

No. 2008–33

**CONDITIONAL COOPERATION: DISENTANGLING
STRATEGIC FROM NON-STRATEGIC MOTIVATIONS**

By Ernesto Reuben, Sigrid Suetens

March 2008

ISSN 0924-7815

Conditional cooperation: Disentangling strategic from non-strategic motivations*

Ernesto Reuben[†] and Sigrid Suetens[‡]

Abstract

We use a novel experimental design to examine the role of reputational concerns in explaining conditional cooperation in social dilemmas. By using the strategy method in a repeated sequential prisoners' dilemma in which the probabilistic end is known, we can distinguish between strategically and non-strategically motivated cooperation. Second movers who are strong reciprocators ought to conditionally cooperate with first movers irrespective of whether the game continues or not. In contrast, strategically motivated second movers conditionally cooperate only if the game continues and they otherwise defect. Experimental results, with two different subject pools, indicate reputation building is used around 30% of the time, which accounts for between 50% and 75% of all realized cooperative actions. The percentage of strong reciprocators varied between 6% to 23%.

This version: March 2008

JEL Codes: C91, D01, D74

Keywords: cooperation, reputation building, strong reciprocity, repeated prisoners' dilemma

*We would like to thank Simon Gächter, Mark Isaac, Jan Potters, Arno Riedl, and seminar participants from the University of Nottingham and Tilburg University and participants of the 2007 International ESA Meeting for useful comments and suggestions.

[†]Northwestern University, e-mail: ereuben@northwestern.edu

[‡]Tilburg University, e-mail: s.suetens@uvt.nl

1 Introduction

A large body of literature has demonstrated that even in one-shot settings a significant number of people are willing to cooperate, or at least conditionally cooperate, in social dilemmas (see e.g., Dawes and Thaler, 1988; Davis and Holt, 1993; Fischbacher et al., 2001). A number of authors argue that this is due to individuals caring about fairness (Rabin, 1993) and has motivated the development of various models of social preferences (for an overview see Fehr and Schmidt, 2006). It has also been conjectured that this is the result of evolutionary forces which have lead humans to exhibit so-called *strong reciprocity*, that is, reciprocity irrespective of potential future interaction (Gintis, 2000; Fehr et al., 2002). An important insight from these models is that, even if small, a fraction of fair-minded individuals can have a considerable effect on aggregate outcomes (Fehr and Gächter, 2000). One such case is a situation with repeated interaction where the presence of a few strong reciprocators gives incentives to rational own-payoff maximizing actors to build a reputation through strategic cooperation (Kreps et al., 1982).

In this paper, we run an experiment to examine the role of strategic and non-strategic concerns in explaining cooperation. We implement a novel within-subjects design that allows us to unambiguously distinguish to what extent individuals cooperate due to strong reciprocity and to what extent due to reputation building.

Various experiments have shown that individuals display higher rates of cooperation in social dilemma games where they repeatedly interact with the same subjects, so-called partners matching, than in games where they repeatedly interact with different individuals, strangers matching.¹ Furthermore, a common observation in partners treatments is a sharp decline in cooperation at the end of the game, particularly when it involves only two players (for a discussion of end-game effects see Selten and Stoecker, 1986). These patterns are consistent with both reputation building and strong reciprocity.² That is, the decrease in cooperation observed when switching from partners to strangers or when a game ends, can be due to two causes: first, own-payoff maximizing players who defect because they no longer have an incentive to cooperate, and second, strong reciprocators who defect because they expect others to stop cooperating. Note that this problem

¹This has been observed in, for example, public good games (Croson, 1996; Keser and van Winden, 2000), bribery games (Abbink, 2004), principal-agent games (Cochard and Willinger, 2005), trust games (Huck et al., 2006), and conflict games (Lacomba et al., 2008). For some exceptions see Andreoni and Croson (2003).

²Some authors argue that strangers matching has the additional disadvantage that some subjects might not fully understand that interaction consists of a series on one-shot encounters (Levitt and List, 2007).

cannot be easily resolved by eliciting beliefs³ as it is only on rare occasions in which reputation builders and strong reciprocators are expected to exhibit different belief-action combinations. Our design allows us to circumvent this problem as we elicit the subjects' strategies.

In our experiment, we can observe the extent to which cooperation is due to strategic considerations, namely reputation building, or to non-strategic ones like strong reciprocity. Subjects play a repeated prisoners' dilemma game with a probability of continuation. We use the contingent response or 'strategy' method developed by Selten (1967) to allow second movers to condition their decision on: (i) whether the period they are playing *is* or *is not* the final period of the game, and (ii) whether the first mover *cooperates* or *defects*. Eliciting the second movers' strategy for the stage game conditioned on whether they is future interaction or not, is enough to give us strong insights on the different motivations behind cooperative behavior.⁴ To facilitate reading, we will often refer to these conditional stage-game strategies simply as strategies.

In this game, second movers who cooperate when they know they are playing the last period of the game must be motivated (at least in part) by non-strategic reasons. In contrast, second movers who defect when they know it is the last period, but who cooperate when they know it is not, are clearly motivated by strategic considerations.⁵ Since we are interested in situations where cooperation by rational own-payoff maximizers is due to reputation building, the probability of continuation and the payoffs of the stage game are chosen such that trigger strategies such as those proposed by Friedman (1971) and Axelrod (1981) do *not* support mutual cooperation as an equilibrium. This reduces the motivations to cooperate to either strong reciprocity or reputation building combined with the expectation that there exist some strong reciprocators.

Our design also has the advantage that it clearly isolates the end-game effect. In games with a finite number of periods, differing abilities to perform backward induction (McKelvey and Palfrey, 1992; Katok et al., 2002) can distribute the end-game effect over the last few periods. In our experiment, cooperation in non-final and final periods are clearly differentiated. Furthermore, since we elicit end-game behavior over various periods, we can observe whether and how it evolves over time.

³Croson (2000) and Gächter and Renner (2006), for example, measure beliefs in social dilemma experiments.

⁴Note that we do not elicit the second movers' strategy for the whole game as this could require an infinite number of questions.

⁵The motivation behind the strategies of first movers cannot be easily discerned since, as in other experiments, they depend on their beliefs concerning the choice of second movers. For this reason we concentrate our analysis on the second movers' strategies.

