Should I organize the conference?

Cut-point belief reciprocity in an experimental public goods game with alternating, single decision makers*

by

Jens Großer[i,ii]

Abstract:

We study a public goods experiment where just one decision maker of a group ('organizer') has to decide whether or not to provide the good. The others in the group ('visitors') do not make a decision. We find that organizers update their beliefs about the next organizer providing the public good based on their own decisions in accordance with learning direction theory. This calls for a more sophisticated learning model than those of adaptive learning. Moreover, our results show that the additional information of reported beliefs about others providing the public good can explain decisions more accurate. We show that a reciprocal decision rule involving a cut-point belief of 50% performs astonishingly well in all treatments.

Keywords: sequential public goods games, voluntary contribution mechanism, experimental economics, belief learning, reciprocity

* My special thanks go to Ronald Bosman, Arthur Schram, and Ingrid Seinen.

[i] Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands, e-mail: j.w.grosser@uva.nl

[ii] Department of Economics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany

1. Introduction

Many public goods are provided voluntarily. This has traditionally surprised economists, since there is usually a strong incentive to free ride on others' contributions. Why then, do people provide voluntarily? One explanation discussed in the experimental literature is that people are simply prone to mistakes (*e.g.*, Palfrey and Prisbrey, 1996, 1997; Goeree and Holt, 2002). A second explanation can be found in peoples' motivations.[1] It has been shown that a substantial number of people are not only concerned about their own well-being but also about the well-being of others (*e.g.*, Kelley and Stahelski, 1970). Social preferences could explain voluntary contributions (*e.g.*, Offerman *et al.*, 1996). Several studies on voluntary contributions use the concepts of conditional cooperation (*e.g.*, Offerman *et al.*, 1996) and reciprocity (*e.g.*, Fehr and Gächter, 1998). Roughly speaking, in both cases people respond friendly to decisions that are perceived as friendly, and hostile to those that are perceived as hostile.

A third explanation is the institutional design of public goods situations. *E.g.*, in a repeated setting, people might want to build a reputation of being cooperative in order to ensure cooperative gains for the future (*e.g.*, Keser and van Winden, 2000; Gächter and Falk, 2002). Moreover, in the presence of monitoring systems people might fear social punishment for non-contribution (*e.g.*, Fehr and Gächter, 2000a, 2002; Sefton *et al.*, 2002). This paper reports results on an experimental public goods game with a new institutional design, in which rather than a whole group just one individual is responsible for provision.

A frequently used public goods setting in laboratory research is the voluntary contributions mechanism (VCM) (see Ledyard, 1995, for an overview). The VCM has almost exclusively maintained a simultaneous-move protocol. All subjects of a group decide at the same point in time, not knowing how many of the other group members will contribute to the public good. However, simultaneous-move is not a necessary condition for pure public goods situations, which only demand non-rivalry and non-excludability. Here, we give up the assumption of simultaneous-move and present a VCM with only a single decision maker of a group having to decide on providing the public good. We call the game underlying such a VCM the 'organizer's game'. Across rounds, players alternate in being responsible for the

[1] For a discussion of individual motivations in experimental public goods games see, *e.g.*, Schram (2000). There is an abundance of literature discussing the mistakes-versus-motivations explanations of voluntary contributions, *e.g.*, Anderson *et al*. (1998), Andreoni (1995), and Brandts and Schram (2001).

provision of the public good ('sequential-move protocol'). The player responsible is called the organizer. All other players, the visitors, will benefit in case of provision but do not face the decision whether or not to provide themselves at that point in time. Organizing a conference and preparing coffee for the department are two examples that can be represented by the organizer's game.

Only few studies report sequential-move public goods experiments.[2] Erev and Rapoport (1990) focus on subjects' coordination on binary contributions in a step-level VCM. All their treatments show more efficient outcomes as compared to experimental step-level public goods games with simultaneous-move protocol. Bardsley (2000) finds more (non-binary) contributions with simultaneous- than with sequential-move. Weimann's (1994) simultaneous-move design catches some aspects that are also present under a sequential-move protocol: subjects receive information about individual (non-binary) contributions of other subjects in the previous round. He finds the same results with and without information about individual behavior.

The focus of our paper is the experimental investigation of i) belief learning, ii) decision-based reciprocity, and iii) belief-based reciprocity in the organizer's game. Moreover, we iv) study decision rules based on cut-point beliefs about others providing the public good. Our main focus is on the links between i) to iv).

i) In the standard VCM, a rational, own payoff maximizing player will make his decisions independent of beliefs. Not providing is a dominant strategy. But experimental results show positive beliefs about others contributing and strong correlation between decisions and beliefs even in the presence of dominant strategies (*e.g.*, Bosman and van Winden, 2002; Großer and Ule, 2002). Subjects who are more cooperative estimate the probability that others will contribute to be higher than subjects who are less cooperative (*e.g.*, Offerman *et al.*, 1996). This can be explained *e.g.* by the false consensus bias (*e.g.*, Ross *et al.*, 1977).[3]

Repetition of encounters allows for belief learning. Here, we elicit subjects' beliefs using an incentive compatible quadratic scoring rule (*e.g.*, Offerman *et al.*, 1996; Nyarko and

[2] The sequential prisoner's dilemma has been experimentally investigated *e.g.* by Bolle and Ockenfels (1990), Clark and Sefton (2001), and Solà (2000).

[3] See Goeree and Großer (2003) and Großer and Kugler (2002) for examples of how the false consensus bias can be incorporated in a rational choice framework.

Schotter, 2002). We use learning direction theory (Selten and Stoecker, 1986) to analyze subjects' belief learning. Furthermore, we test if belief learning of subjects in the organizer's position differs from that of visitors. Due to the sequential-move protocol organizers' and visitors' belief changes can be cleanly isolated. This is of particular interest for the study of the learning process. Models of adaptive learning like fictitious play (Brown, 1951) or naïve Bayesian learning (Eichberger *et al.*, 1993) only allow for learning of others' strategies, not for learning of the effect own decisions have on others' strategies. If belief learning of organizers is observed, a more sophisticated learning model is needed.

ii) Reciprocity has been used to explain decision making in various experimental games (good overviews are given in Fehr and Gächter, 1998, 2000b). However, its exact role still remains a puzzle. Definitions of reciprocity can be based on decisions of others as well as on beliefs about others' decisions.[4] Since we have data for both, we investigate decision- and belief-based reciprocity. Basically, decision-based reciprocity occurs when decisions to reward (punish) are based on observed friendly (hostile) choices by others.[5] In general, there is a problem of applying the notion of reciprocity to experimental public goods games due to externalities of decisions. A subject unavoidably punishes or rewards all members of a group, even if she prefers a more individualistic approach.[6] Nevertheless, we will test our data for the described basic notion of reciprocity.

iii) Next, we will include beliefs in the definition of reciprocity and investigate if belief-based reciprocity occurs. The definition of belief-based reciprocity we will use is a straightforward modification of that of decision-based reciprocity: we simply replace observed by expected decisions. Hence, according to our definition of belief-based reciprocity, a decision maker's choices are related to her beliefs about others' decisions. We will test whether subjects who provide the public good estimate the probability of others contributing to be higher than subjects who do not provide do.

[4] Or preferences (Rabin, 1993) and intentions of others (Falk and Fischbacher, 1998).

[5] Stronger definitions for reciprocity have been proposed, such as 'measure for measure' (Selten, *et al*., 1997). In this definition the amount of punishment or reward depends on the exact size of others' decisions. The organizer's game is not suitable to detect measure for measure due to its binary choices.

[6] For individualistic accounts of reward and punishment in public goods experiments see, *e.g*., Fehr and Gächter (2000a, 2002) and Sefton *et al*. (2002).

iv) Finally, we will elaborate on our findings about beliefs and reciprocity, and investigate whether subjects follow reciprocal decision rules, based on cut-point beliefs about other subjects providing the public good. The cut-point belief hypothesis states that a subject provides the public good if her belief is strictly higher than her cut-point and does not provide if it is strictly lower. In case her belief is equal to her cut-point, she makes either decision.

The paper is organized as follows. The organizer's game and its equilibria are presented in Section 2. Section 3 describes the experimental design. General experimental results are presented in Section 4, results on belief learning in Section 5, and results on reciprocity and decisions rules in Section 6. Section 7 concludes.

