ELSEVIER

Journal of Economic Behavior & Organization
Vol. 56 (2005) 61–76

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Compensating the cooperators: is sorting in the prisoner's dilemma possible?

Iris Bohnet [a,*], Dorothea Kübler [b]

[a] *Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138, USA*
[b] *Department of Economics and Management, Technical University Berlin, Strasse des 17. Juni 135, D-10623 Berlin, Germany*

Received 17 October 2001; accepted 23 April 2003

Abstract

Choice between different versions of a game may provide a means of sorting. We experimentally investigate whether auctioning off the right to play a prisoner's dilemma game in which the cost of unilateral cooperation is lower than in the status quo version separates (conditional) cooperators from money maximizers. After the auction, significantly more subjects cooperate in the modified PD than in the status quo PD, whereas there is no difference between cooperation rates if the two versions of the game were assigned to participants. However, sorting is incomplete and cooperation deteriorates over time.

JEL classification: C72

Keywords: Prisoner's dilemma game; Sorting; Conditional cooperation; Auctions; Experiments

1. Introduction

Experiments on prisoner's dilemma and other public goods games typically reveal cooperation rates higher than the equilibrium prediction in the first round and a decrease over time, leading to cooperation rates closer to the equilibrium prediction in the final round of the game. This pattern applies to repeated games as well as to repeated one-shot games.[1] In this paper, we investigate whether auctioning off the right to play a prisoner's dilemma game may stop this trend. The paper explores whether the choice between two versions of a

* Corresponding author. Tel.: +1-617-495-5605; fax: +1-617-496-5747.
E-mail addresses: iris_bohnet@harvard.edu (I. Bohnet), dkuebler@ww.tu-berlin.de (D. Kübler).

[1] See, for example, Andreoni (1988), Andreoni and Miller (1993), and for surveys, Davis and Holt (1993), Ledyard (1995) and Andreoni and Croson (2004).

0167-2681/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jebo.2003.04.002

prisoner's dilemma game with the same Nash equilibrium but different out-of-equilibrium payoffs provides a means of sorting, allowing players with different preferences to self-select into the version of the game they prefer.

The existing experimental findings on prisoner's dilemma and public goods games overwhelmingly suggest the existence of heterogeneous players, the two most important being conditional cooperators (i.e. subjects who cooperate if others cooperate as well) and egoists (i.e. money maximizers).[2] Fischbacher et al. (2001), for example, find that the majority of their subjects were conditionally cooperative (with a strong correlation between own and other contributions) in a public goods game.[3] However, despite the prevalence of conditional cooperators, high cooperation rates cannot be sustained over time. As conditional cooperators' behavior depends on what others do, the existence of non-cooperative types induces a downward spiral.[4]

We investigate whether an auction for the right to play a modified version of the game rather than the status quo version separates cooperators from defectors. In the modified game the cost of unilateral cooperation is lower, but the Nash equilibrium remains the same as in the status quo game. We choose this specific change in out-of-equilibrium payoffs, because not only theory but also earlier experimental evidence (e.g. Ahn et al., 2001) suggest that when games are assigned, such a change in payoffs does not affect behavior. In addition, it captures important aspects of real life mechanisms employed to sort employees, customers and insurees. As the cost of unilateral cooperation is lower in the modified version than in the status quo version, the change in the payoff structure represents an insurance mechanism by (partially) compensating the cooperator in case his or her counterpart defects. Therefore, the modified game is prima facie more attractive to players who want to cooperate than to money maximizers. Apart from sorting schemes used by insurance companies, for example, clubs often employ such mechanisms to induce self-selection: high membership rates together with certain privileges can deter some people and attract others who value these privileges highly enough. Similarly, buying a house in a good neighborhood is like bidding to become a member of this neighborhood since the supply of houses is (temporarily) fixed.

The experiment is designed as follows. We run two versions of a one-shot two-person prisoner's dilemma game with the same unique Nash equilibrium payoffs. Before playing the game, each subject individually decides which version of the game he or she wants to play. The right to participate in the "insured" instead of the status quo version of the game is sold in an *n*th-price auction. The subjects who win the auction play the game according to the modified payoff structure while for all others the status quo version remains valid.

We test for the effect of two central contextual variables, the number of rights available to play the modified PD and the number of periods played after an auction. The intuition for this is straightforward: if there are more rights available than conditional cooperators present,

[2] See, for example, Brandts and Schram (2001), Croson (1999), Keser and van Winden (2000), and for a recent survey, Fehr and Gächter (2000). Individual heterogeneity in preferences does not exclude the possibility of errors. Rather, studies testing the relevance of errors and of other-regarding preferences in public goods games find that both are present, see Anderson et al. (1998), Andreoni (1995) and Palfrey and Prisbrey (1997).

[3] Of the subjects, 48 percent were conditionally cooperative and 32 percent could be classified as purely selfish. The remaining 20 percent of the subjects displayed an unusual, not easily identifiable pattern of behavior.

[4] Fehr and Schmidt (1999) show theoretically that cooperation cannot be obtained in equilibrium even if the conditional cooperators are in the majority.

full separation of player types is impossible. Even if all cooperators opted for the modified game, some egoists would be able to take advantage of them, inducing the downward spiral. If there are fewer rights available than conditional cooperators present, on the other hand, sustainable sorting seems possible. We vary the number of periods played after an auction to test for different expected values of playing the insured version of the game. The more periods subjects can spend in the "safe(r) haven," the higher their bids should be.