There are other experiments that concentrate as well on reputation building or eliciting types of strategies. Early work in this area focused on testing the model of Kreps et al. (1982). These papers report a close match to the model’s predictions (Camerer and Weigelt, 1988; Andreoni and Miller, 1993; Sonnemans et al., 1999). However, they cannot easily differentiate between types of strategies. A few papers have concentrated on eliciting strategies in one-shot settings (Fischbacher et al., 2001; Fischbacher and Gächter, 2006), which permits them to identify strong reciprocators but not individuals who use reputation building. The study which comes closest to our paper is Muller et al. (2008), where subjects play a two-period linear public good game and submit a strategy for the whole two periods. This allows them to see how individuals condition their second period decision on the mean contribution of the first period. They observe a sharp decline of 45% in cooperation from one period to the next. However, they do not attempt to identify specific types of strategies. Our design improves on theirs for the purpose of identifying reputation building. As in their game groups consist of three players, there is a fairly continuous contribution decision, information is limited to mean group contributions, and strategies are conditioned only on mean contributions. This makes trigger strategies hard to implement and a precise reputation building strategy hard to define.

On the basis of three treatments, with two different subject pools, we find evidence for both reputation building and strong reciprocity with the former being the more common reason why subjects cooperate. Reputation building constitutes around 30% of the second movers’ strategies and accounts for 55% of cooperative outcomes. Strong reciprocity corresponds to between 6% and 23% of the second movers’ strategies and between 19% and 36% of cooperative outcomes, depending on the profitability of cooperation. Furthermore, with the use of two control treatments we find that applying the strategy method does not affect the second movers’ behavior.

The paper is organized as follows. In Section 2 we describe our experimental design in more detail. In Section 3 we present the results, and Section 4 concludes.

2 The Experiment

The game implemented in the experiment is a sequential prisoners’ dilemma with two players that is repeated with a probability of continuation equal to δ . In the stage game, a first mover $i = 1$, and a second mover $i = 2$, choose between cooperating C_i , or defecting D_i . If both players cooperate they both get π^C , if both defect they both get π^D , and if one defects and the other

cooperates the defector gets the temptation payoff π^T and the cooperator gets the sucker payoff π^S . As in all prisoners' dilemma games, payoffs are such that defecting is the dominant strategy: $\pi^T > \pi^C > \pi^D > \pi^S$, and mutual cooperation is the efficient outcome: $2\pi^C > \pi^T + \pi^S$.

By making their decision using the strategy method (Selten, 1967), players can condition their choice on whether the game will continue or end. In each period t a number x_t is drawn from a uniform distribution with support $[0, 1]$. If $x_t \leq \delta$ then t is not the final period of the game and the game continues. If $x_t > \delta$ then the game ends as soon as period t is played. Both first movers and second movers submit an action depending on whether $x_t \leq \delta$ or $x_t > \delta$. In addition, second movers submit an action also depending on whether the first mover cooperates or defects. Thus, whereas first movers submit an action in two cases: (i) $x_t \leq \delta$ and (ii) $x_t > \delta$, second movers submit an action in four cases: (i) C_1 and $x_t \leq \delta$, (ii) D_1 and $x_t \leq \delta$, (iii) C_1 and $x_t > \delta$, and (iv) D_1 and $x_t > \delta$. After both players submit their decisions, they learn the realization of x_t , the corresponding action of the other player, and their payoff.⁶

As is well known, full cooperation in repeated games with an unknown end can be achieved by rational individuals with the use of trigger strategies (Friedman, 1971). In fact, for a sufficiently high δ , any profile of play can be sustained as part of a subgame perfect equilibrium (Rubinstein, 1979; Fudenberg and Maskin, 1986). However, mutual cooperation of rational own-payoff maximizers is no longer supported if the continuation probability falls below the threshold $\delta^* = (\pi^T - \pi^C)/(\pi^T - \pi^D)$.⁷ This follows from the fact that both players always defect if $x_t > \delta$, and consequently, if $x_t \leq \delta$ the game is equivalent to one with an unknown end. In this case, mutual cooperation can be sustained by a trigger strategy only if $\delta \geq \delta^*$ in which case the second mover gets a higher payoff by cooperating than by defecting.⁸ Since we set $\delta < \delta^*$ in all our treatments, we can conclude that observed strategies supporting mutual cooperation cannot be due to rational own-payoff maximizers who think all others are alike.⁹

⁶Note that players are only informed of the other player's action for the *actual realization* of x_t . Thus if $x_t \leq \delta$, the second mover learns whether the first mover chose C_1 or D_1 when the game continues but does not learn the first mover's choice if t had been the last period of the game.

⁷Note that we are assuming no time discounting in the experiment, which we think is reasonable since the time interval between periods is very short and subjects are not paid until the end.

⁸A first mover does not have an incentive to deviate from an equilibrium with mutual cooperation since, given that a second mover would imitate defection, he can never attain the high temptation payoff.

⁹It is still possible to support unilateral cooperation if $\delta \geq (\pi^D - \pi^S)/(\pi^T - \pi^D)$ (see Stahl, 1991). However, these equilibria require a high degree of coordination which is hard to achieve in the laboratory. We report whether there is evidence for these type of strategies in footnote 12.

However, if subjects believe they might play with a strong reciprocator, the fact that in each period t they can condition their choice on whether t is the final period or not opens the possibility for cooperation through reputation building as in Kreps et al. (1982). This is driven by two cases in which cooperation in the last period is possible. First, if the first mover is a strong reciprocator so that he cooperates in period t as long as the other player cooperated in $t - 1$, then a rational second mover who maximizes own payoff gets a higher payoff by defecting in the last period instead of in period t . This way, she gets the temptation payoff and also enjoys the high earnings of mutual cooperation during the game. Second, if a first mover faces a sufficiently high probability that the second mover is a strong reciprocator, he is better off by cooperating even in the last period. If this is the case, an own-payoff maximizing second mover can mimic a strong reciprocator up to the last period and then defect. This again gives her the temptation payoff and the benefits of cooperation. Denoting α_1 and α_2 as the fraction of first and second movers who are strong reciprocators, the two conditions under which mutual cooperation can be sustained in equilibrium are given below (calculations are provided in Appendix A)

$$(a) \alpha_1 \geq \frac{\pi^T - \pi^C - \delta(\pi^T - \pi^D)}{\pi^T - \pi^D - \delta(\pi^T - \pi^D)} \quad \text{or} \quad (b) \alpha_2 \geq \frac{\pi^D - \pi^S}{\pi^C - \pi^S}.$$

As previously mentioned, our design allows us to observe the strategies used by second movers. We call strong reciprocity the strategy that consists of conditionally cooperating with the first mover irrespective of whether they are playing the last period or not. Reputation building corresponds to the strategy of conditionally cooperating only if they are not playing the last period and defecting otherwise.¹⁰ These and other strategies of interest are described in Table 1. For example, it is also informative to know the prevalence of second movers who choose the strategy of unconditional defection. We should note, however, that we cannot differentiate between second movers who are strategic but defect because they play a defection equilibrium and second movers who defect for non-strategic reasons.