2. The organizer's game

2.1 One-shot organizer's game

There is a finite set of N players ($N > 1$). One player i, the 'organizer', is chosen exogeneously. The other $N-1$ players j ($j \neq i$) are 'visitors'. The organizer has to decide whether or not to provide a public good, *e.g.* organize a conference. Costs are given and equal to 0 if the public good is not provided, and to $c > 0$ if it is provided. Costs are private and can only be paid by the organizer. Visitors do not make a decision. If the good is not provided, revenues are zero: $r(0) = r_i(0) = r_j(0) = 0, \forall j$. If the public good is provided at cost c to i, revenues are symmetric: $r(c) = r_i(c) = r_j(c) > 0, \forall j$. In case the public good is not provided, all payoffs are equal to zero. The organizer's payoff $p_i(.)$ in case of provision of the public good is

$$p_i(c) = r(c) - c\,, \tag{1}$$

and that of each visitor is

$$p_j(c) = r(c)\,. \tag{2}$$

Total payoff is given by

$$R(c) = N \cdot r - c\,. \tag{3}$$

The Nash equilibrium with own payoff maximizing players is straightforward. The organizer has a dominant strategy to provide if $c < r(c)$, and a dominant strategy not to provide if $c > r(c)$. In case of $c = r(c)$, every strategy, mixed or pure, is a Nash equilibrium.

2.2 Finitely repeated organizer's game

The game theoretical solution for the finitely repeated organizer's game is also straightforward. Using backwards induction, the unique subgame-perfect Nash equilibrium is that in all rounds the organizer provides the public good if $c < r(c)$, and does not provide it if $c > r(c)$. In case of $c = r(c)$, in all rounds every mixed and pure strategies of organizers can be part of a subgame-perfect Nash equilibrium. Note that these predictions hold for every possible combination of organizers over the rounds and irrespective of the way in which roles are determined.

2.3 Efficiency

If $N \cdot r(c) > c$, provision of the public good is the efficient solution, and if $N \cdot r(c) < c$, non-provision is efficient. If $N \cdot r(c) = c$, any outcome is efficient.

3. Experimental Design

Groups of $N = 4$ subjects were used in all treatments. They were formed from subject pools of 16, 20, or 24 per session. Group composition remained unknown to subjects during and after the experiment. All sessions lasted 32 rounds.

3.1 Procedure within a decision round

Each round of a session starts with each subject's belief report: they have to estimate the probability that the next organizer will provide the public good. Thereafter, one subject per group is randomly determined to be organizer, with equal probability. Hence, on average, each subject decides in 8 rounds. Subjects are only told whether or not they are the organizer. Visitors are not told who the organizer is. After appointment, the organizer makes the decision whether or not to provide the public good. At the end of each round, each subject in the group is informed about this decision.

3.2 *Parameters and treatments*

In all treatments the costs for provision were $c = 120$ tokens, the experimental currency. The exchange rate was 1 Dutch guilder for 30 tokens.[7] As mentioned above, revenues are equal to zero tokens, $r(0) = 0$, in case of no provision. Our first distinction in treatments is with respect to revenues in case of provision. These were either equal to 40 tokens, $r(120) = 40 < c$, or to 60 tokens, $r(0) = 60 < c$. This gives a 'marginal per capita return' (MPCR) (Isaac and Walker, 1988) of $0.\overline{3}$ and 0.5, respectively. The second distinction in treatments studied is related to the matching protocol, distinguishing 'strangers' and 'partners' (Andreoni, 1988). In partners, groups are randomly determined at the beginning of the first round and remain constant thereafter. In strangers, groups are randomly redetermined before each round.[8] All information so far is known to the subjects. Table 1 summarizes the treatments.

Table 1

Summary of treatments

Treatment	***N***	**Costs *c***	**Revenues *r*(0) / *r*(120)**	**MPCR**	**Matching**	**# Rounds**	**# Observations (# Sessions)**
HP	4	120	0 / 60	0.5	Partners	32	10 (2)
LP	4	120	0 / 40	$0.\overline{3}$	Partners	32	10 (2)
LS	4	120	0 / 40	$0.\overline{3}$	Strangers	32	6 (3)

Note: 'H' denotes 'high MPCR', 'L' 'low MPCR', 'P' 'partners', and 'S' 'strangers'.

For HP we organized 2 sessions with 20 subjects each, for LP 2 sessions, one with 16 and one with 24 subjects, and for LS 3 sessions, one with 16 subjects and two with 24 subjects. Hence, for each of the partners treatments we had 10 independent observations. For LS we divided the 16 and 24 subjects into 2 sub-pools of 8 respectively 12 each. Matching only took place

[7] 1 Dutch Guilder is worth approximately 0.45 Euro Cent.

[8] Note that being in a common group in the organizer's game does not necessarily mean that one can always signal a decision, because only one subject makes a choice.

within a sub-pool. Subjects were not informed about this. By this procedure we obtained 2 independent observations per LS session, 6 in total.

3.3 Reporting beliefs

Before determination of the organizer, subjects are asked to estimate the probability that the next organizer will provide the public good. A subject appointed to be organizer is not rewarded for this estimation. Instead, she receives her previous round's reward for estimation again.[9] Visitors' estimations are rewarded according to the quadratic scoring rule

$$Q_{obs}(b) = a + 2b\,b_{obs} - b\sum_{h=0}^{1}\left(b_h^2\right), \tag{4}$$

with $Q_{obs}(b)$ denoting the payoff for reported beliefs b, b_{obs} the probability that the subject estimated for the event that was actually observed, and b_h the reported (deduced) probability that the next organizer will, $h = 1$ (will not, $h = 0$), provide the public good. For the experiment the parameters chosen were $a = b = 20$ for the low MPCR treatments and $a = b = 30$ for the high MPCR treatment.

The quadratic scoring rule is incentive compatible. The closer a subject's belief report is to the observed value the more money she makes (*e.g.*, Murphy and Winkler, 1970; Selten, 1998). With this incentive scheme, true revelation of beliefs is advantageous for subjects. Note that although there is a chance of being appointed organizer, hence, not being paid according to the actual belief report, it is still in the subjects' interest to reveal beliefs truthfully. This is because of the positive probability of becoming a visitor.

3.4 General procedures and earnings

The computerized[10] experiment was carried out at the laboratory of the 'Center for Experimental Economics and Political Decision Making (CREED)' at the University of Amsterdam (*cf*. Appendix B for the read-aloud instructions). Altogether, 144 subjects were recruited from the university's student population. Each session lasted between 2 and 2.5 hours. The average earning differed across treatments including a show up fee of 10 Dutch

[9] Organizers in the first round receive a fixed reward for estimation of 15 tokens in low MPCR treatments and 22.5 tokens in the high MPCR treatment.

[10] For programming we used the experimental toolbox RatImage (Abbink and Sadrieh, 1995).

guilders[11] and the average earning of 11.34 Dutch guilders for a value orientation test.[12] On average, subjects earned a total of 64.57 Dutch guilders in HP, 46.52 in LP, and 39.62 in LS.

4. General results

In this section, we analyze beliefs and provisions at the aggregate and group level. The experiment provides independent observations for beliefs and decisions at the group level for partners and at the sub-pool level for strangers.[13] All our nonparametric statistical analysis is based on these levels of independent observations.[14]

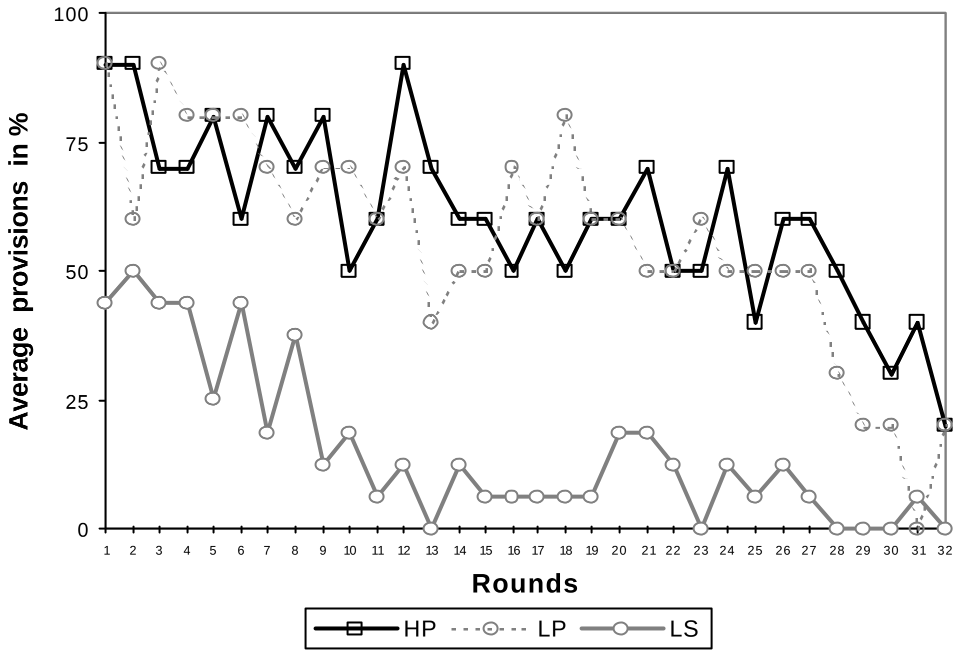

Fig. 1. Average provision (in %) per round at the aggregate level for all treatments.