This is a novel experimental design. In contrast to earlier related experimental studies, our design investigates the choice between two versions of a game instead of between playing and not playing a game. The latter choice situation has been extensively studied for bargaining and coordination games.[5] The exit experiment by Orbell and Dawes (1993) comes closest to our design. In their experiment, subjects could choose between playing a prisoner's dilemma game (where a player's profits were only positive if the other player cooperated, otherwise he or she made a loss) and exiting the game (with payoff zero). The authors report higher cooperation rates with an exit option than in the standard PD and argue that this supports sorting: egoists opt out as they underestimate the probability of cooperation while cooperators choose to play the PD.

In comparison, our design allows for a better test of the sorting hypothesis because we observe the behavior of those who lost the auction, whereas in the design of Orbell and Dawes those who exit have no choice to make. We find that in the first period of the auction treatments significantly more subjects cooperate in the insured game than in the status quo version. Paying for the right to participate in the insured game seems to provide an opportunity for self-selection. However, sorting is incomplete. In most sessions, there is some cooperation in the status quo game and some defection in the insured game. Thus, there are still many cooperating subjects who are "exploited". The experience of being the "sucker" induces the disappointed cooperators to stop cooperating.[6] This dynamic leads to a decrease in cooperation rates in the insured game over time.

The auction price does not correspond to the differences in expected values between the insured and the status quo game. While the laboratory environment seems comparatively simple, the bidding decision is cognitively quite demanding. In order to bid rationally, subjects would have to hold correct beliefs about the cooperation rates in both games. With two player types, conditional cooperators and money maximizers, this requires knowledge of the distribution of types in both games after the auction. However, we find that over time auction prices move towards the expected value of playing the modified game instead of the status quo game.

In the next part of the paper, we outline the experimental design and provide a simple analysis of the game. Section 3 reports the results and Section 4 relates them to the literature. Section 5 concludes the paper.

[5] In bargaining games, proposer competition was analyzed by Güth and Tietz (1986) and responder competition by Prasnikar and Roth (1992). Van Huyck et al. (1993) and Cachon and Camerer (1996) allow players in a coordination game with Pareto-ranked equilibria to opt out of the game. Cooper et al. (1993) investigate the effect of an outside option in the battle-of-the-sexes game. Related to these exit-experiments is a public goods game by Ehrhart and Keser (1999) in which subjects could form new groups.

[6] For early experimental evidence in psychology, see Brubaker (1975). For more recent economic experiments, Isaac et al. (1989).

Table 1
Payoff tables

	X	*Y*
Table A		
X	350; 350	0; 500
Y	500; 0	150; 150
Table B		
X	350; 350	100; 500
Y	500; 100	150; 150

2. The experiment

2.1. Design

Our design consists of a two-person, one-shot prisoner's dilemma game, which is employed in two versions. Table 1 presents Payoff table A, the status quo version, and Payoff table B, the insured version of the game. Numbers represent actual payoffs in cents.[7]

Four different treatments were conducted: The control treatment I where Games A and B were assigned to the participants as well as three different auction treatments. In the auction treatments, all subjects were assigned version A but could bid for the right to play version B of the game. We varied the number of the rights available to play game B as well the number of periods played after an auction. In Treatment conditions II and III, auctions were repeated in every period, in condition II with large B-groups (approximately two-thirds of the participants) and in condition III with small B-groups (approximately one-third of the participants). The group sizes were chosen such that in condition II, B would consist of more players than the number of cooperators we expected in the first period (66% of all participants play B), and in condition III, B would be small enough (33% play B) to consist of cooperators only in the first period.[8] In Treatment IV, also a small-B-group design (33% play B), an auction was only run in the first period, after which subjects remained in the respective games for Periods 2–5. In all treatments, the game was repeated five times, which was common knowledge. Subjects were randomly matched with a new counterpart in each period[9] and privately informed about their individual results after each period. An overview of the experimental design is presented in Table 2.

An nth-price sealed bid auction was used to elicit individuals' willingness to pay for game B. Auction sessions were conducted as follows: after the participants had read the

[7] The payoffs were presented to our experimental subjects in a matrix form; no normative frames were used. The experimental instructions can be found at: http://ksghome.harvard.edu/∼.ibohnet.academic.ksg/papers.html.

[8] In our Control treatment I, 36% of our subjects cooperated in the first round. Reviewing the literature revealed a striking consistency: typically, cooperation rates of about one third are found in the first rounds of anonymous, one-shot prisoner's dilemma games (see, e.g., Andreoni and Miller (1993), Bohnet and Frey (1999a,b), or Dawes et al. (1977).

[9] See Andreoni and Miller (1993) for this repeated one-shot design in two-person prisoner's dilemma games. Due to backward induction, our prediction from above follows even if we acknowledge that random matching does not produce a "true" one-shot environment.

