.

4. Designs I and II

4.1. Design conditions

Designs I and II utilize the same random matching protocol to pair subjects in each decision round of both phases and differ only in the manner in which Phase 2 was conducted.

4.1.1. Phase 1

Prior to Phase 1 of both designs, the subjects were informed that: (a) Phase 1 would consist of 12 decision rounds; (b) the decision game would vary between decision rounds; (c) all players would play the same game in each round; (d) each subject would be randomly matched with one of seven other subjects prior to each round, with replacement; and (e) the subjects would never know the identity of the person with whom they were matched in any particular decision round. Further, subjects were explicitly informed of the conditions for Phase 2.

4.1.2. Phase 2

Prior to Phase 2 of Design I, the subjects were informed that: (a) Phase 2 would consist of 12 decision rounds; (b) the decision game would vary between decision rounds; (c) every player would play the same game in each round; (d) following Phase 1, but prior to decisions in Phase 2, information would be publicly displayed about the decisions made in Phase 1; (e) prior to play in Phase 2 the group of eight subjects would be divided randomly into two groups of four subjects; and (f) the person with whom each subject would be matched each round would be randomly drawn from the three other subjects in their group. In particular, the subjects were informed that the public display would include, organized by subject identification number, the number of times a subject chose C, and the number of times that person chose D. Also displayed would be the number of times the subjects with whom a subject was matched chose C or chose D. The random assignment of subjects to Phase 2 groups was also publicly displayed. However, players were identified by subject number only.

Design II is identical to Design I, except for the way the two subgroups were chosen for Phase 2. Prior to Phase 1, subjects were informed that the grouping for Phase 2 would be based on the decisions of each subject in Phase 1. They were informed that the four subjects with the most choices of C would be grouped together and the four subjects with the most choices of D would be grouped together, and any “ties” would be dealt with by random assignment. No terms such as “cooperators” or “non-cooperators” were ever used.

These designs allow us to consider issues related to the pecuniary benefits, player types and information about player types, and linkages between players. The motivating idea behind the creation of these two designs was to see how behavior in Phase 2 would differ when subjects were grouped with individuals whose history of play was similar to their own, relative to the case where they were randomly grouped, which can be seen as relating to information about player types. In addition to the obvious analysis of varying the pecuniary benefits across the games, the two designs also permit investigation of additional questions. Do players cooperate more in Phase 1 if it means they will be likely to be put into a group of cooperators in Phase 2 (Design II) than if cooperation has no such affect (Design I)?

If so, do players respond differently to a particular record of Phase 1 cooperation under the two design conditions because the reasons for the Phase 1 cooperation may be different?

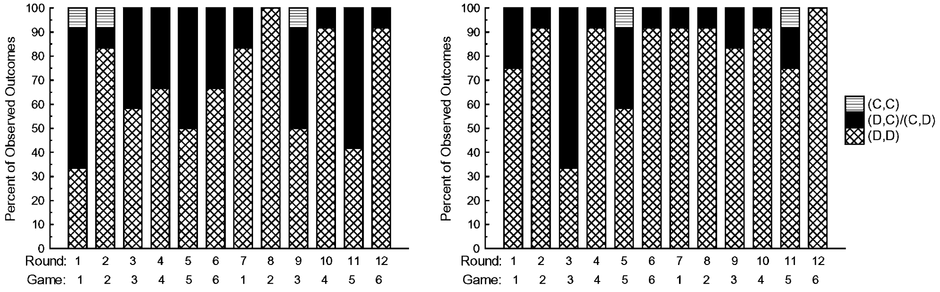
4.2. Results: Designs I and II

The presentation of results will first focus on several summary descriptive observations related to outcomes across decision rounds and across design conditions. This analysis will be followed by a more formal multivariate logistic regression analysis.