[11] In two LS session subjects received a show up fee of 20 Dutch guilders. For the calculation of the averages we treated them 'as if 10 Dutch guilders'.

[12] In all treatments subjects went through a so-called value orientation test (see, *e.g.*, Liebrand, 1984; Offerman *et al*. 1996) before the experimental organizer's game started. This computerized test, which lasted on average 30 minutes, measures subjects' concerns about own and others well-being. The results were reported to the subjects at the end of the whole session. Since we do not use the data for the main analysis reported in this paper, we do not discuss the test in detail here.

[13] To simplify things, we use the term 'group level' to refer to groups of 4 in partners and sub-pools of 8 and 12 in strangers.

[14] All nonparametric statistical tests that we use are described in Siegel and Castellan, Jr. (1988).

4.1 Aggregate level

Fig. 1 presents the aggregate average provision (in %) per round for all three treatments. Contrary to the usual simultaneous move results (again, *e.g.*, Isaac and Walker, 1988) average provision does not monotonically decline across rounds in partners: it settles at relatively high levels (56-60%) during the rounds 11-27. Finally, we observe exceptionally high initial contribution levels in partners (90% in round 1 of both HP and LP).

RESULT 1: There are no systematic differences in provision between high and low MPCRs.

The patterns of average provisions of both HP and LP look virtually the same. A Wilcoxon-Mann-Whitney test cannot reject the null hypothesis of no difference (10% significance level, two-tailed test). This distinguishes our result from other studies, where a treatment effect is often observed. *E.g.*, Isaac and Walker (1988) find that a higher MPCR results in higher contributions.[15]

RESULT 2: Partners provide (much) more often than strangers do.

The patterns for LP and LS look very different, with substantially higher levels of provision in partners. A Wilcoxon-Mann-Whitney test clearly rejects the null hypothesis of no difference (1% significance level, two-tailed test). Contrary to LP, we find a fast decline in average provision for LS at the beginning followed by relatively stable, low contributions in rounds 11-27 (LP: 56%; LS: 9%). The strangers result is comparable to that in, *e.g.*, Isaac and Walker (1988) with a MPCR of .3. Nevertheless, the partners-strangers effect in standard public goods experiments is controversial in the literature (*cf*. Andreoni and Croson, 2002, for an overwiew). Some results show that strangers contribute more than partners (*e.g.*, Andreoni, 1988), some find no effect (*e.g.*, Weimann, 1994), and other studies report higher contributions by partners than strangers (*e.g.*, Croson, 1996). Our results support the latter.

RESULT 3: There is a strong end-effect for partners, but no end-effect for strangers.

A strong end-effect occurs in the last 5 rounds of HP and LP. The difference between average provision in the first 27 rounds and the last 5 rounds is high (HP: 65% vs. 36%; LP: 63% vs.

[15] Compared to previous studies, our variation in MPCR is limited. Nevertheless, a priori we would have expected to find higher provision for higher MPCR.

18%). No end-effect is observed in strangers: the decline of average provisions starts early and reaches an average of 9% in rounds 11-27 followed by 1% in the last 5 rounds. Our results on end-effects are comparable to studies with simultaneous-move protocol (*e.g.*, Keser and van Winden, 2000).

RESULT 4: Average expected provisions are close to average actual provisions.

Consider Figs. 2(a)-(c). Average beliefs correspond closely to average provisions across rounds. A closer look reveals that average beliefs tend to be higher than average decisions, indicating that beliefs are following decisions. The average belief lines are smoother due to the higher number of observations, because both organizers and visitors report beliefs, whereas only organizers decide whether or not to provide.[16] A Wilcoxon signed ranks test rejects the null hypothesis of no difference between beliefs and contributions in favor of higher beliefs than provisions for LP (1% significance level) and LS (5% level), but not for HP (10% level) (all one-tailed tests).

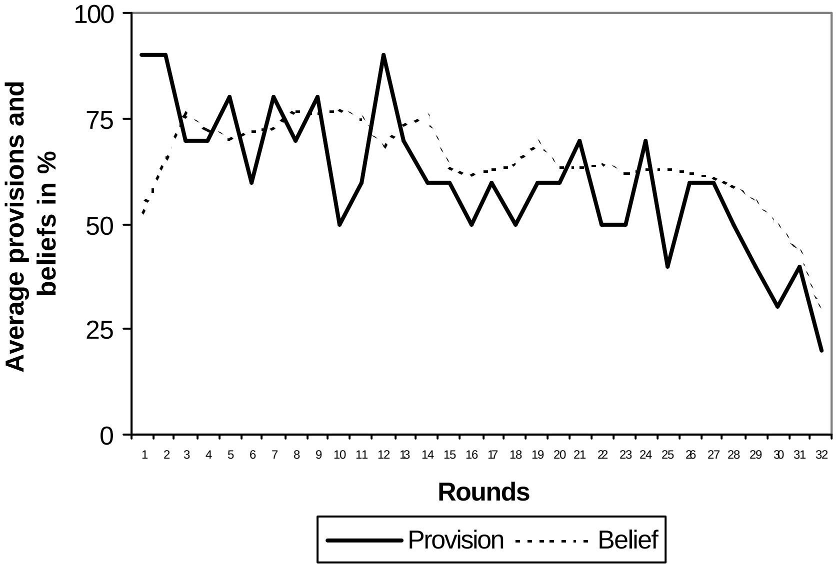

Fig. 2(a). Average aggregate provision and belief for HP.

[16] The null hypothesis of no difference between beliefs of organizers and visitors cannot be rejected (Wilcoxon signed ranks test, 10% significance level, two-tailed).

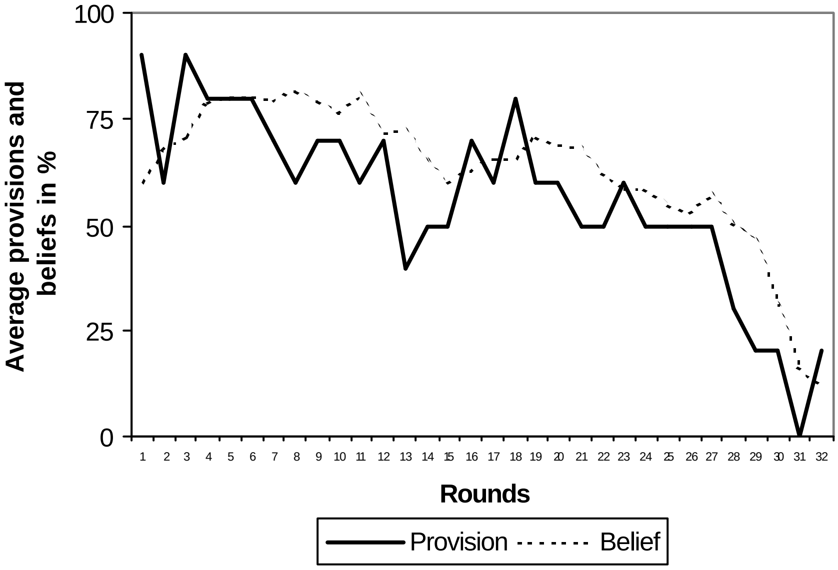

Fig. 2(b). Average aggregate provision and belief for LP.

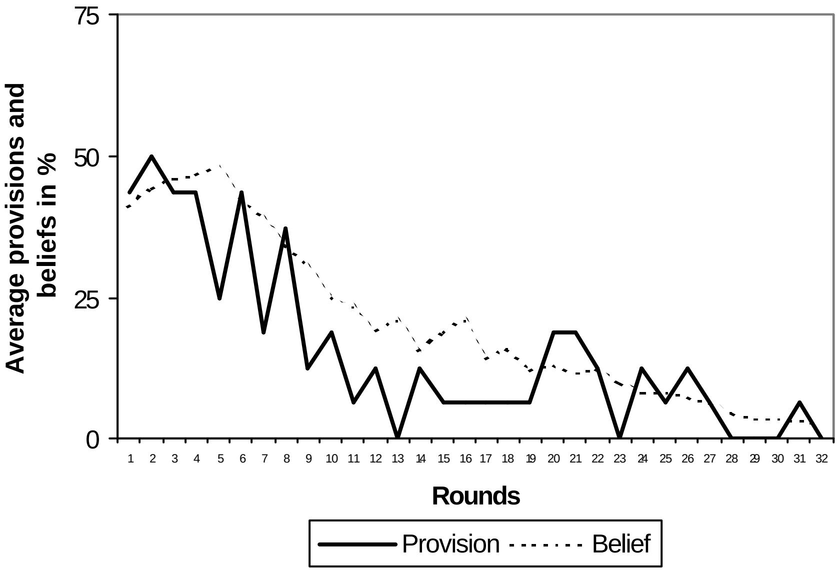

Fig. 2(c). Average aggregate provision and belief for LS.