Table 2
Experimental design

Treatment conditions	N	n(A)	n(B)
I. Control	48		
Assigned A		24	
Assigned B			24
II. Repeated auction, B large	72		
Auction 1	26	10	16
Auction 2	24	10	14
Auction 3	22	8	14
III. Repeated auction, B small	78		
Auction 4	30	20	10
Auction 5	18	12	6
Auction 6	30	20	10
IV. First-round auction, B small	54		
Auction 7	28	18	10
Auction 8	26	16	10

instructions and questions had been answered, we ran a practice auction where subjects were assigned their true private valuation for a hypothetical good beforehand.[10] We then started with the experiment: each participant had to indicate his or her bid; the auction price was determined and written on the blackboard. Subjects were allocated to Game A or Game B depending on their bid, randomly paired with another player in the same version of the game and informed about which game they were playing. Then each subject had to decide whether to cooperate or to defect; finally everybody learned about his or her earnings in the current period. In Treatment conditions II and III, this procedure was repeated five times; in Treatment condition IV, the auction was only conducted in Period 1 and players remained in either Game A or B for all five periods.

The experiments were conducted with students from various universities in the Boston area. Participants were paid a show-up fee of $5. The experiment was conducted anonymously and took approximately 45 min. After the experiment, participants were paid in cash and earned $15 on average (including the show-up fee). Subjects were identified by code numbers only and care was taken that neither other subjects nor the experimenter could observe private decisions.

2.2. Analysis of the game

Defection is the unique Nash equilibrium in both versions of the game, which only differ in the out-of-equilibrium payoffs for unilateral cooperation. Thus, if all subjects are rational

[10] Participants were asked to indicate their willingness to pay for one of three identical hypothetical goods on a piece of paper, which we collected. We wrote all the bids on the blackboard and demonstrated how an nth-price auction works: the three highest bidders would each win one of the identical hypothetical goods and pay the price that the fourth highest bidder bid. In case of a tie, a random device was used to determine who won the auction, and the price announced before the tie occurred had to be paid.

money maximizers, and if this is common knowledge, the equilibrium prediction is identical in both versions of the game: nobody cooperates. However, there is much empirical evidence for prisoner's dilemma and public goods games suggesting that at least one other type of player is present whose behavior depends on others. Such interdependence may be caused by intentions (Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 1998; Levine, 1998; Rabin, 1993), preferences (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), or behavior (Brandts and Schram, 2001) and is often labeled "conditional cooperation".

We use a simple but straightforward definition of conditional cooperation: as money maximizing rational individuals should not cooperate in one-shot prisoner's dilemma games, we consider all players (and only those) who cooperate in the first period to be conditional cooperators. While unconditional cooperators continue to cooperate in later periods independent of their first-period experience, conditional cooperators stop cooperating if their counterpart defected but cooperate again in later periods if their counterpart cooperated in the first period. Based on this definition of types, we expect the proportion of conditional cooperators to be constant in all treatment conditions.

The auction we employ in this experiment neither qualifies as a private value nor as a common value auction.[11] Our bidders do not know the value of the item to themselves with certainty as the value depends on other participants' choices in the game. Allowing for subject heterogeneity, the value of the item is also not the same to everyone: one's own preferences and expectations about other people's behavior determine bidding. Van Huyck et al., who used an auction to sell the right to play a game (eliciting individuals' willingness to pay for a coordination game), call it a "game form auction". They state that "the value of the object being auctioned is determined by the strategic interaction of the owners and this strategic interaction can depend on the price generated in the auction" (Van Huyck et al., 1993, p. 493).

If subjects bid rationally, their bids should reflect the expected value of playing version B rather than version A. In order to evaluate how much better it is to play version B rather than version A, a participant has to form beliefs about what strategy his or her counterpart will adopt, consider that sorting may occur, and choose a strategy. In particular, every subject has to compute the expected payoff from playing B minus the expected payoff from playing A given his or her strategy. If subjects do not assume that only subjects of their own type are present[12] and if egoists do not assume that the likelihood of being paired with a cooperator is the same in both versions of the game (in which case they bid zero cents), then both conditional cooperators and egoists are willing to pay a positive price to participate in version B rather than version A of the PD.

While bids most likely do not reflect the true expected value initially due to the complexity of the task, we expect them to move towards it as subjects learn and update their beliefs about other players' behavior over time. Subjects are informed about the auction price and

[11] Comparatively stable behavior is reported in private and common value second-price auctions (Kagel, 1995) although the experimental evidence does not fully support the theoretical predictions (Vickrey, 1961).

[12] If this is common knowledge, everybody bids 0 cents and egoists defect and cooperators cooperate in both games. This is an extreme version of the "false consensus effect" (Dawes, 1989).

their individual earnings after each period, so that both money maximizers and conditional cooperators can learn and adapt their behavior accordingly.

In small B-groups, complete sorting with only conditional cooperators playing game B is possible and full cooperation could be sustained over time. With fewer rights to play B than conditional cooperators present, the auction price should approach the expected payoff difference for cooperators. However, in large B-groups, sorting can never be complete. With more rights to play game B than conditional cooperators present, the auction price should be equal to the expected payoff difference between game A and game B for defectors. Conditional cooperators whose counterpart defects in the current period stop cooperating, increasing the number of defecting subjects in the next period and thus causing a downward spiral over time.[13] Thus, auction prices should be higher when group B is smaller and the more periods of version B are played after the auction.[14]

Apart from the sorting explanation, we want to explore an alternative hypothesis as well. Forward induction has been used to explain why auctions could lead to increased efficiency in coordination games.[15] If all our individuals were conditional cooperators, the prisoner's dilemma would be transformed into a coordination game. If the auction price was above 150 cents for game B, this would help such players to coordinate on the cooperative outcome in game B but not in game A. Differences in cooperation rates between versions A and B of the game could thus be accounted for.