The experiment consists of two sets of treatments, each set run with a different subject pool. In all cases subjects were randomly paired and assigned to either the first or the second mover role. Roles and pairs were kept constant throughout the experiment giving us one independent

¹⁰A general worry with playing games with an unknown end is that subjects know the experiment cannot last for an extremely long time. Thus, they might discount future interactions at a rate that is lower than δ . However, for the purpose of our experiment, this can at most induce a very small decrease in the frequency of strategically motivated cooperation (subjects who are not playing strategically and those who are already defecting unconditionally are not affected by more discounting). Thus, if we do observe reputation building, it is still clearly strategic behavior in line with Kreps et al. (1982).

Table 1: Strategies of second movers

Strategy	<i>Not the last period</i>		<i>Last period</i>	
	First mover	First mover	First mover	First mover
	Cooperates	Defects	Cooperates	Defects
Reputation building	Cooperate	Defect	Defect	Defect
Strong reciprocity	Cooperate	Defect	Cooperate	Defect
Unconditional defection	Defect	Defect	Defect	Defect
Unconditional cooperation	Cooperate	Cooperate	Cooperate	Cooperate

observation per second mover. The commonly used experimental procedures for anonymity, incentivized payments, and neutrally worded instructions were implemented. Next, we briefly describe each of the treatments. A detailed description of the experimental procedures, including all treatment parameters and instructions are available in Appendix B.

The first set of treatments was run in CentERlab at Tilburg University. It consists of one treatment where we implement the game described above, which we refer to as *Tilburg*, and two control treatments. In all cases, we chose a continuation probability of $\delta = 0.60$ and the payoffs of the stage game were selected so that $\delta^* = 0.61$. The value of x_t was generated by the computer, and in order keep the three treatments as comparable as possible, the random sequence of each pair in the Tilburg treatment was used in one pair in each control treatment. The purpose of this treatment is to identify the various strategies used by second movers, whereas the purpose of the control treatments is to test the validity of the strategy method.

In principle, it is possible that the use of the strategy method induces a change in behavior. In the experimental literature there is yet no consensus if this is indeed the case. Various authors report no significant differences in, for example, sequential dictator games (Cason and Mui, 1998), trust games (Vyrastekova and Onderstal, 2005), and, closest to our study, chicken and prisoners' dilemma games (Brandts and Charness, 2000). However, there are also studies that do find differences in behavior. For instance, some authors have found less punishment with the use of the strategy method than without it (Brosig et al., 2003; Falk et al., 2005). For this reason, we use two control treatments to ensure that the strategy method does not affect behavior in our setting. In the first control treatment, *Control I*, subjects play the same game of Tilburg except that they no longer submit a decision for both final and non-final periods. In other words, subjects are *first* told whether $x_t \leq \delta$ or $x_t > \delta$ and then they make their decision. Note that

second movers still submit separate choices depending on whether the first mover cooperates or not. By comparing choices, and, in particular, the stage-game strategy of second movers, between this control and Tilburg we can test whether behavior is affected by conditioning decisions on whether it is the last period or not. In the second control, *Control II*, we again implement the Tilburg game but this time without the use of the strategy method. In other words, subjects are told whether $x_t \leq \delta$ or $x_t > \delta$ before they decide and second movers learn what the first mover did before they make their choice. By comparing behavior between the two control treatments we can test whether the decisions of second movers are affected by the possibility to condition their choice on whether the first mover cooperated or defected.

In the second set of treatments, we implement the game with the full use of the strategy method but this time we vary the value of δ^* between treatments. We used a coin toss to determine whether the game continues or ends, which makes the continuation probability equal to $\delta = 0.50$. These treatments are run for two reasons. The first is that it is interesting to see how the elicited strategies change as one varies the benefit of cooperation. It is well established that subjects are more willing to cooperate in non-strategic settings if it is more profitable to do so (e.g., Goeree et al., 2002). In our experiment, an increase in non-strategic cooperation, if foreseen, ought to also increase the frequency of reputation building. The second reason is that, although the evidence is mixed, there is some worry that the use of economics and business students biases results in experiments involving cooperative behavior (e.g., Marwell and Ames, 1981; Engelmann and Strobel, 2006; Fehr et al., 2006). Since Tilburg's subject pool consists mainly of such students, we ran these two treatments in Northwestern University and excluded students who study economics or a related field. In the first of these two treatments, *Northwestern High*, we use a high payoff for mutual cooperation such that $\delta^* = 0.56$. In the second treatment, *Northwestern Low*, we use a low payoff of mutual cooperation (keeping everything else constant) such that $\delta^* = 0.72$.

In summary, we ran three treatments with which we can observe the prevalence of strategies involving different motivations for cooperation: Tilburg, Northwestern High, and Northwestern Low. Furthermore, comparing the Northwestern treatments to each other allows us to see how the use of strategies change with a change in the profitability of cooperation, and lastly, using Tilburg and the two control treatments we can identify any behavioral changes induced by the strategy method. In the next section we present the results.

3 Results

In this section we present the experimental results. The main results are discussed in subsection 3.1, where we focus on the second movers' cooperation rates and strategies. End-game effects are discussed in subsection 3.2. In subsection 3.3 we show—with the use of the control treatments—that the strategy method does not affect behavior in this game. Throughout this section, we test differences in frequencies using likelihood-ratio χ^2 -tests. Since this is our most common test, in order to avoid unnecessary repetition, we simply report the resulting p -values when it is used. Moreover, since different second movers played the game a different number of periods, all tests, figures, and tables based on aggregate data across periods are adjusted by the inverse number of periods played. This way, each second mover, as one independent observation, receives an equal weight.¹¹

3.1 Disentangling conditional cooperation

In this section we give an overview of second movers' cooperation rates and we disentangle their motivations for conditionally cooperating. Overall, if we look at realized outcomes, the second movers' cooperation rate equals 0.251 in Tilburg, 0.395 in Northwestern High, and 0.147 in Northwestern Low. We find it to be significantly higher in Northwestern High vis-à-vis Northwestern Low ($p = 0.021$), showing that the occurrence of mutual cooperation is sensitive to its profitability.

Conditional cooperation

Second movers conditionally cooperate: their cooperation rates are much higher when first movers cooperate than when first movers defect. In Tilburg, if the first mover cooperates, the second movers' cooperation rate is 0.366, and if he defects, it is 0.084. In Northwestern High, the respective rates are 0.570 and 0.074, and in Northwestern Low 0.206 and 0.015. In all treatments the difference in cooperation rates is statistically significant (Wilcoxon signed-rank tests, $p < 0.004$).