4.2 Group level

RESULT 5: Average group provisions and average group beliefs are highly correlated for all treatments.

For all treatments average beliefs and provisions across rounds are correlated. The Spearman rank-order correlation coefficients are $r_S = .96$ for HP, $r_S = .98$ for LP, and $r_S = .93$ for LS. The former two coefficients are significant at the 1% level and the latter coefficient at the 5% level (all two-tailed tests).

RESULT 6: Average group provisions in rounds 1-4 and rounds 5-32 are correlated in both partners treatments but not in the strangers treatment.

Average group provisions in the first 4 rounds are positively correlated with average provisions in the last 28 rounds for partners, but not for strangers. The Spearman rank-order correlation coefficients are $r_S = .82$ for HP (1% significance level), $r_S = .66$ for LP (5% level), and $r_S = .14$ for LS (insignificant) (all two-tailed tests). This result suggests that early rounds have an important impact on further decisions in partners, but not in strangers.

5. Belief learning

In experiments learning is often observed that cannot be rationalized in the Bayesian-Nash framework (see, *e.g.*, Camerer *et al.*, 2001, for functional experience-weighted attraction learning in various experiments; Offerman *et al.*, 1998, for adaptive learning in experimental step-level public goods games; Roth and Erev, 1995, for reinforcement learning in public goods, market, and ultimatum bargaining experiments; Selten and Stoecker, 1986, for learning direction theory in repeated experimental prisoners' dilemma games). Assuming own payoff maximizing players and parameters as in our experimental organizer's game, theory predicts no provision of the public good and that everybody believes that no provision will occur. Hence, rational inference is sufficient and learning behavior excluded.

Nevertheless, there is an opportunity to learn in public goods experiments. If one drops the assumption that (everyone knows that) everyone is an own payoff maximizer, the learning can be related to others' preferences. Hence, the opportunity to learn can motivate subjects to try to influence other subjects' beliefs and decisions, *e.g.*, by building a reputation of cooperation. If reputation building is successful and (some) subjects provide jointly, they can earn substantially more money than in the Nash equilibrium.

An important distinction often made is between adaptive and sophisticated learning (*e.g.*, Milgrom and Roberts, 1991).[17] In adaptive learning models, *e.g.* fictitious play (Brown 1951) and naïve Bayesian learning (Eichberger *et al.*, 1993), decision makers take strategies of others as given and seek to learn more about these. In sophisticated learning models decision makers do not take others' strategies as given but learn more complex patterns in behavior of others (see, *e.g.*, Brandts and Holt, 1995; Fudenberg and Levine, 1998, chapter 8; Milgrom and Roberts, 1991). If experimental subjects indeed succeed in influencing each other and building reputations, their behavior can be called sophisticated.

For our purposes, learning direction theory as introduced in Selten and Stoecker (1986) turns out to be a simple but appropriate model of learning behavior.[18] Here, we use learning direction theory to test for belief learning in the experimental organizer's game. To the best of our knowledge, we are the first to apply the theory to data on beliefs. We do so, as follows: First, we present subjects' average scores for their belief reports. Second, we briefly introduce learning direction theory in the context of the experimental organizer's game and analyze whether it can explain our belief data. Third, the theory is applied to detect if there are differences between belief learning of organizers and visitors.

5.1 Average scores for belief reports

RESULT 7: On average, beliefs become more accurate over time.

Table 2 shows the average earnings as a proportion of the maximum possible, for belief reports by subjects in blocks of 8 rounds. The lowest values per treatment are found in early rounds. For LS the performance improves monotonically over the rounds. For both partner treatments there is a drop in average scores in late rounds. This can be attributed to the end-effect discussed earlier (see Result 3), the beginning of which was not anticipated by all subjects. A Page test for ordered alternatives rejects the null hypothesis of no difference in favor of the above-mentioned ordering in strangers and in favor of '1-8' < '9-16' < '25-32' < '17-24' in partners at least at the 5% significance level in all cases (all one-tailed tests).

[17] A good and extensive overview of learning in games is given in Fudenberg and Levine (1998).

[18] Further applications and extensions of the theory in various fields can be found *e.g.* in Mitzkewitz and Nagel (1993), Nagel (1995), Kuon (1994), Selten and Buchta (1998), and Anderson *et al.* (1999).

Table 2

Average scores for subjects' belief estimations

Treatment	Rounds				Total
	1-8	9-16	17-24	25-32	
HP	.740	.744	.855	.798	.784
LP	.782	.826	.890	.871	.842
LS	.606	.833	.853	.935	.807

5.2 Belief learning and learning direction theory

In the learning direction theory of Selten and Stoecker (1986) the direction of learning is explained, but not the extent of changes. Predictions are qualitative rather than quantitative. In the context of belief learning in the experimental organizer's game, the direction of belief changes is predicted in response to two possible observations: provision and no provision of the public good. This can be formalized as follows. For the sequence of rounds $t = 1,2,\dots,T$, the decision maker has to report her belief b_t (in %) that the next organizer will provide the public good. Before she reports a belief, she observes whether or not the public good was provided in the previous round (except in round 1). Learning direction theory predicts:

- $b_t \leq b_{t-1}$ in case no provision is observed in $t-1$, and
- $b_t \geq b_{t-1}$ in case provision is observed in $t-1$.

Note that the theory allows a decision maker not to respond to observations. However, if she decides to change her belief report, the change is expected to be in the right direction. The interpretation is that the decision maker could have yielded a higher or equal score for her previous belief report, if she would have stated it according to the hypothesis.

Tables 3(a)-(c) present the belief data at the aggregate level for each treatment by providing an overview of the number of changes and no changes of beliefs after observing provision or no provision for rounds 2-32. In each cell frequencies and proportions (in brackets) are shown. For the second to the fourth column proportions relate to the row sums.

Table 3(a)

Learning direction of beliefs for HP

Observation	Direction of belief change			Row sums
	Decrease (-)	Unchanged (0)	Increase (+)	
No provision	189 (.40)	223 (.47)	60 (.13)	472 (.38)
Provision	92 (.12)	424 (.55)	252 (.33)	768 (.62)
Column sums	281 (.23)	647 (.52)	312 (.25)	1.240 (1.00)

Table 3(b)

Learning direction of beliefs for LP

Observation	Direction of belief change			Row sums
	Decrease (-)	Unchanged (0)	Increase (+)	
No provision	166 (.31)	319 (.60)	43 (.08)	528 (.43)
Provision	30 (.04)	554 (.78)	128 (.18)	712 (.57)
Column sums	196 (.16)	873 (.70)	171 (.14)	1.240 (1.00)

Table 3(c)

Learning direction of beliefs for LS

Observation	Direction of belief change			Row sums
	Decrease (-)	Unchanged (0)	Increase (+)	
No provision	358 (.22)	1181 (.71)	129 (.08)	1.668 (.84)
Provision	37 (.12)	143 (.45)	136 (.43)	316 (.16)
Column sums	395 (.20)	1.324 (.67)	265 (.13)	1.984 (1.00)

RESULT 8: Subjects change their beliefs in accordance with learning direction theory.

The numbers in Tables 3(a)-(c) are in agreement with the belief change hypothesis for all treatments. In HP we find more decreases than increases in beliefs after no provision has been observed (HP: 40% vs. 13%; LP: 31% vs. 8%; LS: 22% vs. 8%). After observing provision, subjects increased their beliefs more often than they decreased them (HP: 33% vs. 12%; LP: 18% vs. 4%; LS: 43% vs. 12%). A Wilcoxon signed ranks test clearly rejects the null hypothesis of no difference for all treatments: there is no group with more increases (decreases) after observing no provision (provision).

5.3 Belief learning of organizers and visitors

The sequential-move protocol of the repeated experimental organizer's game is particularly useful to isolate learning behavior of subjects in the organizer's and visitor's position. For the former, learning can only be due to the possible influence of the own decision on another, next organizer's future decision. This is because organizers report their beliefs about others before and after they make a decision themselves. The case is different for visitors, where learning can only be due to the observation of another subject's decision. According to adaptive learning models, subjects in the organizer's position should not change their beliefs about the next organizer providing the public good, simply because they do not get further information about others' strategies. Visitors on the other hand could change their beliefs according to an adaptive learning rule.

RESULT 9: In aggregate, there is no difference in directional belief learning of organizers and visitors.