4.2.1. Observations across decision rounds

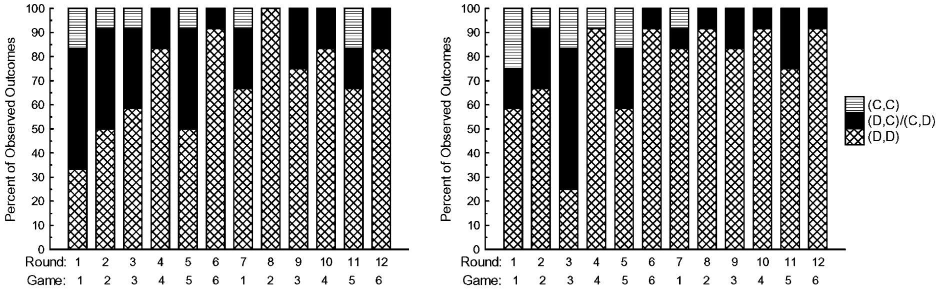
We begin the presentation of results with an overview of the outcomes observed across decision rounds in the two designs. Fig. 2 presents a summary for each decision round depicting the percentage of outcomes that fall into each of the possible categories: (C, C), (D, D), and either (C, D) or (D, C). Since the crosshatched areas represent both players defecting, and the black areas represent one of the two players defecting, overall defection rates can easily be obtained from these graphs by taking the midpoint of the black area. For instance, in Design I, Phase 1, Round 1, the black area goes from roughly 33 to 92%, so the overall defection rate for individuals is around 63%, and the overall cooperation rate is 37%.

Design I (12 observations per round)



Phase 1 Phase 2

Design II (12 observations per round)



Phase 1 Phase 2

Fig. 2. Game outcomes across all decision rounds.

Several observations are clear. The frequency of outcomes in which both players cooperate is generally very low, and the frequency of outcomes in which both players defect is very high. Note that cooperation rates are generally higher in odd-numbered rounds (Games 1, 3, and 5) than in even-numbered rounds (Games 2, 4, and 6). Recall from Fig. 1 that Cooperators' Gain is higher in the odd-numbered games than in the even-numbered games. This pattern may be caused, however, by some players responding in a particular way to the play they observed in the previous round, and not by the pecuniary benefits. Consequently, the regression analysis discussed below will attempt to control for such issues. In addition, there is no apparent difference in cooperation rates across the two design conditions. It appears that the endogenous group formation used in Phase 2 of Design II did not motivate subjects in Phase 1 of Design II to behave any more cooperatively than the subjects in Phase 1 of Design I. We now turn to a more detailed look at the behavior by each of the Phase 2 subgroups.

4.2.2. Phase 2 cooperation rates: Designs I and II

Recall, for matching purposes in Phase 2, that the group of eight subjects in a particular session was separated into two groups of four subjects. In Design I, subjects knew that the others with whom they would be randomly grouped in Phase 2 would be informed of the history of choices made and observed by each of the subjects in their group. So, subjects had a weak incentive to cooperate in Phase 1 in order to send a signal to potential Phase 2 matches. Subjects never knew the identity of their match, however. If a subject cooperates often in Phase 1 and randomly ends up in a group of low cooperators in Phase 2, the signal is likely to have little value. In Design II, the signaling component is also present and there is an additional reason why subjects may choose to cooperate in Phase 1. In this design condition, the Phase 2 groups were formed on the basis of grouping together subjects with similar histories.

Fig. 3 displays summary data on cooperation rates in Phase 2 for each group, for each design condition. Each data point in the figure displays the level of cooperation for a group

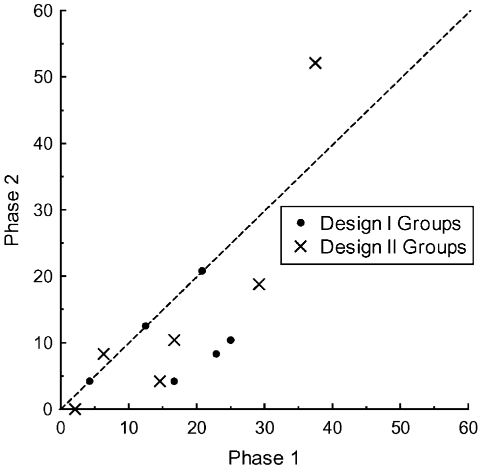


Fig. 3. Average cooperation rates (%) by Phase 2 groups.