3. Experimental results

In the following, the main findings are presented.

3.1. Observation 1 (types)

There is evidence for both types of players, egoists and conditional cooperators.

In our experiments, 86 out of 252 subjects cooperated in the first period of the game. One out of the 86 continued to cooperate in all the remaining four periods. Most other first-period cooperators' behavior is contingent on their counterpart's type: 75 percent of the cooperators meeting another cooperator in the first period ($N = 32$) are willing to cooperate at least once again in the future. On the other hand, only 24 percent of the cooperators meeting an egoist in the first period ($N = 54$), are ever willing to cooperate again in the remaining four periods.

[13] A similar spiral has been observed in continuous-choice public goods environments where conditional cooperators' contributions do not quite match the average contributions of others, see Fehr and Schmidt (1999).

[14] This prediction holds under the assumption that about one-third of the group are conditional cooperators, sorting takes place, and both are anticipated by the subjects.

[15] Van Huyck et al. (1993) and Cachon and Camerer (1996) allowed players of a coordination game with Pareto-ranked equilibria to opt out of the game. The price for the right to play the game served as an efficiency-enhancing coordination device if the price was high enough to exclude inefficient equilibria. While Van Huyck et al. argued that in line with forward induction, auction prices provide a means of tacit communication, Cachon and Camerer showed that 'better' equilibria were reached even if fees were imposed—which cannot be accounted for by forward induction. Cooper et al. (1993) and Offerman and Potters (2001) also find only limited support for forward induction.

Table 3
Share of players who cooperate at least once in Periods 2–5

	Treatments			
	C meets C in Period 1	C meets D in Period 1	D meets C in Period 1	D meets D in Period 1
I. Assigned games	75% ($N = 8$)	33% ($N = 9$)	22% ($N = 9$)	41% ($N = 22$)
II. Mult. large B-auctions	63% ($N = 8$)	43% ($N = 14$)	29% ($N = 14$)	25% ($N = 36$)
III. Mult. small B-auctions	70% ($N = 10$)	17% ($N = 18$)	28% ($N = 18$)	41% ($N = 32$)
IV. Single small B-auction	100% ($N = 6$)	8% ($N = 13$)	15% ($N = 13$)	5% ($N = 22$)

Of the 112 egoists who meet another defector in the first period, 29 percent cooperate at least once in Periods 2–5 (and 24 percent of the egoists meeting a cooperator in the first period ($N = 54$) are willing to cooperate in later rounds). First-period cooperators thus do not behave differently in later periods than first-period defectors if their counterpart defects but are much more likely to cooperate again if their counterpart also cooperates (Fisher's Exact test, $p < 0.01$). Table 3 presents the results for the four different treatments.

3.2. Observation 2 (cooperation rates)

The proportion of cooperators in the first period is similar in all our treatments. Cooperation rates are higher in version B than in version A of the game in the auction treatments, but not in the control treatment. Cooperation rates in all treatments and in all games decrease over time.

In the first period, there are no significant differences between the overall-cooperation rates in the control treatment and the eight auction sessions (Fisher's Exact tests, $p > 0.1$).[16] This is consistent with our simple definition of types. At the same time, substantial differences between first-period cooperation rates in versions A and B can be observed in the auction treatments. For a graphic representation of cooperation rates in Games A and B see Fig. 1 below. (Table 4 reports the cooperation rates for each session.) In all auction treatments, but not in the control treatment, cooperation rates in versions B are significantly higher than in versions A. A Fisher's Exact test yields $p = 0.547$ for Treatment I, $p = 0.000$ for Treatment II, $p = 0.025$ for Treatment III, and $p = 0.046$ for Treatment IV.[17] While the first-period results suggest that most subjects sort themselves into version A and B according to their type, in none of the auction sessions sorting is complete. While

[16] Comparing the overall first-period cooperation rate of each auction session with the cooperation rate in the assigned games, no significant differences can be found (using a Fisher's Exact test): Auct-1: $p = 1.00$, Auct-2: $p = 0.431$, Auct-3: $p = 1.000$, Auct-4: $p = 0.805$, Auct-5: $p = 0.167$, Auct-6: $p = 0.805$, Auct-7: $p = 0.626$, Auct-8: $p = 0.799$.

[17] At the level of individual auction sessions, the difference between versions A and version B is significant at a 10%-level in the three sessions of Treatment II with large B and in one of the sessions of Treatment IV: Auct-1(A–B): $p = 0.087$, Auct-2(A–B): $p = 0.024$, Auct-3(A–B): $p = 0.022$, Auct-4(A–B): $p = 0.115$, Auct-5(A–B): $p = 0.638$, Auct-6(A–B): $p = 0.115$, Auct-7(A–B): $p = 0.050$, Auct-8(A–B): $p = 0.664$ (Fisher's Exact test). The data are pooled as there are no significant differences between the sessions in each treatment.