Table 4 shows the aggregate proportions of changes of reported beliefs for organizers and visitors. The numbers indicate no difference in learning behavior between organizers and visitors: both change their beliefs in accordance with learning direction theory. The proportion of decreases is higher than that of increases after no provision has been observed, and vice versa after provision has been observed. A Wilcoxon signed ranks test clearly rejects the null hypothesis of no difference for visitors: there is no counter-observation. The same test rejects the null hypothesis of no difference for organizers for all situations at the 10% significance level or better, except for HP after no provision is observed (insignificant).[19] We conclude that in most situations organizers change their beliefs based on their own decisions. In a way, like visitors, they update as if they observe the decision of another organizer. This cannot be explained by adaptive learning, but asks for more sophisticated learning models. I.e., a model that implies some kind of expected influence of own decisions on others' decisions.

[19] For both organizers and visitors we had to exclude groups from the statistical analysis due to zero cells.

Table 4

Learning direction of beliefs for organizers and visitors for all treatments

Treatment	Observation	Direction of belief change*	Organizers	Visitors	Total
HP	No provision	–	.27	.44	.40
		0	.49	.47	.47
		+	.24	.09	.13
	Provision	–	.14	.11	.12
		0	.58	.54	.55
		+	.28	.35	.33
LP	No provision	–	.32	.31	.31
		0	.58	.61	.60
		+	.10	.08	.08
	Provision	–	.04	.04	.04
		0	.79	.78	.78
		+	.17	.18	.18
LS	No provision	–	.18	.23	.22
		0	.73	.70	.71
		+	.09	.07	.08
	Provision	–	.17	.10	.12
		0	.49	.44	.45
		+	.34	.46	.43

*A decrease in belief is represented by ‘–’, an unchanged belief by ‘0’, and an increase by ‘+’.

6. Reciprocity

In this section, we analyze whether subjects behave reciprocally. Fehr and Gächter (2000b) draw a line between reciprocal and cooperative behavior. According to them (p. 160), the former holds if “the actor is responding to friendly or hostile actions even if no material gains can be expected”, whereas the latter arises “because actors expect future material benefits from their actions.” Following this definition, reciprocal behavior can be spotted best if subjects encounter once, whereas cooperative behavior needs repeated encounters (both one-shot and repeated encounters are compared in Gächter and Falk, 2002, in the context of a experimental gift exchange game). A similar notion for cooperative behavior often used in the literature is that of conditional cooperation (*e.g.*, Offerman *et al.*, 1996; Keser and van Winden, 2000). There, subjects provide as long as others do the same or are expected to do so.

Moreover, it has been shown that cooperative behavior increases dramatically if there is a opportunity to reward or punish other subjects individually for their behavior (see, *e.g.*, Fehr and Gächter, 2000a, 2002; Sefton *et al*., 2002).

Here, our aim is to systematically study a specific type of reciprocity, which we call cut-point belief reciprocity. It is applicable to both one-shot and repeated encounters. This concept does not distinguish between reciprocal and cooperative motivations. Hence, we prefer to use the notion of reciprocity in a more general sense. In a broad definition, a decision maker behaves positively reciprocally, if she rewards observed or expected choices of others that are considered to be friendly. She behaves negatively reciprocally, if she punishes observed or expected choices of others that are considered to be hostile (*cf*. Abbink *et al.*, 2000). Note that in the experimental organizer's game a subject cannot distinguish among decisions of others. Notice further that our definition allows reciprocity to be based not only on observed but also on expected decisions. The latter can be particular useful in one-shot encounters (*e.g.*, Bosman and van Winden, 2002; Großer and Ule, 2002). We use both our data on decisions and beliefs to analyze decision-based and belief-based reciprocity. In order to obtain deterministic predictions when beliefs are incorporated, a decision rule has to be specified on how decisions are based on beliefs. We suggest and analyze a cut-point belief reciprocal decision rule.

This section proceeds as follows. First, we analyze the data according to decision-based reciprocity. Second, belief-based reciprocity is tested for. Third, we introduce and investigate cut-point belief reciprocity, using a simple error classification analysis. Finally, performances of decision-based and cut-point belief reciprocity are compared in terms of error rates.

6.1 Decision-based reciprocity

In this section we investigate if the experimental data support a simple definition of decision-based reciprocity (*cf*. Clark and Sefton, 2001). We define

- positive decision-based reciprocity as an organizer's choice to provide the public good after observing provision by another organizer in the previous round, and
- negative decision-based reciprocity as an organizer's choice not to provide the public good after observing no provision by another organizer in the previous round.

Note that the definition only considers the previous round. Data where organizers were also organizers in the previous round are not included in the analysis. The results are presented in Tables 5(a)-(c), which show provision and no provision in response to the previously observed decision made by another organizer. Each cell shows frequencies and proportions (in brackets). The proportions of the second and third columns relate to the row sums.

Table 5(a)

Decision-based reciprocity for HP

Observation	**Decision**		**Row sums**
	No provision	**Provision**	
No provision	53 (.63)	31 (.37)	84 (.37)
Provision	35 (.24)	109 (.76)	144 (.63)
Column sums	88 (.39)	140 (.61)	228 (1.00)

Table 5(b)

Decision-based reciprocity for LP

Observation	**Decision**		**Row sums**
	No provision	**Provision**	
No provision	74 (.79)	20 (.21)	94 (.41)
Provision	22 (.16)	116 (.84)	138 (.59)
Column sums	96 (.41)	136 (.59)	232 (1.00)

Table 5(c)

Decision-based reciprocity for LS

Observation	**Decision**		**Row sums**
	No provision	**Provision**	
No provision	260 (.87)	38 (.13)	298 (.85)
Provision	35 (.67)	17 (.33)	52 (.15)
Column sums	295 (.84)	55 (.16)	350 (1.00)

RESULT 10: We observe positive and negative decision-based reciprocity for both partners and only negative decision-based reciprocity for strangers.

In aggregate, we find negative decision-based reciprocity in all treatments: after having observed no provision, organizers more often choose no provision than provision. For partners we also find positive reciprocity: after having observed provision, provision is more frequently chosen by organizers than no provision. However, there is no positive reciprocity in LS, where no provision occurs more often than provision after having observed provision. The aggregate results are confirmed for LP and LS, and partly for HP. A Wilcoxon signed ranks test rejects the null hypothesis of no difference at the 10% significance level or better in favor of positive reciprocity for HP, positive and negative reciprocity for LP and negative reciprocity in LS, but not (10% significance level) for negative reciprocity in HP or positive reciprocity in LS (all one-tailed tests).[20]

6.2 Belief-based reciprocity

Here, we investigate the relation between subjects' decisions and their beliefs about the next organizer providing the public good. One relationship that has been found in experiments is that subjects who provide the public good estimate the probability that others will contribute to be higher than those who do not provide (*e.g.*, Offerman *et al.*, 1996). We define

- positive belief-based reciprocity as an organizer's choice to provide the public good if she has a high belief about others providing, and
- negative belief-based reciprocity as an organizer's choice not to provide the public good if she has a low belief about others providing.

The cells in Tables 6(a)-(c) show average organizers' beliefs for both previously observed decisions and for actual decisions (*cf*. Appendix A, Fig. A-1 and Table A-2, for proportions of provisions and numbers of observations per belief category).

[20] For some situations we had to exclude groups from the statistical analysis because of no observation.

Table 6(a)

Belief-based reciprocity for HP

Observation	Decision		Row average belief
	No provision	Provision	
No provision	20.04 (53)	60.68 (31)	35.04 (84)
Provision	54.89 (35)	90.17 (109)	81.60 (144)
Column average belief	33.90 (88)	83.64 (140)	64.44 (228)

Table 6(b)

Belief-based reciprocity for LP

Observation	Decision		Row average belief
	No provision	Provision	
No provision	16.04 (74)	64.95 (20)	26.45 (94)
Provision	54.91 (22)	94.93 (116)	88.55 (138)
Column average belief	24.95 (96)	90.52 (136)	63.39 (232)

Table 6(c)

Belief-based reciprocity for LS

Observation	Decision		Row average belief
	No provision	Provision	
No provision	8.38 (260)	53.74 (38)	14.16 (298)
Provision	39.03 (35)	71.41 (17)	49.62 (52)
Column average belief	12.02 (295)	59.20 (55)	19.43 (350)

RESULT 11: There is negative and positive belief-based reciprocity in all treatments.

We observe positive and negative belief-based reciprocity in all treatments. A Wilcoxon signed ranks test rejects the null hypothesis of no difference in favor of higher beliefs for providers than non-providers at the 10% significance level or better for both previously observed choices in all treatments (all one-tailed tests). Hence, the additional information of organizers' beliefs can explain the decisions for both observation conditions in all treatments. Note the particular order in the data. In all three tables the following order of average beliefs (in %) that another, next organizer will provide the public good can be observed: 'provision after observing provision' > 'provision after observing no provision' > 'no provision after observing no provision' > 'no provision after observing provision'. This ordering is not easily tested at the group level, since there are 12 groups with no observations in at least one of the four possible situations. The ordering raises the question whether some sort of cut-point belief separates the decision to provide from that to not provide.