Table 2
Logit analysis: Design I[a]

Variable	Phase 1 — fixed effects		Phase 2 — fixed effects	
	Coefficient	P-value	Coefficient	P-value
Constant	1.84	0.30	−2.63	0.98
Fear	−0.07	0.01	−0.13	0.00
Greed	−0.05	0.03	−0.14	0.00
PrevRound-(D, C)	0.19	0.77	–	–
PrevRound-(C, D)	−1.43	0.02	−0.27	0.83
PrevRound-(C, C)	0.60	0.62	–	–
No. of observations	165[b]		136[b]	
Nagelkerke R^2	0.27		0.43	
$-\ln L$	75.2		45.2	

[a] Coefficients for fixed effects dummy variables for each subject are not reported.

[b] In Phase 2 of Design I, no player who observed (D, C) ever cooperated in the following round (25 observations). The best estimate of the coefficient for PrevRound-(D, C) is thus negative infinity, and these observations have no influence on the estimates of the other parameters. Likewise, no cooperation was observed after a player observed (C, C) in Phase 2 of Design I (four observations). Consequently, we drop any observations in which either PrevRound-(D, C) or PrevRound-(C, C) takes on a value of 1, resulting in the omission of 29 observations. Furthermore, nine subjects in Phase 1 and nine subjects in Phase 2 played D in every round, so they were dropped from the dataset for purposes of this estimation because their maximum likelihood fixed effect is negative infinity.

of four in Phase 2 against the cooperation rate of those same four individuals in Phase 1 prior to them being split into subgroups. Several interesting patterns are apparent in the data. The most obvious observation is that there is no clear pattern of stratification between the Design I groups (dots) and the Design II groups (crosses). Further, there is evidence consistent with player type effects. A pronounced positive correlation exists between the level of cooperation of subgroups in their Phase 2 decisions and the level of cooperation

Table 3
Logit analysis: Design II[a]

Variable	Phase 1— fixed effects		Phase 2 — fixed effects	
	Coefficient	P-value	Coefficient	P-value
Constant	3.35	0.08	6.78	0.01
Fear	−0.10	0.00	−0.13	0.00
Greed	−0.09	0.00	−0.15	0.00
PrevRound-(D, C)	1.37	0.05	0.29	0.79
PrevRound-(C, D)	−0.94	0.20	1.13	0.18
PrevRound-(C, C)	2.35	0.00	2.53	0.01
No. of observations	198[b]		154[b]	
Nagelkerke R^2	0.31		0.45	
$-\ln L$	79.2		57.6	

[a] Coefficients for fixed effects dummy variables for each subject are not reported.

[b] Six subjects in Phase 1 and 10 subjects in Phase 2 played D in every round, so they were dropped from the dataset for purposes of this estimation because their maximum likelihood fixed effect is negative infinity.

these group members had displayed in Phase 1. This seems to hold for both of the two designs with no clear difference between the two.

4.2.3. Results based on a multivariate analysis

Tables 2 and 3 present results from logistic models designed to predict cooperative play in Designs I and II, respectively. Results are reported separately for each phase. Coefficient values and corresponding *P*-values are reported for each variable. The dependent variable, which can be thought of as the expected likelihood of cooperative play, is coded as 1 if the decision was C and 0 if the decision was D. It is regressed on the following set of variables, created to capture key elements of the game parameters, history of play within a decision environment, and subject type effects.

Based on game parameters:

Fear	the value of Fear in the specific game being played;
Greed	the value of Greed in the specific game being played.[7]

Based on decision environment:

PrevRound-(D, C)	1 if subject *i* did not cooperate in the previous round, but the player with whom he/she was paired did cooperate;[8]
PrevRound-(C, D)	1 if subject *i* cooperated in the previous round, but the player with whom he/she was paired did not cooperate;
PrevRound-(C, C)	1 if subject *i* and the player with whom he/she was paired both cooperated in the previous round.

Based on subject type:

Individual specific dummy variables specified to capture any idiosyncratic propensity to cooperate.