In what follows, we take a detailed look at the behavioral patterns by analyzing the actual strategies chosen by second movers. This allows us to see the amount of conditional cooperation that is due to reputation building and the amount that is due to other motivations. Figure 1

¹¹We should note, however, that none of the reported significant or not significant results change if we use unweighted data or if we concentrate on the first period (which was played once by all second movers).

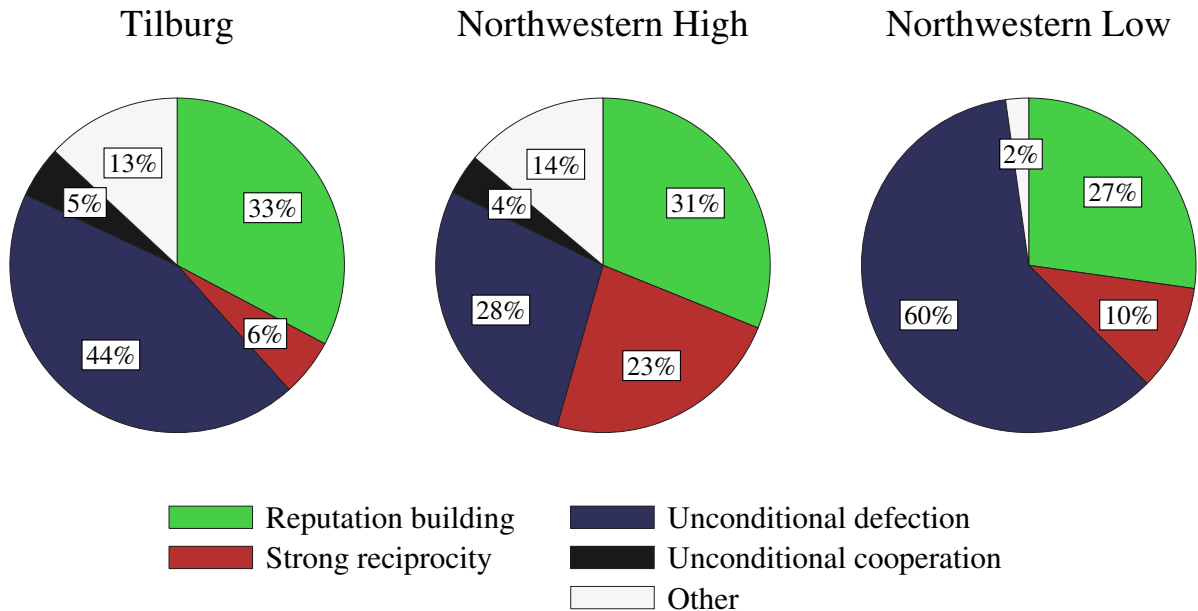


Figure 1: Frequency of strategies

Note: The pie charts show, for each treatment and across all periods, the frequency of strategies used by second movers classified according to Table 1. Strategies are weighted by the inverse number of periods played by each subject.

presents the distribution of second movers’ strategies in Tilburg and the two Northwestern treatments using the classification of Table 1. Overall, unconditional defection is the most common strategy. It is chosen from 28% to 60% of the time. However, this still leaves considerable space for strategies that are predicted only in the presence of individuals with social preferences.¹²

In all treatments, the most frequent strategy that includes some cooperation is reputation building. It accounts for around 30% of all strategies. This is even the case in Northwestern Low where the continuation probability is well below the threshold required for mutual cooperation. The third most common strategy is strong reciprocity, whose frequency varies between 6% and 23%. Unconditional cooperation is used less than 5% of the time and other strategies in between 2% and 14%.¹³

¹²One can expect some degree of unilateral cooperation if subjects play a correlated equilibrium (Stahl, 1991). If this is the case, some second movers must choose the strategy: D_2 if C_1 and $x_t \leq \delta$, C_2 if D_1 and $x_t \leq \delta$, and D_2 whenever $x_t > \delta$. We found only one second mover who chose this strategy once. Hence, we don’t find support for these type of equilibria, which can be due to the lack of a suitable coordination device.

¹³Two strategies account for around 70% of those in the category ‘other’. The first is always defecting if it is not the last period and conditionally cooperating if it is. The second is always cooperating if it is not the last

Reputation building explains most of the second movers' cooperation. If we look at the fraction of their realized cooperative actions that are due to reputation building, we find it to be 57% in Tilburg, 56% in Northwestern High, and 70% in Northwestern Low. In contrast, strong reciprocity accounts for 14% of the second movers' cooperation in Tilburg, 25% in Northwestern High, and 20% in Northwestern Low.

Another way of observing the importance of reputation building is to look at the second movers' realized stage-game strategies (i.e. the strategies we would have observed if subjects could not condition on whether it was the last period or not). We observe that conditional cooperation is relatively common: it accounts for 31% of all stage-game strategies in Tilburg, 52% in Northwestern High, and 21% in Northwestern Low. Of the stage-game strategies that imply conditional cooperation, in Tilburg 64% are due to subjects playing reputation building, in Northwestern High it is 48%, and in Northwestern Low 64%. The respective percentages for strong reciprocity are 12%, 44%, and 36%.

In Figure 1 one can also see a noticeable change in the frequency of strategies across the two Northwestern treatments (the distributions are significantly different, $p = 0.043$). This reveals that the observed decrease in cooperation rates in Northwestern Low vis-à-vis Northwestern High is driven by a sharp increase in the frequency of unconditional defection at the expense of strong reciprocity, unconditional cooperation, and strategies under 'other'. Contrary to what we expected, the frequency of reputation building is almost identical.

Stability of strategies

Next, we briefly analyze the stability over time of the distribution of strategies. In all treatments, the frequencies of strategies do not change considerably across periods. If we test, in each treatment, for equality of distributions between the first three periods we find no significant differences ($p > 0.570$). This tells us that the relative influence of reputation building and strong reciprocity vary little with repetition. Next, we check whether this relative stability is hiding substantial changes at the individual level.

In order to analyze the stability of strategies within each subject, we take a look at how often subjects choose the same strategy. Specifically, we calculate the probability that a second mover picks in period t the same strategy that she picked in period $t - 1$. Overall, second movers pick the same strategy for consecutive periods 64% of the time in Tilburg, 72% in Northwestern High, and 86% in Northwestern Low. The stability of individual strategies can be seen in Table 2 where

period and conditionally cooperating if it is.

Table 2: Stability of strategies

Note: The table shows, for each strategy and treatment, the probability that a second mover picks in period t the same strategy that she picked in period $t - 1$. Probabilities are weighted by the inverse of the number of periods played by each subject.