6.3 *Cut-point belief reciprocity*

We define cut-point (b_{cut}) belief reciprocity as a decision where the organizer

- provides the public good if $b_t > b_{cut}$,
- does not provide the public good if $b_t < b_{cut}$,
- and either provides or not if $b_t = b_{cut}$,

where b_t is the reported belief in round t. The first case represents positive and the second negative reciprocity.

We investigate cut-point beliefs by using a simple error classification analysis (in a different context, see Palfrey and Prisbrey, 1996, 1997). This analysis calculates for each possible b_{cut} upper and lower relative errors for the experimental data. A lower relative error e_{low} is defined as the observed number of unexpected provisions for all beliefs lower than b_{cut}, divided by the number of observed beliefs lower than b_{cut}. Similarly, a upper relative error e_{up} is defined as the observed number of unexpected decisions not to provide for all beliefs higher than b_{cut}, divided by the number of observed beliefs higher than b_{cut}. The total relative error e_{total} is simply the sum of e_{low} and e_{up}. Formally,

$$e_{low}(b_{cut}) = \frac{\sum(provision|b < b_{cut})}{\sum(provision|b < b_{cut}) + \sum(no\ provision|b < b_{cut})}, \tag{5}$$

$$e_{up}(b_{cut}) = \frac{\sum(no\ provision|b > b_{cut})}{\sum(provision|b > b_{cut}) + \sum(no\ provision|b > b_{cut})}, \tag{6}$$

and

$$e_{total}(b_{cut}) = e_{low}(b_{cut}) + e_{up}(b_{cut}). \tag{7}$$

The optimal cut-point belief is determined as the one with the lowest total relative error. Figs. 3(a)-(c) depict total, lower, and upper relative errors for the three treatments. The straight horizontal lines show the aggregate relative errors produced by decision-based reciprocity.

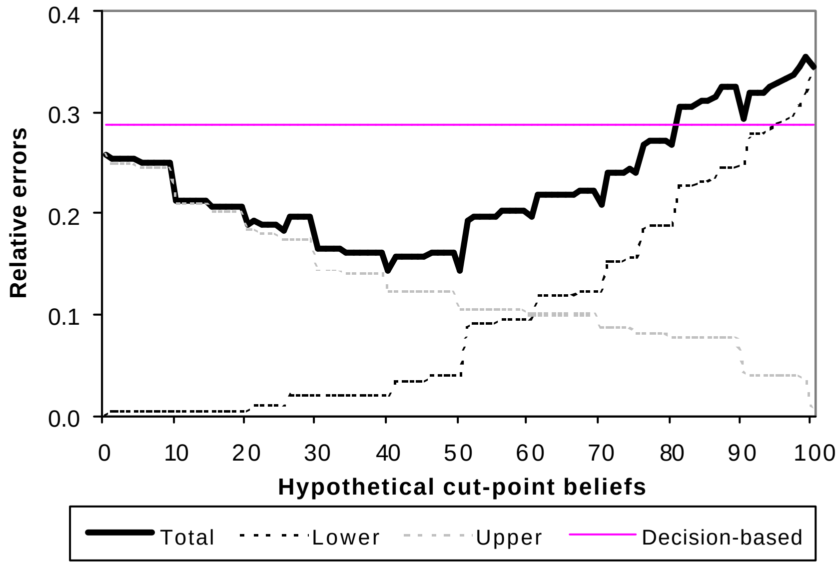

Fig. 3(a): Total, lower, and upper relative errors for hypothetical cut-point beliefs in HP.

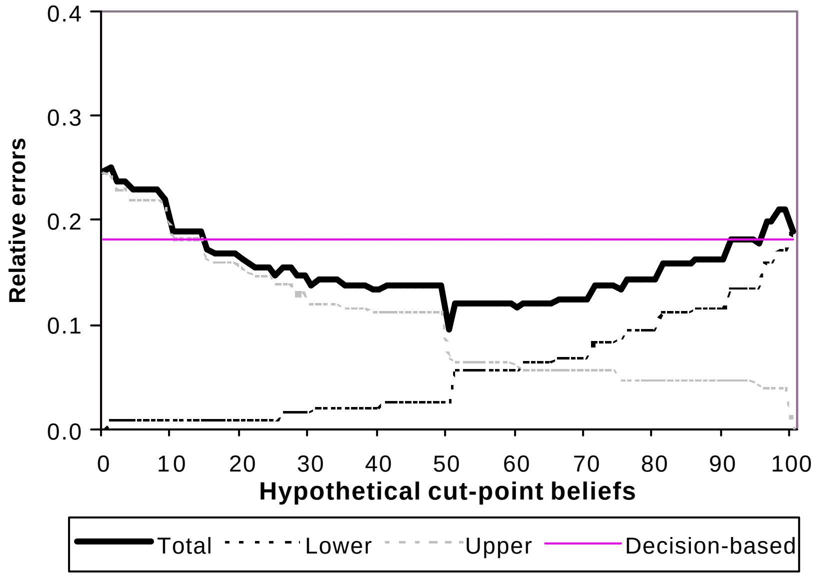

Fig. 3(b): Total, lower, and upper relative errors for hypothetical cut-point beliefs in LP.

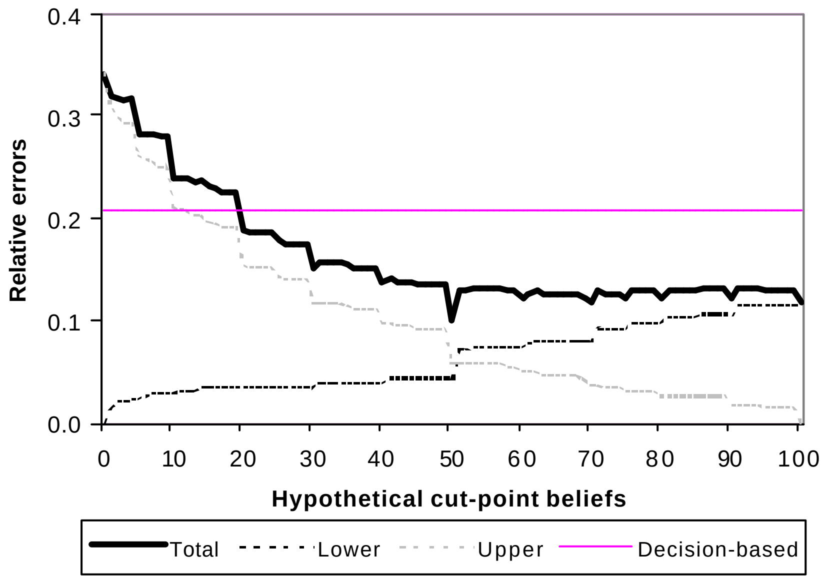

Fig. 3(c): Total, lower, and upper relative errors for hypothetical cut-point beliefs in LS.

RESULT 12: In all treatments there is the same optimal cut-point belief of 50% for a simple error classification analysis. HP has an additional cut-point belief of 40%.

Figs. 3(a)-(c) show that in all treatments a wide range of cut-point beliefs (about 80% in each) performs better than decision-based reciprocity. Moreover, all treatments have the same optimal cut-point belief of 50%. Only HP has an additional optimal cut-point belief of 40%. Hence, we observe a tendency towards the underlying norm that at least half of the possible other next organizers should be willing to provide the public good. Note the total relative error lines can be quite flat in some belief regions due to relatively low numbers of observations (*cf*. Appendix A, Table A-2). This asks for further research with more observations in these regions.

We now compare the performances of decision-based reciprocity and cut-point belief reciprocity. Table 7 presents the respective total relative errors for all treatments.

Table 7

Comparison of total relative errors for decision- and belief-based reciprocity

Treatment	Reciprocity	
	Decision-based	Cut-point belief (50%)
HP	.29	.14
LP	.18	.09
LS	.21	.10

RESULT 13: Cut-point belief reciprocity describes decisions better than decision-based reciprocity.