In each table, we report the results of two regressions meant to predict individual play of a game in a particular round. The first regression in each table considers only Phase 1 data, regressing the dummy variable indicating whether an individual cooperated in a particular round on the game parameters, the outcome of the previous round, and a fixed effect dummy variable for each of the 24 subjects.[9] The second regression in each table is identical in

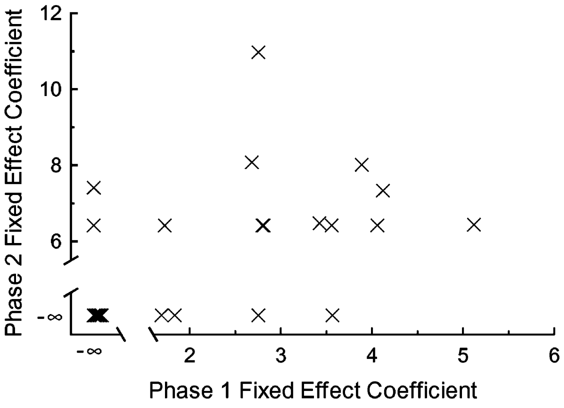
---

[7] Cooperators' gain was not included in the regression to avoid multicollinearity with the intercept term because Fear, Greed and Cooperators' Gain always sum to 100 in our games.

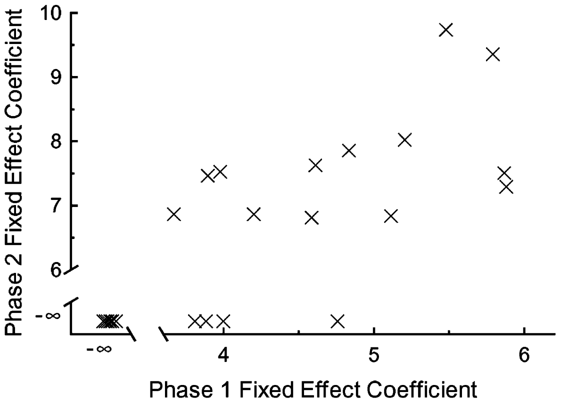
[8] We investigated the impact of lagging outcomes more than one round, but found nothing statistically significant.

[9] The individual fixed effects are being estimated based on 11 observations per subject. The maximum likelihood procedure only guarantees that these will be consistent asymptotically, so these estimates should be interpreted with care. However, we feel that the large differences we see between some individuals' fixed effects are not merely the result of small samples. A preferred method to deal with the heterogeneity in discrete choice models is to run a conditional logistic regression, which essentially estimates the regression coefficients by regressing the likelihood of observing a particular path of play for an individual, controlling for the total number of times the player cooperated, on the common explanatory variables. This regression yields parameter estimates on the common independent variables (Fear, Greed, and the prevround variables) that are consistent under less stringent conditions. We estimated a conditional logit for each phase of each design. The estimated coefficients and their significance levels were virtually identical to our fixed effects results, so we report only the fixed effects regressions. See Hsiao (1996) for further discussion of this issue.

structure to the first, but it is based on Phase 2 data. We should note that some observations had to be dropped from the dataset for this regression analysis. Some individuals did not cooperate in any round of a particular phase, which would not allow us to estimate a fixed effect for them with a logit regression. So we only analyzed the actions of individuals whose actions varied within a phase. Note that these subjects are behaving consistent with how we would predict that both PD types and assurance-types would act if they believed most other subjects were PD types or assurance-types that felt the equilibrium selection problem was best solved by playing (D, D). However, past experimental research shows that players in an assurance game tend to respond to the previous actions they have observed, whereas these players do not seem to respond to observed history of play (Schmidt et al., 2000).



Design I (Coefficient of Correlation = 0.50†)



Design II (Coefficient of Correlation = 0.68†)

† For fixed effects of negative infinity, the coefficients of correlation are calculated using asymptotic values.

Fig. 4. Correlation between Phase 1 fixed effects and Phase 2 fixed effects — Designs I and II.

Consequently, we view these players' actions as being more in line with PD-type preferences.

Observation 1: In both phases of Designs I and II, changes in Fear or Greed, with an accompanying change in Cooperators' Gain, have a significant negative impact on cooperation levels.

Recall, our game parameters were designed such that Fear, Greed, and Cooperators' Gain always equaled 100. Thus, by the nature of our design, the impact of changes in Fear and Greed should be interpreted in the context of an accompanying change in Cooperators' Gain. Our results suggest that when Fear (Greed) is increased and Cooperators' Gain is decreased by the same absolute amount, the combined effect is a decrease in the probability of cooperation, and vice versa.

Observation 2: In both phases of Designs I and II, subjects encountering a cooperative play by another subject in the previous round tend to respond with a higher probability of cooperative play in the next round. Subjects encountering a play of defection tend to respond with a lower probability of cooperative play.