	Reputation building	Strong reciprocity	Unconditional defection	Unconditional cooperation	Other
Tilburg	62%	100%	77%	–	0%
Northwestern High	84%	88%	68%	80%	35%
Northwestern Low	88%	80%	88%	–	0%
All treatments	77%	89%	78%	80%	21%

this probability is calculated separately for each strategy and treatment. From the table, one can see that the three main strategies: reputation building, strong reciprocity, and unconditional defection are quite stable. A second mover who chooses one of these strategies has around an 80% chance of choosing the same strategy in the next period. In comparison, the strategies that fall within ‘other’ are considerably less robust. In most cases, these strategies are chosen for only one period at a time.¹⁴ With respect to the motivation of second movers to switch strategies, besides choosing a strategy under ‘other’, we do not find that either the previously chosen strategy or the outcome in the stage game has a significant effect.¹⁵ In summary, strategies are fairly stable both across periods and within subjects. A majority of subjects consistently chose one of the strategies in Table 1, while other strategies are chosen less consistently.

¹⁴Using binomial probability tests and the null hypothesis that the probability of choosing the same strategy in period t and $t - 1$ is less than 50% (i.e. a subject is more likely to switch than to choose the same strategy), we can (weakly) reject it in all treatments for reputation building ($p < 0.001$), unconditional defection ($p < 0.001$), and strong reciprocity ($p < 0.056$). For unconditional cooperation it is rejected in Northwestern High ($p = 0.032$). Treating strategies under ‘other’ as a group, we cannot reject the null in any treatment ($p > 0.998$).

¹⁵We ran a probit regression with a binary variable indicating whether a subject changes strategy from period t to $t + 1$ as the dependent variable. We used the following independent variables: dummy variables for the strategy chosen in t , dummy variables for the realized outcome in t , treatment indicator variables, and the period number. We find that choosing a strategy from ‘other’ in period t is associated with a 32% higher probability of choosing a different strategy in $t + 1$ ($p = 0.001$, using White’s heteroscedasticity consistent covariance matrix estimator to cluster on each subject). However, we find no other significant effect. We get the same result if we run a separate regression for each strategy or for each treatment.

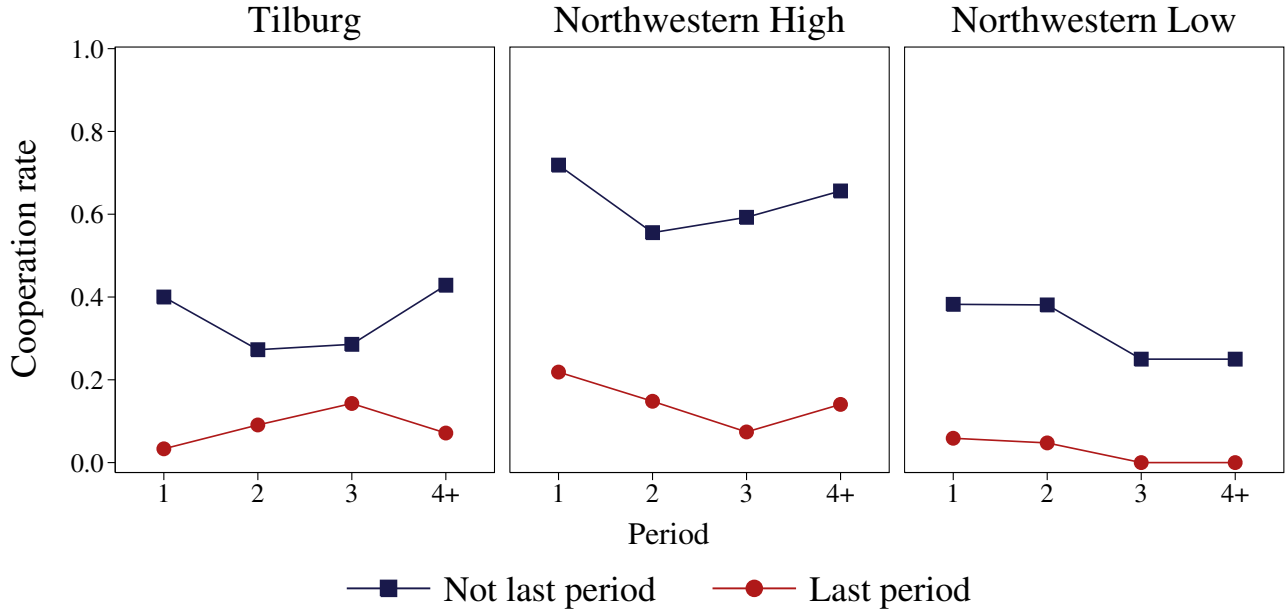


Figure 2: Mutual Cooperation Rates

Note: The figures show, for each treatment, the rate of mutual cooperation per period using the subjects' strategies to determine what their action would be if a given period is the last period or not.

3.2 End-game effects

If we look at mutual cooperation in final periods and non-final periods then a clear end-game effect is observed.¹⁶ The realized rate of mutual cooperation depending on whether it is the last period or not is 0.363 and 0.033 in Tilburg, 0.590 and 0.000 in Northwestern High, and 0.333 and 0.000 in Northwestern Low. The sharp difference between the two indicates a strong effect of removing the strategic incentive to cooperate.

A clearer picture can be observed in Figure 2. In it, we show the rate of mutual cooperation in each period $t \in \{1, 2, 3, 4+\}$ ¹⁷ using the subjects' strategies to determine what their action would be if t is the final period or not. As can be seen, mutual cooperation is higher when subjects are not playing the last period of the game. This can be statistically confirmed for each treatment with Wilcoxon signed-rank tests ($p < 0.001$). Notably, a significant difference between non-final and final periods is already present in the first period (Wilcoxon signed-rank tests, $p < 0.001$). This indicates there is already a clear difference in behavior between the two

¹⁶All results in this subsection also hold if we concentrate on the second movers' cooperation rates.

¹⁷After period 3 the number of observations decreases considerably in some treatments.

situations before subjects have the opportunity to interact (as in Keser and van Winden, 2000). We should also note that we do not find time trends in the rates of mutual cooperation for any of the treatments, neither for non-final periods nor for final periods.¹⁸

Next, we look at the extent to which end-game effects are due to first movers switching from cooperation to defection. A large fraction of the first movers who cooperate do so only in non-final periods. In other words, their stage-game strategy consists of cooperating if it is not the last period and defecting otherwise. In Tilburg, 73% of first movers submit this strategy, in Northwestern High it is 41% and in Northwestern Low 67%. This is interesting as in this game, first movers have an incentive to defect in the last period only if they anticipate that a large fraction of second movers are cooperating strategically. Given that defection by first movers induces defection among strongly reciprocal second movers, the first movers' expectation of reputation building is responsible for part of the observed drop in cooperation.