6.4 Cut-point belief reciprocity and early rounds

As already seen in Result 6, early rounds matter in partners. Another look at Figs. 2(a)-(c) makes clear that cut-point belief reciprocity plays a crucial role for both partners and strangers. In the first round we find for partners average beliefs larger than the cut-point belief of 50% whereas for LS average beliefs are lower than 50%. In other words, if subjects decide according to a cut-point belief reciprocal rule, early beliefs in strangers are generally too low, in fact lower than the cut-point belief of 50%. For partners on the other hand, early beliefs often exceed the cut-point belief. There were 9 (1) organizers in the first rounds in HP with a belief higher than or equal to (smaller than) 50%, of whom 8 (1) provided the public good [LP: 8 (2) and 8 (1); LS: 7 (9) and 6 (0)]. Although the differences between actual organizers' beliefs and optimal cut-point belief are not too large (HP: 9%; LP: 18%; LS: -13%), they seem to be sufficient to determine the development of contribution in subsequent rounds. Because starting beliefs are close to the cut-point, early round decisions are very sensitive because next organizers may update their beliefs below 50%.

7. Conclusions

We report and analyze a repeated experimental organizer's game where just one decision maker of a group, the organizer, has to decide whether or not to provide the public good. The others in the group, the visitors, do not make a decision. Because organizers change across rounds, the experiment has a sequential- rather than simultaneous-move character. We find higher provision rates in the experimental organizer's game than is observed in more common simultaneous-move counterparts for groups that stay together for all rounds. For these groups, we also find a sharp end-effect, similar to that observed in simultaneous-move experiments.

We have shown that subjects update their beliefs about others providing the public good in accordance with learning direction theory. Of particular interest is the finding that this holds for both organizers and visitors. However, adaptive learning models cannot explain the observation that organizers learn from their own decisions. This calls for a more sophisticated learning model, which takes the strategic influence of own decisions on others' beliefs and decisions into account. This finding deserves further future research.

We also find that negative decision-based reciprocity occurs. Except for the case when groups are remixed at the beginning of each round, there is also positive decision-based reciprocity. When beliefs are included in the analysis, we show that subjects decide in accordance with belief-based reciprocity. They provide substantially more often if they have higher beliefs about others providing the public good rather than if they have lower beliefs, independent of the previously observed decision. The data also led us to a cut-point belief as a turning point that separates provision from no provision. 50% is the observed optimal cut-point belief in all treatments. The analysis showed that cut-point belief reciprocity describes actual decisions astonishingly well.

Our aim was to provide a simple but straightforward explanation of why people provide voluntarily in public goods environments. In order to focus on the main line of argument, we restricted ourselves to simple tools of analysis. In this respect, we find strong evidence for our concepts, in particular for sophisticated belief learning and cut-point reciprocity. We consider it fruitful to experimentally test and improve our concepts further by applying them to various decision situations and using also other methods, *e.g.*, simulation studies. In particular, the sequential move protocol turned out to be crucial in gaining new insights into the decision making process. Many of these insights could not have been reached under the more standard simultaneous move protocol.

Appendix A: Data

Table A-1

Average provisions and reported beliefs (in %) per treatment

Group	HP			LP			LS		
	Provi-sion	Total beliefs	Organi-zer's. beliefs	Provi-sion	Total beliefs	Organi-zer's. beliefs	Provi-sion	Total beliefs	Organi-zer's. beliefs
1	12.50	22.32	21.69	3.13	13.46	14.56	7.29	11.06	11.59
2	28.13	46.11	41.72	25.00	46.23	52.66	7.82	4.91	3.08
3	31.25	27.04	28.72	34.38	34.49	34.22	14.07	20.96	22.92
4	43.75	62.76	63.59	34.38	46.63	44.81	16.67	21.03	22.96
5	59.38	62.04	60.66	43.75	51.18	59.41	16.67	23.34	24.05
6	68.75	72.13	72.47	68.75	76.07	73.31	27.09	36.27	43.56
7	81.25	77.18	78.94	81.25	85.94	90.63			
8	87.50	90.60	93.78	84.38	86.20	85.53			
9	93.75	93.18	91.53	90.63	91.88	95.00			
10	100.00	93.02	96.34	96.88	95.51	95.00			

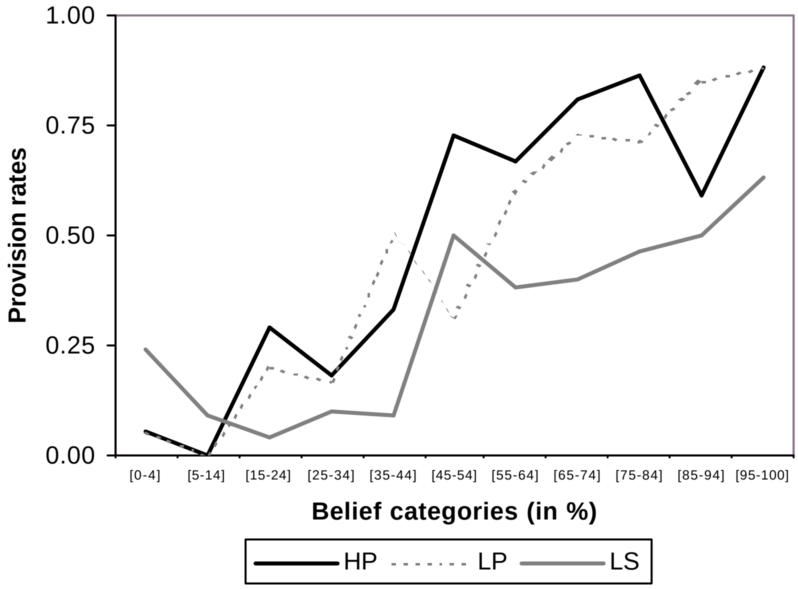

Fig. A-1. Provision rates per belief category.

Table A-2

Numbers of observations per belief category

Treatment	Belief categories (in %)											Total
	0-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95-100	
HP	44	12	14	11	12	26	9	26	22	32	112	320
LP	64	15	10	6	2	32	5	11	7	13	155	320
LS	293	43	23	21	11	30	8	20	13	12	38	512

Appendix B: Instructions for treatment HP [*LP*, *LS*]

General instructions

Welcome to our experiment on decision making. Everybody will receive a **base payment of 10 Guilders** for participation in the experiment. Depending on your *own choices*, *own probability estimations*, and the *choices of other participants*, you may earn more money today. The experimental currency is *tokens*. **30 tokens are worth one Guilder**. At the end of the experiment your token earnings will be exchanged into Guilders and paid to you in cash. The payment will remain *anonymous*. No other participant will be informed about your payment.
(…) We will now start reading the instructions (…). After that, you will have the opportunity to ask questions. If you have any questions, please raise then your hand. One of the experimenters will come to you to answer them. **You are not allowed to communicate with other participants during the experiment**. (…).

Instructions part 1 (…) [value orientation test, not reported here]

Instructions part 2

We are ready with part 1 of the experiment. As you know, you will get to know your earnings from part 1 only after part 2 has been completed. We will now read out the instructions for the second part.

Rounds and groups of participants

Part 2 consists of *32 rounds*. At the beginning of the *first* [*LS*: *each*] round you will be matched with three other participants (*P1*, *P2*, and *P3*). Hence, you are part of a group of four participants. The matching is random and will be conducted by the computer program. **Note that you will be part of the same group in all rounds. Hence, you are always matched with the same participants [*LS: Note that the groups will be re-matched at the beginning of each round*]**.

Determination of the decision maker

In *each round* one participant in your group will be selected to make a choice. This selection is random and will be done by the computer program. The selected participant is called *decision maker* in the further course. Every participant in a group has the same chance of ¼ to become the decision maker. The chance will be represented by a *wheel of fortune*. (…)

Choices of the decision maker and round earnings for choices

In each round the decision maker has to choose between two alternatives, *choice 'A'* and *choice 'B'*. If the decision maker chooses 'A', then everybody in the group, inclusive the decision maker, receives a *revenue* of 0 tokens in this round. There are no costs related to choice 'A'. If the decision maker chooses 'B', then everybody in the group receives a revenue of 60 [*LP and LS*: *40*] tokens. There are, however, costs related to choice 'B'. These costs are 120 tokens and are *only* paid by the decision maker. Independent of whether you are the decision maker or not, in each round your *choice earning* will be calculated as revenue minus costs. In case of choice *A*, the choice earning is equal to 0 for everybody. In case of choice *B*, the choice earning of the decision maker is minus 60 [*LP and LS*: *minus 80*] tokens and for the others plus 60 [*LP and LS*: *plus 40*] tokens. (…)

Probability estimations

Before the decision maker is selected and the choice is made, everybody in the group will be asked to estimate *how big the probability is that the decision maker will choose 'B' in case oneself will **not** be the decision maker*. The estimate will be expressed in percentage and can be any number from 0 to 100. Note that at the moment you enter your probability estimation, you do not know whether you or one of the other three participants will be the decision maker in that round. (…)

Round earning for probability estimations

Depending on the actual choice of the decision maker, you will receive a *earning for your probability estimation*. You will earn at least 0 tokens and at most 30 [*LP and LS*: *20*] tokens for your probability estimation per round. We do not go any further into the calculation of the earning for your probability estimation, but if you are interested, we will explain the formula to you after the experiment. **It is important to know that it is in your own interest to report your true probability estimation**. In this case your expected earning is highest. The decision maker her- or himself will *not* be paid for her or his probability estimation. She or he will receive an earning, which corresponds to the most recent earning received for probability estimation when she or he was not the decision maker. Note that this earning is independent of her or his current probability estimation. If there is no earlier round in which the current decision maker was not the decision maker, for instance in the first round, then the current decision maker gets a round earning for probability estimation of 22.50 [*LP and LS*: *15*] tokens.