From the estimated coefficients on PrevRound-(C, C), it appears that an encounter in which both subjects cooperated in the previous round (compared to the benchmark where both defected) had the largest and most systematic impact on the probability of cooperation in the next round in Design II. Note that this is exactly the behavioral response that Kreps et al. (1982) showed could generate cooperative play in finitely repeated PD games. If there is some sort of signaling of types going on in Design II, one explanation for why it does not show up in Design I is that the greater degree of randomness involved in the matching procedure makes it less likely that pairs of subjects who play (C, C) will end up in the same Phase 2 group, and hence have less incentive to signal their types.

If these fixed effects coefficients capture aspects of player types, then we would expect to see correlation between the estimated fixed effects coefficients for a particular individual across the two phases of an experiment. Fig. 4 plots estimated fixed effects coefficients for each subject, estimated in Phase 1 (first regression in Tables 2 and 3) and in Phase 2 (second regression). For this analysis, we also include maximum likelihood fixed effects coefficients of negative infinity for individuals who always defected. Considerable correlation between the fixed effects estimated in the two phases is apparent from this figure, especially in Design II.[10] Interpreting these coefficients as evidence of subject types suggests that subjects may indeed play these games in a manner dependent on particular subject characteristics. We view a player who always defected as being most consistent with the individual having PD-type preferences. The players with higher estimated coefficients responded to game parameters and/or previous outcomes, which would be consistent with players who may view some of the games as assurance games. These players' actions are also consistent

---

[10] We perform a more rigorous test of this correlation by pooling the data from each of the two phases within each design, and estimating one fixed effects coefficient per subject in a regression identical to those reported in Tables 2 and 3, except that we allow for different coefficient estimates for Fear, Greed, and the PrevRound variables for each of the two phases. We then perform a likelihood ratio test to see if restricting these fixed effects coefficients is a significant restriction on the parameters. We would reject the hypothesis that these fixed effects coefficients are the same across the two phases of Design I at a 10% significance level (but not at 5%), but would not reject the same hypothesis for Design II at any significance level below 90%. This corresponds with the higher correlation coefficient reported for Design II in Fig. 4 than those reported for Designs I and III.

with some notions of reciprocity, or tit-for-tat strategies, or any number of other potential “player types”. Given the great number of reasonable player types that may exist, we do not seek to identify an actual player type for each individual, but are simply looking for signs of consistency.

5. Design III

Designs I and II provided several insights into how game parameters and history of play within a repeated decision setting impact cooperation rates. These designs, however, also produced extremely low levels of cooperation, declining with repetition of the decision environment. This observation led us to consider a fixed matching protocol that is a logical extension of Design II.

In Design III, all design aspects of Design II remain the same, except for the fact that in round 1 of Phase 1, each subject is randomly matched with another subject and those subjects remain matched for all further rounds of Phase 1. In round 1 of Phase 2, each subject is randomly matched with another subject from his or her group, and each subject pair remains matched for all remaining rounds of Phase 2. As with both of the earlier designs, the matching procedure is public information. This design was constructed to reduce the coordination and signaling problems created with a random matching protocol.

5.1. Design III results

5.1.1. Observations across decision rounds

Fig. 5 displays summary measures on outcomes across decision rounds. In many ways, the patterns are similar to those depicted in Fig. 2, for Designs I and II. What is different, however, is the significant increase in the play of C.

5.1.2. Results based on multivariate analysis

Table 4 reports results related to a logit analysis of the data from Design III.

Observation 3: In Design III, cooperation rates are negatively correlated with changes in Fear and Greed.

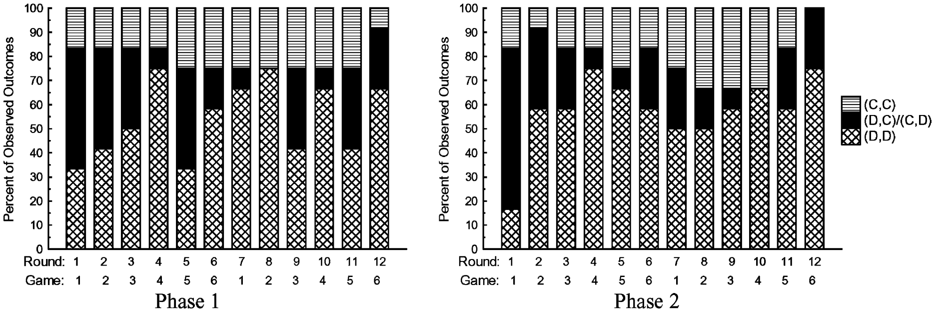


Fig. 5. Game outcomes across all decision rounds — Design III (12 observations per round).