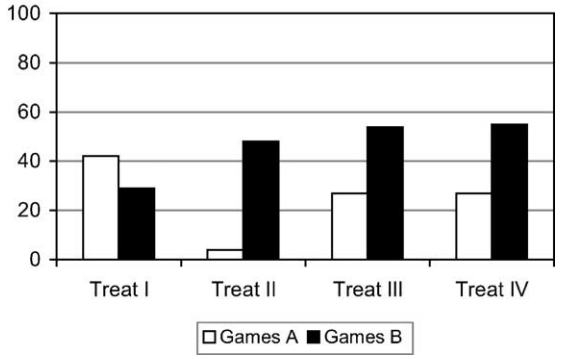

Fig. 1. First-period cooperation rates in Games A and B (%).

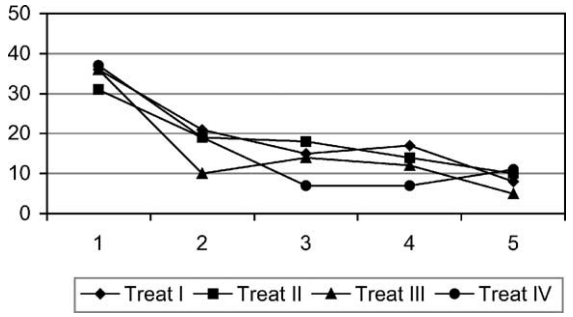

Fig. 2. Overall-cooperation rates over time (%).

first-period cooperation rates in large B-groups are about 50% in all auction sessions, they do not approach 100% in any of the small B-groups.

The opportunity to self-select into groups of similar types in Period 1 does not prevent conditional cooperators from defecting in later rounds. Fig. 2 presents this trend graphically. The decrease in cooperation rates in auction Treatments II and III is very similar to the decay in the assigned Treatment I. The differences between the overall-cooperation rates in the control Treatment I and the two Auction treatments II and III are not significant in almost all periods.[18] Finally, no difference between Treatments II and III can be observed (χ^2-tests: $p > 0.1$). Thus, incomplete sorting in both large and small B-groups induces a downward spiral.

Moreover, the decrease in cooperation is similarly strong in Treatment IV with a single auction and in Treatments II and III with an auction in each period. The differences between the overall cooperation rates in the three auction treatments are not significant in almost all periods.[19] As the decrease in cooperation in Treatment IV cannot be attributed

[18] χ^2-tests: $p > 0.1$ for all treatment comparisons in all periods, with the exception of Treatments I and III in Period 2 where we find a marginally significant difference with $p = 0.099$.

[19] χ^2-tests: $p > 0.1$ for all treatment comparisons in all periods, with the exception of Treatments II and IV in Period 3 where we find a marginally significant difference with $p = 0.083$.

Table 4
Cooperation rates in all periods

	Periods				
	1	2	3	4	5
Assig. I (%)					
A + B	36	27	15	17	8
A	42	21	21	13	8
B	29	33	8	21	8
Auct. II-1					
A + B	35	31	23	27	15
A	10	0	0	10	0
B	50	50	38	38	25
Auct. II-2					
A + B	25	12	12	4	8
A	0	10	0	0	10
B	43	14	21	7	7
Auct. II-3					
A + B	32	13	18	9	8
A	0	0	13	0	0
B	50	21	21	14	13
Auct. III-4					
A + B	30	13	23	10	3
A	20	5	10	5	0
B	50	30	50	20	10
Auct. III-5					
A + B	56	11	11	17	11
A	50	17	8	17	0
B	67	0	17	17	33
Auct. III-6					
A + B	30	7	7	10	2
A	20	0	5	5	0
B	50	20	10	20	5
Auct. IV-7					
A + B	43	18	11	7	14
A	28	6	6	6	11
B	70	40	20	10	20
Auct. IV-8					
A + B	31	19	4	8	8
A	25	19	6	0	0
B	40	20	0	20	20

to invaders who play B in later rounds in order to exploit the cooperators, it seems unlikely that this is the case in Treatments II and III. Rather, incomplete sorting, which induces conditional cooperators to defect in later rounds, provides a unified explanation for all three treatments.

Table 5
Probability of moving between games in Periods 2–5 in Treatments II and III

Move in Period t	C meets C in Period t-1	C meets D in Period t-1	D meets C in Period t-1	D meets D in Period t-1
A → B	0% ($N = 2$)	40% ($N = 15$)	12% ($N = 17$)	23% (237)
B → A	7% ($N = 28$)	26% ($N = 53$)	25% ($N = 47$)	25% ($N = 128$)

3.3. Observation 3 (moving between games)

Subjects are most likely to move from Game A to Game B after having unilaterally cooperated in A in the previous period. Subjects are least likely to move from Game B to Game A after having mutually cooperated in B in the previous period.