Computer screen

(1) The *Status* window depicts the *current* round and your *total earnings in part 2*, which is the sum of all round earnings so far.

(2) The *Previous round* window shows all events that occurred in the previous round. There you can see whether you or one of the other three participants in your group was the decision maker, the choice that was made, and your revenue and costs. Moreover, the window depicts your probability estimation of the previous round and your related earning. Finally, your *total round earning* is shown. The total round earning is the sum of the earnings for choice and probability estimation.

(3) At the upper bound of the screen you will find the *Menu bar*. You can use this to access the **Calculator** and **History** functions. The calculator can be handled with the number pad on the right hand side of you keyboard or with the mouse buttons. The function 'history' shows all information of the last *sixteen rounds* as this had appeared in the window 'Previous round'. At the bottom of your screen the *Information bar* is located. There you are told the current status of the experiment.

Further procedures part 2

Part 2 starts with *three training-rounds*. Choices and probability estimations in these training-rounds have no impact on your earnings. Only after the training-rounds we will start with the 32 rounds that determine your earnings. If you have any question during the experiment, raise your hand. One of the experimenters will come to you. If you have any questions at the moment, you can ask them now. After answering all of your questions we will begin with part 2 of the experiment.

References

Abbink, K., Irlenbusch, B., Renner, E., 2000. The moonlighting game. An experimental study on reciprocity and retribution, *Journal of Economic Behavior and Organization* 42, 265-277

Abbink, K, Sadrieh, A., 1995. RatImage, Research Assistance Toolbox for Computer-Aided Human Behavior Experiments, Discussion Paper No. B-325, University of Bonn

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

Anderson, S., Goeree, J., Holt, C.A., 1999. Stochastic game theory: adjustment to equilibrium under noisy directional learning, Working paper, University of Virginia

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., 2002. Partners versus strangers: random matching in public goods experiments, prepared for Handbook of Experimental Economics Results, (eds) C.R. Plott and V.L. Smith

Bardsley, N., 2000. Control without deception: Individual behaviour in free-riding experiments revisited, *Experimental Economics* 3, 215-240

Bolle, F., Ockenfels, P., 1990. Prisoner's dilemma as a game with incomplete information, *Journal of Economic Psychology* 11, 69-84

Bosman, R., van Winden, F., 2002. Emotional hazard in a power-to-take experiment, *Economic Journal* 112, 147-169

Brandts, J., Holt, C.A., 1995. Naïve Bayesian learning and adjustment to equilibrium in signaling games, Working paper, Universitat Autonoma de Barcelona

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

Brown, G.W., 1951. Iterative solutions of games by fictitious play, in: T.C. Koopmans (ed.), Activity analysis of production and allocation, New York: McGraw-Hill

Camerer, C.F., Ho, T.-H., Chong, J.K., 2001. Behavioral game theory: thinking, learning, and teaching, paper presented at the Nobel Symposium

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

Croson, R., 1996. Partners and strangers revisited, *Economics Letters* 53, 25-32

Eichberger, J., Haller, H., Milne, F., 1993. Naïve Bayesian learning in 2×2 matrix games, *Journal of Economic Behavior and Organization* 22, 69-90

Erev, I., Rapoport, A., 1990. Provision of step-level public goods - the sequential contribution mechanism, *Journal of Conflict Resolution* 34, 401-425

Falk, A., Fischbacher, U., 1998. A theory of reciprocity, Working paper, University of Zurich

Fehr, E., Gächter, S., 1998. Reciprocity and economics: The economic implications of Homo Reciprocans, *European Economic Review* 42, 845-859

Fehr, E., Gächter, S., 2000a. Cooperation and punishment in public goods experiments, *American Economic Review* 90, 980-994

Fehr, E., Gächter, S., 2000b. Fairness and retaliation: the economics of reciprocity, *Journal of Economic Perspectives* 14, 159-181

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 137-140

Fudenberg, D., Levin, D.K., 1998. The theory of learning in games, Cambridge (Mass.): MIT Press

Gächter, S., Falk, A., 2002. Reputation and reciprocity: consequences for the labour relations, *Scandinavian Journal of Economics* 104, 1-26

Goeree, J.K., Holt, C.A., 2002. An explanation of anomalous behavior in binary-choice games: entry, voting, public goods, and the volunteers' dilemma, Working paper, University of Virginia

Goeree, J.K., Großer, J., 2003. Costly voting with correlated preferences- a rational analysis of the false consensus bias, Working paper, CREED, University of Amsterdam

Großer, J., Kugler, T., 2002. The impact of group size uncertainty on voter turnout: an experimental study, Working paper, CREED, University of Amsterdam

Großer, J., Ule, A., 2002. Trust, reciprocity, and social preferences, Working paper, CREED, University of Amsterdam

Isaac, R.M., Walker, J.M., 1988. Group size effects in public goods provision: the voluntary contribution mechanism, *Quarterly Journal of Economics* 103, 179-200

Kelley, H.H., Stahelski, A.J., 1970. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology* 16, 66-91

Keser, C., Van Winden, F., 2000. Conditional cooperation and voluntary contributions to public goods, *Scandinavian Journal of Economics* 102, 23-39

Kuon, B., 1994. Two person bargaining experiments with incomplete information, Berlin, Heidelberg, New York: Springer Lecture Notes in Economics and Mathematical Systems, No. 412

Ledyard, J.O., 1995. in: Kagel, J.H., Roth, A.E. (eds), Handbook of experimental economics, Princeton, New Jersey: Princeton University Press

Liebrand, W.B.G., 1984. The effect of social motives, communication and group size on behaviour in an n-person multi stage mixed motive game, *European Journal of Social Psychology* 14, 239-64

Milgrom, P., Roberts, J., 1991. Adaptive and sophisticated learning in normal form games, *Games and Economic Behavior* 3, 82-100

Mitzkewitz, M., Nagel, R., 1993. Experimental results on ultimatum games with incomplete information, *International Journal of Game Theory* 22, 171-198

Murphy, A.H., Winkler, R.L., 1970. Scoring rules in probability assessment and evaluation, *Acta Psychologica* 34, 273-286

Nagel, R., 1995. Unraveling in guessing games: an experimental study, *American Economic Review* 85, 1313-1326

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs, *Econometrica* 70, 971-1005

Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations, and voluntary contributions in public goods, *Economic Journal* 106, 817-845

Palfrey, T.R., Prisbrey, J.E., 1996. Altruism, reputation and noise in linear public goods experiments, *Journal of Public Economics* 61, 409-427

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

Rabin, M., 1993. Incorporating fairness into game theory and economics, *American Economic Review* 83, 1281-1302

Ross, L., Greene, D., House, P., 1977. The ‘false consensus effect’: an egocentric bias in social perception and attribution processes. *Journal of Experimental Psychology* 13, 297-301

Roth, A.E., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term, *Games and Economic Behavior* 8, 164-212

Schram, A., 2000. Sorting out the seeking: The economics of individual motivations, *Public Choice* 103, 231-258

Sefton, M., Shupp, R., Walker, J., 2002. The effect of rewards and sanctions in provision of public goods, Working paper, CeDEx, University of Nottingham

Selten, R., 1998. Axiomatic characterization of the quadratic scoring rule, *Experimental Economics* 1, 43-62

Selten, R., Buchta, J., 1998. Experimental sealed bid first-price auctions with directly observed bid functions, in Games and human behavior: Essays in honor of Amnon Rapoport. Mahwah, NJ: Lawrence Erlbaum Associates, 105-142

Selten, R., Mitzkewitz, M., Uhlich, G.R., 1997. Duopoly strategies programmed by experienced players, *Econometrica* 65, 517-555

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoner's dilemma supergames - a learning theory approach, *Journal of Economic Behavior and Organization* 7, 47-70

Siegel, S., Castellan, Jr., N.J., 1988. Nonparametric statistics for the behavioral sciences, New York: McGraw-Hill

Solà, C., 2002. The sequential prisoner's dilemma: reciprocity and group size effects, unpublished manuscript

Weimann, J., 1994. Individual behaviour in a free riding experiment, *Journal of Public Economics* 54, 185-200