Table 4
Logit analysis: Design III[a]

Variable	Phase 1 — fixed effects		Phase 2 — fixed effects	
	Coefficient	*P*-value	Coefficient	*P*-value
Constant	2.12	0.20	0.86	0.63
Fear	−0.07	0.01	−0.02	0.42
Greed	−0.04	0.08	−0.05	0.05
PrevRound-(D, C)	0.36	0.59	0.49	0.47
PrevRound-(C, D)	0.07	0.90	0.14	0.84
PrevRound-(C, C)	2.77	0.00	2.66	0.00
No. of observations	198[b]		165[b]	
Nagelkerke R^2	0.50		0.58	
$-\ln L$	81.8		67.1	

[a] Coefficients for fixed effects dummy variables for each subject are not reported.

[b] Four subjects in Phase 1 and nine subjects in Phase 2 played D in every round, so they were dropped from the dataset for purposes of this estimation because their maximum likelihood fixed effect is negative infinity. Furthermore, two subjects played C in every round of Phase 1, so they were also dropped from the dataset because their optimal fixed effects coefficients are known to be positive infinity.

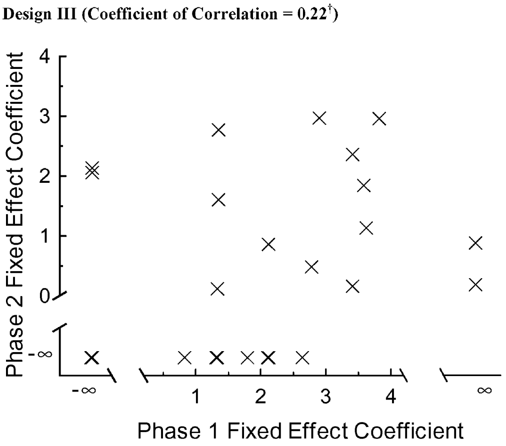
Although the patterns of effects of Fear and Greed are similar to those observed in Designs I and II, the impact in Design III appears to be smaller than in those designs.

Observation 4: Play in the next round is dependent on play in the previous round.

The interpretation of Observation 4 is straightforward. When subjects are re-matched each round with the same person, what matters most is the play of the pair in the previous round. From the results presented in Table 4, however, one sees that the dominant effect (in comparison to the benchmark where both subjects defect) is in regard to whether both subjects cooperated in the last round. This result points to the strong effect of reciprocity in a setting where subjects are repeatedly matched with the same person. This reciprocity effect also seems to overpower the “player type” effects seen in the previous designs. Fig. 6, which plots each subject’s Phase 1 and Phase 2 fixed effects coefficients, displays much less correlation than appeared in Fig. 4.[11] Just as in the two previous designs, several subjects never cooperated, so the best estimate of their fixed effects coefficient would be negative infinity. Also note that two subjects had fixed effects coefficients of positive infinity in Phase 1. This pair of subjects cooperated in every round. The maximum likelihood estimate of each of their intercept terms in a logistic model would be positive infinity since the dependent variable takes on a value of one for all values of the independent variables.

---

[11] As with Designs I and II, a regression was estimated using data pooled from the two phases of this design in which only one fixed effects coefficient was estimated per subject. Comparing the results of this regression to those of Table 4, we would reject the hypothesis that these fixed effects coefficients are the same across the two phases.



† The coefficient of correlation is an asymptotic value calculated by taking the limit as the infinite fixed effects coefficients become either very small or very large as appropriate.

Fig. 6. Correlation between Phase 1 fixed effects and Phase 2 fixed effects — Design III.