We focus on Treatments II and III to explore switching behavior. Table 5 reports the likelihood of switching from one game to the other based on previous period history. The data suggests that "previous-period suckers" (cooperating subjects who were paired with a defecting subject in the previous period) are most likely to switch from A to B and "previous-period mutual cooperators" are least likely to switch from B to A. A regression on bidding behavior controlling for round and treatment effects confirms these findings.[20] Subjects who cooperated in the previous period but were exploited bid more than anyone else who had played A and thus are most likely to move to Game B. Subjects who cooperated in the previous period and met another cooperating subject bid more than anyone else who had played B and thus are most likely to remain in B. Bids decrease over time and are higher in Treatment III than in Treatment II because groups B are smaller in the former. The pattern suggests that cooperating subjects try to sort themselves into Game B. However, they are unable to achieve complete sorting in both Treatments II and III.

3.4. Observation 4 (auction prices)

In the first period, auction prices are higher than if no sorting had been anticipated, but they do not correspond to the expected value of playing Game B minus the expected value of playing Game A. Auction prices move towards the difference in expected values between Games B and A over time.

If no sorting were expected, egoists should have bid 0 and conditional cooperators maximally 66 cents in the first period of both Auction treatments II and III (assuming that they expect about one-third of the group to cooperate). If subjects had rationally anticipated the distribution of types, first-period auction prices should be just equal to the expected advantage of playing B instead of A. Computing the expected value of B over A for defecting subjects by using the realized cooperation rates and subtracting the auction price for each session yields the graphs of Fig. 3.[21] It shows the difference between the expected value

[20] Regression results and additional analyses of the data can be found at: http://ksghome.harvard.edu/~.ibohnet.academic.ksg/papers.html.

[21] We compute the expected value of playing B instead of A for players who defect in both games. For example, in the first period of Auct-1 the expected value of playing B instead of A for a defector is $(0.5 \times 500 + 0.5 \times$

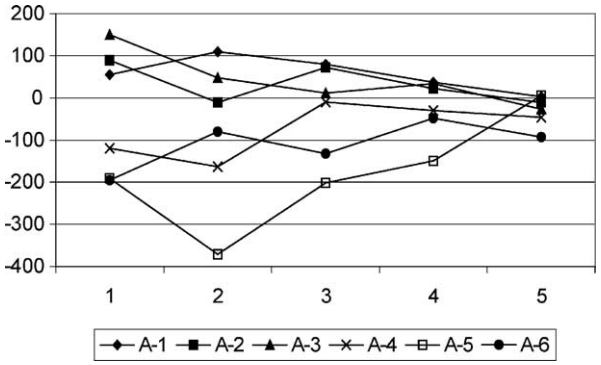

Fig. 3. Expected value of playing B instead of A minus auction price (in cents).

Table 6
Auction prices in all periods (in cents)

	Periods				
	1	2	3	4	5
Auction II-1	85	66	68	76	84
Auction II-2	62	25	2	2	0
Auction II-3	25	26	20	15	15
Auction III-4	225	250	150	82	81
Auction III-5	250	275	233	150	110
Auction III-6	300	150	150	100	110
Auction IV-7	450	–	–	–	–
Auction IV-8	350	–	–	–	–

of playing Game B and Game A for defectors minus the auction price in each period (i.e. $[\mathrm{EV(B)} - \mathrm{EV(A)}]_{\mathrm{Def}} - P$). In the first period, the auction price is too low in Treatment II, whereas it is too high in Treatment III. Table 6 reports the realized auction prices in all periods and all sessions. Over the five periods, as cooperation rates decrease, subjects adjust their bids to the decreasing true value of Game B.

Finally, the data do not support the forward induction explanation of higher cooperation rates in Game B than in A. First, the overall-cooperation rate is insensitive to the size of group B. Forward induction would imply that with large B, more subjects cooperate than when B is small. Second, auction prices in all sessions of Treatment II are too low (below 150 cents) for a forward induction argument to work. Although first-period auction prices in Treatment III were sufficiently high to induce cooperative behavior in Game B through forward induction, they did not result in systematically different behavior than the low auction prices in Treatment II.

$150) - (0.1 \times 500 + 0.9 \times 150) = 140$. Note that the expected values of defectors used in Fig. 3 are lower than the expected values of conditional cooperators, but our conclusions hold for cooperators a well.

4. Discussion of the literature

While it has not yet been investigated how changes in the payoff structure affect sorting in the prisoner's dilemma, the few papers studying the impact of relative payoffs on cooperation rates in assigned prisoner's dilemma games support our results in the control treatment. Introducing an insurance mechanism in a two-person PD that decreased the cost of unilateral cooperation from 40 to 10 cents, Ahn et al. (2001) did not find any effect on cooperation. Only when both the benefit of unilateral defection and the cost of unilateral cooperation were decreased from 40 to 10 cents, did they find an increase in cooperation.[22]

Schotter (1998) tested the effect of an insurance mechanism in a profit sharing game with two equilibria, a high effort and a low effort equilibrium. Subjects' payoffs depended on their own effort levels, those of the other group members and on the group incentive formula. Two such formulas were compared: a high-vulnerability plan A and a low-vulnerability plan B in which a subject's payoff fell less steeply than under plan A for identical reductions in others' efforts. Thus, plan B compensated high effort workers to some degree, providing an insurance mechanism similar to that employed in our game B. The author found that subjects' behavior in plans A and B did not differ (unless they shared a common history of shirking).[23]

Similarly, in our experiment the provision of an insurance mechanism by itself does not lead to higher cooperation rates. Only when the right to play the insured version of the game is auctioned off is more cooperation observed in the insured version of the game compared to the status quo version. This may not come as a surprise: privileges in clubs are not just randomly given to people but sold and sometimes even auctioned off.