An insightful exercise is to isolate the effect of the first movers' expectations in order to compare the effect of strategic and non-strategic second movers on cooperation in final periods. By using the subjects' strategies, we can measure the increase in the rate of mutual cooperation in last period if first movers switch from their final period to their action in non-final periods. In Tilburg and Northwestern Low, the rate of mutual cooperation increases only slightly. In the former it changes from 0.033 to 0.067 and in the latter from 0.000 to 0.059. In Northwestern High there is a considerable difference as it changes from 0.000 to 0.344. This reveals that the extent to which end-effects are driven by reputation building by second movers or by the first movers' expectation of reputation building depend on the profitability of cooperation (and possibly the subject pool). This can be important as, for example, manipulating first movers' expectations would have an effect on cooperation in final periods when the profitability of cooperation is high but less so when it is low.

3.3 Control treatments

To ensure that the use of the strategy or contingent response method does not result in different behavior than the direct response method, we use this subsection to compare behavior in Tilburg and the two control treatments.

¹⁸In all treatments, Spearman's rank correlation coefficients between cooperation rates and periods are not significantly different from zero ($p > 0.178$).

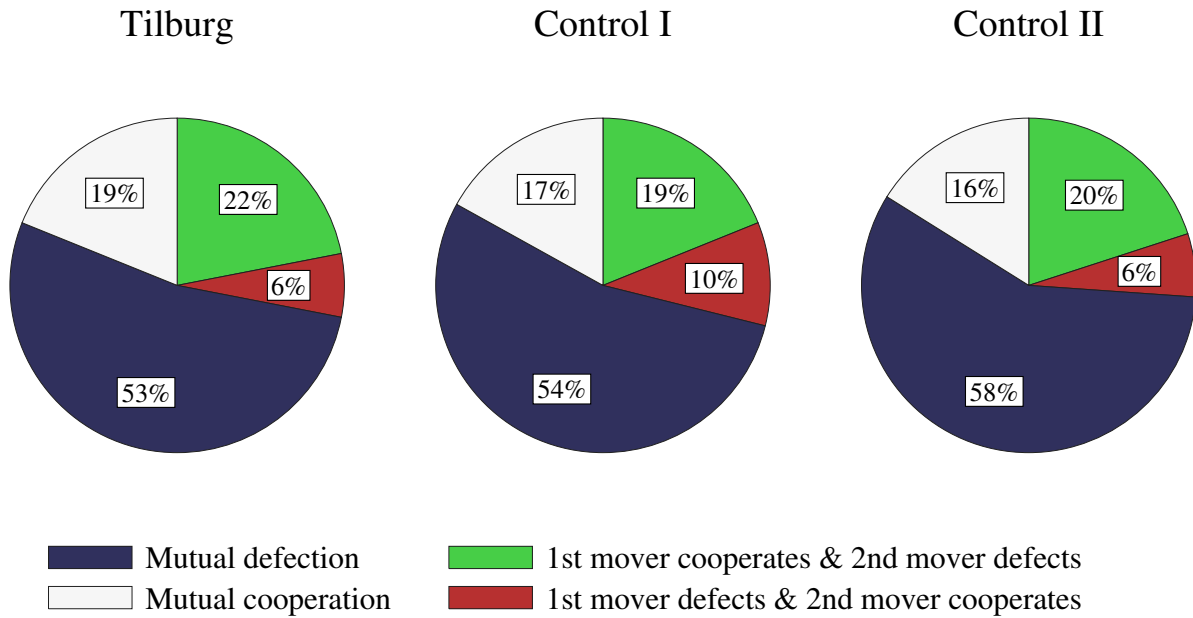


Figure 3: Comparing outcomes with control treatments

Note: The pie charts show the frequency of each of the four possible outcomes in Tilburg and the two control treatments. Outcomes are weighted by the inverse of the number of periods played by each pair.

Figure 3 gives an overview of the distribution of *realized* outcomes in the three treatments. It is clear that outcomes are highly similar. This is corroborated if we test for equality of distributions across the three treatments ($p = 0.992$).¹⁹

Next, we compare Tilburg with Control I to determine whether the elicited stage-game strategies change when they are conditioned on whether it is the last period or not. One could worry that conditioning on the final period might trigger more strategic thinking than otherwise, and therefore, if the strategy method is used, there could be less conditional cooperation in the last period. However, we do not find this to be the case. As is shown in the top part of Table 3, the frequencies of stage-game strategies for both treatments for non-final and final periods are very similar (distributions are not significantly different: $p = 0.642$ for non-final periods and $p = 0.581$ for final periods). Furthermore, in Control I we do not see a higher frequency of conditional cooperation in final periods.²⁰

¹⁹We do not get statistical significance either if we do pairwise comparisons between treatments ($p > 0.910$), or if we compare separately across treatments the frequency of each strategy ($p > 0.745$).

²⁰There are no significant differences if we compare separately the frequency of each stage-game strategy across

Table 3: Comparing strategies with control treatments

Note: The top half of the table shows, for Tilburg and the control treatments, the distribution of the second movers' stage-game strategies depending on whether it is the last period or not. The bottom half shows the second movers' cooperation rates depending on the first mover's choice (actual cooperation rates for Control II and cooperation rates implied by the stage-game strategies in the other two treatments). Strategies and rates are weighted by the inverse of the number of periods played by each subject.

	<i>if not last period</i>			<i>if last period</i>		
	Tilburg	Control I	Control II	Tilburg	Control I	Control II
<i>Comparing stage-game strategies</i>						
Always cooperate	9%	2%		5%	10%	
Cooperate if first mover cooperates	39%	50%		16%	17%	
Cooperate if first mover defects	3%	5%		1%	7%	
Always defect	49%	44%		78%	67%	
<i>Comparing cooperation rates</i>						
When the first mover cooperates	48%	52%	54%	21%	27%	13%
When the first mover defects	12%	7%	8%	6%	17%	14%

Lastly, we compare cooperation rates between Control II and the other two treatments to test whether second movers' choices are affected by conditioning them on the first movers' action. The actual cooperation rates in Control II and the ones implied by the strategies in the other treatments are seen in the bottom part of Table 3. We do not find statistically significant differences between the frequencies of the three treatments when running tests that compare separately non-final periods and final periods depending on whether the first mover cooperates or defects ($p > 0.428$). This also holds if we do pairwise tests between treatments ($p = 0.202$). Thus, we tentatively conclude that the strategy method is a valid technique for the elicitation of strategies in this game.