6. Conclusions

A growing debate in the empirical and theoretical literature relates to play in game situations in which players make strategy selections that are at odds with theoretical predictions based purely on selfish motives related to pecuniary payoffs in the games. The theoretical literature has seen an explosion of models designed to capture alternative explanations for play, based on normative concepts such as altruism, fairness, and income inequality aversion. Similarly, the experimental literature has seen many new studies designed to examine the impact of variations in game parameters, as well as other dimensions of the decision environment.

We introduced four basic elements related to the structure and linkages within any experimental game situation that we posit to affect behavior: pecuniary benefits, player types, information about player types, and linkages that occur between players in repeated game situations. The reported experimental evidence was drawn from decision environments aimed at examining aspects of these elements in the setting of a repeated PD game. Our results can be summarized as follows:

- In these decision settings, overall cooperation rates are quite low.
- Overall cooperation rates increase when subjects are matched with the same individual within Phase 1 and Phase 2 relative to situations where they are randomly rematched each decision round.
- Rates of cooperation increase when subjects face a decrease in Fear or Greed that is accompanied by an increase in Cooperators’ Gain.
- The evidence supports the notion of subject types, including one class of subjects who defect in every round and others whose propensity to cooperate or defect is responsive to game parameters and history of play.

These summary results point to the complicated nature of the linkages among the elements of social dilemma settings. Some subjects simply defected in every round, regardless of the pecuniary benefits or observed history. This behavior is most consistent with "PD-type" preferences. However, players with inclinations toward cooperation face a more complex strategy selection problem. Even if a player has "assurance-type" preferences over the pecuniary payoffs in a particular game, cooperation may not be their optimal strategy. They must also try to predict the likely behavior of the player with whom they are matched based on the pecuniary payoffs, their own experience of past play, and the specific structure of the repeated game including the matching procedure. It is perhaps not surprising that players cooperate most in Design III, where the structure of the interaction provides players with information about the previous plays of the player with whom they are matched, in addition to an enhanced ability to signal a willingness to cooperate. It is perhaps somewhat surprising that cooperation rates are so responsive to changes in pecuniary benefits when controlling for previous history of play, even in the fixed match environment of Design III. Does a player choose to defect because the pecuniary benefits make defection more attractive if the other person were to cooperate (Greed) or because cooperation is more costly if the other person defects (Fear)? Based on the near equality of the coefficients on Fear and Greed in each of the regressions reported in Tables 2–4, our evidence suggests that players, on average, consider both factors.

Acknowledgements

The authors thank the American Academy of Arts and Sciences and the National Science Foundation (Grant no. SBR-9319835) for funding support for this project. We would also like to thank the Russell Sage Foundation for its support for two conferences and the participants of those conferences for their useful comments on this paper. We appreciate the helpful comments of the editor and an anonymous referee. Robert Shupp and Pam Schmitt developed the software used here.

References

Ahn, T.K., Ostrom, E., Walker, J., 1999. Trust and reciprocity: experimental evidence from PD games. Working Paper, Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington.

Andreoni, J., Miller, J., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. The Economic Journal 103, 570–585.

Binmore, K.G., 1998. The evolution of fairness norms. Rationality and Society 10, 275–302.

Bohnet, I., Frey, B.S., 1998. The sound of silence in prisoner's dilemma and dictator games. Working Paper, Haas School of Business, University of California, Berkeley.

Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity reciprocity and competition. American Economic Review 90, 166–193.

Cain, M., 1998. An experimental investigation of motives and information in the prisoner's dilemma game. Advances in Group Processes 15, 133–160.

Camerer, C., 1990. Behavior game theory. In: Hogarth, R. (Ed.), Insights in Decision Making: A Tribute to Hillel J. Einhord. University of Chicago Press.

Clark, K., Sefton, M., 2001. The sequential prisoner's dilemma: evidence on reciprocation. Economic Journal 111, 51–68.

Coleman, J., 1988. Free riders and zealots: the role of social networks. Sociological Theory 6, 52–57.

Crawford, S.E.S., Ostrom, E., 1995. A grammar of institutions. American Political Science Review 89, 582–600.

Croson, R., 1999. Theories of altruism and reciprocity: evidence from linear public goods games. Working Paper, OPIM Department, The Wharton School, University of Pennsylvania.

Falk, A., Fischbacher, U., 1998. A theory of reciprocity. Working Paper, Institute for Empirical Research, University of Zurich, Switzerland.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition and cooperation. Quarterly Journal of Economics 114, 817–868.