Successful sorting mechanisms, however, are rare. In Ehrhart and Keser's public goods experiment where subjects could leave their groups to form new ones, "cooperating subjects are on the run from less cooperative ones who follow them around" (p. 9). As cooperative subjects formed new groups, hoping to meet other cooperators, egoists constantly invaded and decreased cooperation rates over time. Using a similar approach, Page et al. (2002) found that such endogenous group formation is more successful if subjects can choose other group members based on their past contributions, significantly increasing contributions and efficiency. The decrease in contributions over time can be stopped if such endogenous group formation is combined with the ability to punish free-riders.

A similar effect of sorting based on the knowledge of others' types has been reported in a bargaining game (Charness, 2000). Subjects were sorted by the experimenter according to their offers in a first-stage dictator game. In the second stage, people of the same 'type' were paired to play a bargaining game and informed of each other's type. Pairs of generous types

[22] Their average cooperation rate over all random matching treatments is also very close to ours, namely 31.6 percent. A change in payoff structures was first studied by Rapoport (1967) who called the relative cost of unilateral cooperation "fear" (payoff for mutual defection minus the payoff for unilateral cooperation) and the relative benefit of unilateral defection "greed" (payoff for unilateral defection minus payoff for mutual cooperation).

[23] While subjects could not choose between plans but were assigned to them, the author affects expectations about others' likelihood of choosing high effort levels using a different mechanism: prior to playing the profit sharing game, subjects either participated in the minimum or the median coordination game. While subjects typically managed to coordinate on the payoff-dominant equilibrium in the median game (no shirking experience), they typically decreased to the minimum in the minimum game (shirking experience).

bargained more efficiently than all other pairs. Finally, Offerman and Potters experimentally examined how auctioning off entry licenses (e.g. for oil drilling or airport slots) affects pricing behavior, also finding support for first-round sorting only. The most collusive players, setting the highest prices and earning the largest profits, self-selected into the market game in the first auction. However, in later auctions, no signs of such a selection effect were found. Sorting, thus, was not sustainable.

5. Conclusions

We have run an experiment that differs from past prisoner's dilemma studies in that our experimental subjects could choose between two PD payoff structures, the original Game A and a modified version B in which the payoff from unilateral cooperation is increased. The right to participate in version B rather than in the original Game A could be bought in an *n*th-price, sealed-bid auction. The specific design was chosen for two reasons: first, we wanted to provide subjects with an "institutional choice" implying that they could decide which version of a game they wanted to play but could not just opt out. Taking the criticism seriously that game theory is of limited practical relevance because it does not allow for individuals to change the games they play (Brandenburger and Nalebuff, 1996), we tested for the implications of transforming payoff structures in the laboratory. Secondly, the specific change in the payoff structure, the decrease in the cost of unilateral cooperation in Game B, was chosen in order to reflect schemes used by various organizations to sort their employees, customers or insurees.

We compare various auction treatments with a control treatment in which versions A and B were assigned to the subjects. After the auction, significantly more subjects cooperate in the modified PD than in the status quo PD, whereas there is no difference between cooperation rates if the two versions of the game were assigned to participants. Individuals willing to cooperate in Period 1 thus self-select into the insured version B while defectors bid less and play version A. Such a segmentation of types may be especially interesting from an evolutionary point of view as it facilitates the "proliferation of nice traits" (Bowles, 1998, p. 93).

First-period cooperators continue to cooperate if they have been paired with another cooperator in the first period. If their counterpart defected in the first period, the likelihood of ever cooperating again is as low for first-period cooperators as for first-period defectors, indicating that our subjects are either conditional cooperators or egoists. If sorting in the first period is incomplete, we should expect the dynamics of conditional cooperation to lead to a decrease of cooperation rates over time. This is what we observe: first-round cooperators who meet a defector in Period 1 stop cooperating in later rounds. The decrease of cooperation rates makes the differences in expected values between playing version B and version A smaller and smaller. Auction prices reflect this trend. While we observe over- and under-bidding in the beginning, the differences between expected values and auction prices are close to zero in the last period.

We find that auctioning off the right to play a prisoner's dilemma game provides a means of sorting in the beginning of the experiment but that sorting is not sustainable over time. More research remains to be done in order to understand better under which conditions

sorting can be stabilized and cooperation among those who are more inclined towards cooperative outcomes can be maintained. Our experiments shed light on some aspects of such a mechanism.

Acknowledgements

We thank Colin Camerer, Gary Charness, Bruno S. Frey, Simon Gächter, George Loewenstein, Felix Oberholzer-Gee, Dani Rodrik, Georg Weizsäcker, the participants of seminars at Harvard, Wharton, the University of Bern, the University of Bielefeld, Viadrina University Frankfurt/Oder, the University of Bergen, Humboldt University Berlin, and of the Economic Science Association Meetings and ESEM for their helpful comments. We gratefully acknowledge the financial support of the Swiss National Science Foundation, the Kennedy School of Government Dean's Research Fund, and the hospitality of the Haas School of Business and Boalt Hall, University of California at Berkeley.