4 Conclusions

In this paper we provide evidence for the importance of reputation building for cooperation in social dilemmas. We report the results of a laboratory experiment where subjects play a repeated sequential prisoners' dilemma with a probabilistic end. Choices can be conditioned on treatments ($p > 0.268$).

whether the period of play is the final period or not. This design allows us to separate strategic from non-strategic cooperation. Since in our design mutual cooperation is not an equilibrium when all individuals are rational own-payoff maximizers, strategic cooperation is consistent with reputation building and the belief that some players cooperate for non-strategic reasons.

We find that subjects use the reputation building strategy around 30% of the time, which accounts for between 50% and 75% of all realized cooperative actions. We also find that the frequency of reputation building is largely unaffected by changes in the subject pool or by repetition. The other two commonly-used strategies are unconditional defection (used between 28% and 60% of the time) and strong reciprocity (between 6% and 23%). The latter accounts for between 19% and 36% of cooperative actions. We find the distribution of strategies to be fairly stable in time and within subjects.

The observed end-game effects are partly due to second movers playing strategically and partly due to the belief by first movers that second movers are playing strategically. Abstracting from first movers' beliefs, we find that end-game effects are robust to first-mover actions in Tilburg and Northwestern Low due to the relatively large fraction of cooperation that is due to reputation building by second movers. In Northwestern High, the end-game effect would be considerably smaller if first movers were to cooperate more often in the final period.

Lastly, we see a decline in the cooperation rate as the payoff from mutual cooperation decreases. The difference is driven by a decrease in the number of subjects cooperating non-strategically. This is consistent with subjects receiving utility from money and from complying with a 'fairness norm', as in the various models of social preferences (Fehr and Schmidt, 2006). However, it would be interesting to know precisely who changes strategies. Do subjects who play as strong reciprocators switch straight to defection or do they switch to reputation building? Knowing this would give us a better understanding of the various motivations behind the subjects' actions. In this sense, running an experiment where the profitability of cooperation is varied within subjects would be an interesting line of future research.

We should note, however, that if one assumes rational expectations some of our results are not consistent with reputation building as modeled in Kreps et al. (1982). Although the fraction of subjects playing the strong reciprocity strategy (or other strategies that produce cooperation in the final period) is large enough in Tilburg and Northwestern High to support reputation building, it is too small in Northwestern Low. According to the theoretical model, for reputation building to be profitable the percentage of strong reciprocators in the population ought to be at least 31%, which is well above the one observed. In this case, if subjects in infinitely repeated

games move toward the supported equilibria when the super-game is repeated many times (as in Dal Bo and Fréchet, 2007), first movers with enough experience might also stop cooperating in our game. However, another possibility is that individuals might build a reputation due to evolved affective responses (Fessler and Haley, 2003), and thus, in some cases, reputation building might still occur even when it is not profit-maximizing.

A Equilibrium Calculations

For the reported equilibria, we rely on strong reciprocators applying a grim trigger strategy. That is, a deviation from the mutual cooperation equilibrium implies defection until the game ends (as in Kreps et al., 1982). Since this is the harshest punishment for deviating, it gives the most favorable conditions for mutual cooperation to arise. We consider two types of players, rational own-payoff maximizers and strong reciprocators.

Note that, irrespective of their type, first movers will never deviate from a mutual cooperation equilibrium as such an action would be followed with defection by the second mover, which implies first movers cannot get the higher temptation payoff. Equilibria are thus driven by the second movers' incentive to deviate.

Reputation building equilibria require the possibility that there is cooperation by first movers in the last period. This can happen for two reasons: (i) the fraction of strong reciprocators among second movers is high enough that rational own-payoff maximizing first movers are better off cooperating in the last period, or (ii) the fraction of strong reciprocators among first movers is high enough that a second mover has a very high chance that if the last period is reached without a deviation, the first mover cooperates.

Rational own-payoff maximizing first movers cooperate in the last period if no deviation has occurred and the probability that the second mover is a strong reciprocator is high enough. That is, $\alpha_2\pi^C + (1 - \alpha_2)\pi^S \geq \pi^D$, which, solving for α_2 gives the condition

$$\alpha_2 \geq \frac{\pi^D - \pi^S}{\pi^C - \pi^S}. \quad (1)$$

In this case, all first movers cooperate in the last period irrespective of their type, and rational own-payoff maximizing second movers prefer to cooperate instead of defecting in period $t = 0$ if:

$$\begin{aligned} \pi^T + \sum_{t=1}^{\infty} \delta^t \pi^C &> (1 - \delta)\pi^T + \delta\pi^T + \sum_{t=1}^{\infty} \delta^t \pi^D \\ \pi^T + \frac{\delta}{1 - \delta} \pi^C &> (1 - \delta)\pi^T + \delta\pi^T + \frac{\delta}{1 - \delta} \pi^D \end{aligned} \quad (2)$$

which is true if $\pi^C > \pi^D$. Thus, if (1) holds, mutual cooperation is an equilibrium until the last period of the game.

Even if rational own-payoff maximizing first movers defect in the last period, for instance because (1) does not hold, mutual cooperation in non-final periods is still possible if the fraction of strong reciprocators among first movers is high enough. Given that α_1 is the fraction of strong reciprocators among first movers who do not defect in the last period if no previous defection has occurred, and that $1 - \alpha_1$ is the fraction of rational own-payoff maximizing first movers who do defect in the last period, rational own-payoff maximizing second movers prefer to cooperate instead of defecting in period $t = 0$ if:

$$\begin{aligned} \alpha_1 \pi^T + (1 - \alpha_1) \pi^D + \sum_{t=1}^{\infty} \delta^t \pi^C &> (1 - \delta)(\alpha_1 \pi^T + (1 - \alpha_1) \pi^D) + \delta \pi^T + \sum_{t=1}^{\infty} \delta^t \pi^D \\ \alpha_1 \pi^T + (1 - \alpha_1) \pi^D + \frac{\delta}{1 - \delta} \pi^C &> (1 - \delta)(\alpha_1 \pi^T + (1 - \alpha_1) \pi^D) + \delta \pi^T + \frac{\delta}{1 - \delta} \pi^D \end{aligned} \quad (3)$$

which, solving for α_1 , gives the condition

$$\alpha_1 \geq \frac{\pi^T - \pi^C - \delta(\pi^T - \pi^D)}{\pi^T - \pi^D - \delta(\pi^T - \pi^D)}. \quad (4)$$

B Supplementary Material

B.1 Experimental procedures

The 45-minute experiment was programmed and conducted with z-Tree (Fischbacher, 2007). Subjects were recruited through online recruitment systems. In total, 312 subjects participated in the experiment. Each subject played in only one treatment.