Gunnthorsdottir, A., Houser, D., McCabe, K., Ameden, H., 2000. Excluding free-riders improves reciprocity and promotes the private provision of public goods. Working Paper, University of Arizona, Tucson.

Güth, W., Kliemt, H., 1998. The indirect evolutionary approach: bridging the gap between rationality and adaptation. Rationality and Society 10, 377–399.

Güth, W., Yaari, M., 1992. An evolutionary approach to explaining reciprocal behaviour in a simple strategic game. In: Kliemt, H. (Ed.), Explaining Process and Change. University of Michigan Press, Ann Arbor.

Hayashi, N., Ostrom, E., Walker, J., Yamagishi, T., 1999. Reciprocity, trust, and the sense of control: a cross-societal study. Rationality and Society 11, 27–46.

Hoffman, E., McCabe, K., Smith, V., 1996. Social distance and other regarding behavior in dictator games. American Economic Review 86, 653–660.

Hoffman, E., McCabe, K., Smith, V., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Economic Inquiry 36, 335–352.

Hsiao, C., 1996. Logit and probit models. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Kluwer Academic Publishers, Dordrecht.

Kahneman, D., Knetsch, J.L., Thaler, R.H., 1986. Fairness and the assumptions of economics. Journal of Business 59, S285–S300.

Kelley, H.H., Thibaut, J.W., 1978. Interpersonal Relations. Wiley, New York.

Kollock, P., 1998. Transforming social dilemmas: group identity and co-operation. In: Danielson, P.A. (Ed.), Modeling Rationality, Morality, and Evolution. Oxford University Press, Oxford, pp. 186–210.

Komorita, S.S., Sweeney, J., Kravitz, D.A., 1980. Cooperative choice in the *N*-person dilemma situation. Journal of Personality and Social Psychology 38 (3), 504–516.

Kramer, R.M., McClintock, C.G., Messick, D.M., 1986. Social values and cooperative response to a simulated resource conservation crisis. Journal of Personality 54, 576–592.

Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory 27, 245–252.

Kuhlman, D.M., Camac, C.R., Cunha, D.A., 1986. Individual differences in social orientation. In: Wilke, H., Messick, D.M., Rutte, C.G. (Eds.), Experimental Social Dilemmas. Peter Land, Frankfurt.

Ochs, J., Roth, A.E., 1989. An experimental study of sequential bargaining. American Economic Review 79, 355–384.

Ostrom, E., 1998. A behavioral approach to the rational choice theory of collective action. American Political Science Review 92, 1–22.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: how much and why? American Economic Review 87, 829–846.

Palfrey, T.R., Rosenthal, H., 1994. Repeated play, cooperation, and coordination: an experimental study. Review of Economic Studies 61, 545–565.

Rabin, M., 1993. Incorporating fairness in game theory and economics. American Economic Review 83, 1281–1302.

Rapoport, A., 1988. Experiments with *n*-person social traps I and II. Journal of Conflict Resolution 32, 457–488.

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

Raub, W., 1990. A general game-theoretic model of preference adaptations in problematic social situations. Rationality and Society 2, 67–93.

Saijo, T., Nakamura, H., 1995. The 'spite' dilemma in voluntary contributions mechanism experiments. Journal of Conflict Resolution 39, 535–560.

Schmidt, D., Shupp, R., Walker, J.M., Ostrom, E., 2000. Playing safe in coordination games: the role of risk dominance, payoff dominance, social history, and reputation. Working Paper, Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington.

Sen, A., 1974. Choice, orderings and morality. In: Körner, S. (Ed.), Practical Reason: Papers and Discussions. Blackwell, Oxford.

Wagner, T., 1998. Reciprocity and efficiency. Rationality and Society 10, 347–376.

Watabe, M., Terai, S., Hayashi, N., Yamagishi, T., 1996. Cooperation in the one-shot prisoner's dilemmas based on expectations of reciprocity. Japanese Journal of Experimental Social Psychology 36, 183–196 (in Japanese with an English summary).

Yamagishi, T., Cook, K.S., 1993. Generalized exchange and social dilemmas. Social Psychology Quarterly 56, 235–248.