References

Ahn, T.-K., Ostrom, E., Schmidt, D., Shupp, R., Walker, J., 2001. Cooperation in PD games: fear, greed and history of play. Public Choice 106, 137–155.

Andreoni, J., 1988. Why free-ride? strategies and learning in public goods experiments. Journal of Public Economics 37, 291–304.

Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion. American Economic Review 85, 891–904.

Andreoni, J., Croson, R., 2004. Partners versus strangers: Random rematching in public goods experiments. In: Smith, V., Plott, C. (Eds.). Handbook of Experimental Economic Results. Forthcoming.

Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated Prisoner's dilemma: experimental evidence. The Economic Journal 103, 507–585.

Anderson, S.P., Goeree, J.K., Holt, C.A., 1998. A theoretical analysis of altruism and decision error in public goods games. Journal of Public Economics 70, 297–323.

Bohnet, I., Frey, B.S., 1999a. The sound of silence in prisoner's dilemma and dictator games. Journal of Economic Behavior and Organization 38, 43–58.

Bohnet, I., Frey, B.S., 1999b. Social distance and other-regarding behavior in dictator games: comment. American Economic Review 89, 335–339.

Bolton, G., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. American Economic Review 9, 166–193.

Bowles, S., 1998. Endogenous preferences: the cultural consequences of markets and other economic institutions. Journal of Economic Literature 36, 75–111.

Brandenburger, A.M., Nalebuff, B.J., 1996. Co-opetition. Doubleday, New York.

Brandts, J., Schram, A., 2001. Cooperation and noise in public goods experiments: applying the contribution function approach. Journal of Public Economics 79, 399–427.

Brubaker, E.R., 1975. Free ride, free revelation or golden rule. Journal of Law and Economics 18, 147–161.

Cachon, G., Camerer, C., 1996. Loss-avoidance and forward induction in experimental coordination games. Quarterly Journal of Economics 111, 165–194.

Charness, G., 2000. Bargaining efficiency and screening: an experimental investigation. Journal of Economic Behavior and Organization 42, 285–304.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics 117, 817–869.

Cooper, R., DeJong, D.V., Forsythe, B., Ross, T.W., 1993. Forward induction in the battle-of-the-sexes game. American Economic Review 83, 1303–1316.

Croson, R., 1999. Theories of altruism and reciprocity: evidence from linear public good games. Working Paper, University of Pennsylvania, Wharton.
Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press, Princeton.
Dawes, R.M., 1989. Statistical criteria for establishing a truly false consensus effect. Journal of Experimental Social Psychology 25, 1–17.
Dawes, R.M., McTavish, J., Shaklee, H., 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. Journal of Personality and Social Psychology 35, 1–11.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 4, 268–298.
Ehrhart, K.-M., Keser, C., 1999. Mobility and cooperation: on the run. Working Paper, CIRANO, University of Montreal.
Falk, A., Fischbacher, U., 1998. A theory of reciprocity. Working Paper No. 6, University of Zurich, Zurich.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114, 817–868.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. Journal of Economic Perspectives 14 (3), 159–181.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71, 397–404.
Güth, W., Tietz, R., 1986. Auctioning ultimatum bargaining positions. In: Scholz, R.W. (Ed.), Current Issues in West German Decision Research. Peter Lang, Frankfurt, pp. 173–185.
Isaac, R.M., Schmidt, D., Walker, J.M., 1989. The assurance problem in a laboratory market. Public Choice 62, 217–236.
Kagel, J.H., 1995. Auctions: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.). Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 501–586.
Keser, C., van Winden, F., 2000. Conditional cooperation and voluntary contributions to public goods. Scandinavian Journal of Economics 102, 23–39.
Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–194.
Levine, D., 1998. Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1, 593–622.
Offerman, T., Potters, J., 2001. Does auctioning of entry licenses affect induce collusion? an experimental study. Working Paper, University of Amsterdam and Tilburg University.
Orbell, J.M., Dawes, R.M., 1993. Social welfare, cooperators' advantage, and the option of not playing the game. American Sociological Review 58, 787–800.
Page, T., Putterman, L., Unel, B., 2002. Voluntary association in public goods experiments: reciprocity, mimicry and efficiency. Working Paper, Brown University.
Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: how much and why. American Economic Review 87, 829–846.
Prasnikar, V., Roth, A.E., 1992. Considerations of fairness and strategy: experimental data from sequential games. Quarterly Journal of Economics 107, 865–888.
Rabin, M., 1993. Incorporating fairness into game theory and economics. American Economic Review 83, 1281–1302.
Rapoport, A., 1967. A note on the index of cooperation for the prisoner's dilemma. Journal of Conflict Resolution 11, 101–103.
Schotter, A., 1998. Worker trust, system vulnerability, and the performance of work groups. In: Ben-Ner, A., Putterman, L. (Eds.). Economics, Values, and Organization. Cambridge University Press, Cambridge, pp. 364–407.
Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1993. Asset markets as an equilibrium selection mechanism: coordination failure, game form auctions, and tacit communication. Games and Economic Behavior 5, 485–504.
Vickrey, W., 1961. Counterspeculation and competitive sealed tenders. Journal of Finance 16, 8–37.