The details of the five treatments are summarized in Table B1. As can be seen, the Tilburg treatment and the two controls were run with students of economics and business administration in the CentERlab in Tilburg University. The Northwestern High and Low treatments were run with students from neither economics nor business in the laboratory of the Kellogg School of Management in Northwestern University.²¹ Average earnings in Tilburg treatments were €9.38 and \$11.20 in Northwestern treatments (10 points equaled €1.50 in Tilburg and \$2.00 in Northwestern, amounts exclude a showup fee of \$6 in Northwestern).

²¹In Northwestern, the areas of study were: journalism/communication (21%), engineering (21%), biology/chemistry/physics (16%), anthropology/political science/sociology (15%), history/languages/philosophy, (12%), arts (8%), and others (8%).

Table B1: Experimental treatments

	Northwestern High	Northwestern Low	Tilburg	Control I	Control II
<i>Parameters</i>					
Mutual defection π^D	15	15	15	15	15
Unilateral defection π^T	33	33	33	33	33
Unilateral cooperation π^S	6	6	10	10	10
Mutual cooperation π^C	23	20	22	22	22
Probability of continuation δ	0.50	0.50	0.60	0.60	0.60
Threshold for cooperation δ^*	0.56	0.72	0.61	0.61	0.61
Threshold for cooperation α_1	0.11	0.44	0.03	0.03	0.03
Threshold for cooperation α_2	0.53	0.64	0.42	0.42	0.42
<i>Characteristics</i>					
Location	Northwestern	Northwestern	Tilburg	Tilburg	Tilburg
Field of study	Not economics	Not economics	Economics	Economics	Economics
Number of second movers	32	34	30	30	30

After their arrival, subjects drew a card to be randomly assigned to a seat in the laboratory, and consequently to a role and a treatment. Once everyone was seated, subjects were given the instructions for the experiment. The instructions are written with neutral language and clearly explain the decision of first and second movers, including an explanation of the strategy method (see below). Thereafter, roles were revealed and subjects had to answer a few control questions to corroborate their understanding of the game. Next, they played the game until the random draw indicated that it ended. Once the game finished, subjects answered a debriefing questionnaire after which they were paid in cash and dismissed.

Experimental procedures were identical in both locations with one exception. Namely, the method used to randomly determine whether the experiment continued or not. In Tilburg we used the computer to determine for each group and in each period whether the game continued or not. Then, in order to facilitate treatment comparisons, we used the same sequence of random numbers in the control treatments. For the Northwestern treatments, we used a simpler random number generator: a coin toss. In this case, a toss of the coin determined whether the game

continued or ended for all groups in a given session. The average number of periods played in Tilburg equaled 2.73, in Northwestern Low and High it was 2.32 and 4.06, respectively.

B.2 Experimental Instructions

These are the instructions for second movers in the Northwestern High treatment. The instructions for first movers and those of other treatments are very similar and available from the authors upon request.

General

You are participating in an experiment on economic decision making and will be asked to make a number of decisions. If you follow the instructions carefully, you can earn money. At the end of the experiment, you will be paid your earnings in private and in cash.

You are not allowed to communicate with other participants. If you have a question, raise your hand and one of us will help you.

During the experiment your earnings will be expressed in points. Points will be converted to US dollars at the following rate: $10 \text{ points} = \$2.00$.

The experiment is *strictly anonymous*: that is, your identity will not be revealed to others and the identity of others will not be revealed to you.

In the experiment, participants will be randomly divided into groups of 2 participants. You will therefore be in a group with one other participant. The composition of the groups will *remain the same during the entire experiment*.

In each group, one participant will be randomly assigned to the *first mover* position. The other participant in the group will be in the *second mover* position. Your position as first or second mover will remain the same during the entire experiment.

Your decision in each period

The experiment is divided into periods. In each period, both the first and the second mover make a choice between option A and option B. The first mover makes his/her decision first. Thereafter the second mover makes his/her decision. The following table shows what the first and second movers earn (in points) depending on their choices:

Number of Periods

For each group, the number of periods of in the experiment is determined *randomly*. At the end of each period, we will throw a coin to determine whether that period was the last period of the experiment or whether the experiment continues (heads means the experiment continues and tails means the experiment ends). Thus, in every period the probability that the experiment

	first mover's earnings	second mover's earnings
both choose A	23	23
first mover chooses A and the second mover chooses B	6	33
first mover chooses B and the second mover chooses A	33	6
both choose B	15	15

continues is 50% and the probability that the experiment ends is 50%. Your total earnings in the experiment will equal the sum of earnings across all periods.

After each period, you will receive feedback concerning the decision of the other participant in your group and on your earnings.

The decision of the first mover

In each period, the first mover makes his/her decision in each of the two following situations:

- Do you choose A or B if the current period *is not* the final period (in other words the experiment proceeds to a next period)?
- Do you choose A or B if the current period *is* the final period (in other words the experiment does not proceed)?

If the result of the coin toss is that the experiment continues (heads), then earnings in that period will depend on the answer to the first question. If the result of the coin toss is that the experiment ends (tails), then earnings depend on the answer to the second question.

The decision of the second mover

In each period, the second mover makes his/her decision in each of the four following situations:

- *If the first mover chooses A:* Do you choose A or B if the current period *is not* the final period (in other words the experiment proceeds to a next period)?
- *If the first mover chooses A:* Do you choose A or B if the current period *is* the final period (in other words the experiment does not proceed)?
- *If the first mover chooses B:* Do you choose A or B if the current period *is not* the final period (in other words the experiment proceeds to a next period)?

- *If the first mover chooses B:* Do you choose A or B if the current period *is* the final period (in other words the experiment does not proceed)?

If the result of the coin toss is that the experiment continues (heads) and the first mover chooses A, earnings will depend on the answer to the first question. If the result of the coin toss is that the experiment ends (tails) and the first mover chooses A, earnings will depend on the answer to the second question. If the result of the coin toss is that the experiment continues (heads) and the first mover chooses B, earnings will depend on the answer to the third question. If the result of the coin toss is that the experiment ends (tails) and the first mover chooses B, earnings will depend on the answer to the fourth question.

References

- Abbink, K. (2004). Staff rotation as an anti-corruption policy: An experimental study. *European Journal of Political Economy*, 20:887–906.
- Andreoni, J. and Croson, R. (2003). Partners versus strangers: Random rematching in public goods experiments. In Plott, C. R. and Smith, V. L., editors, *Handbook of Experimental Economics Results*, volume forthcoming.
- Andreoni, J. and Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner’s dilemma: Experimental evidence. *The Economic Journal*, 103:570–585.
- Axelrod, R. (1981). The emergence of cooperation among egoists. *American Political Science Review*, 75:306–18.
- Brandts, J. and Charness, G. (2000). Hot vs. cold: sequential responses and preference stability in experimental games. *Experimental Economics*, 2:227–238.
- Brosig, J., Weimann, J., and Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, 6:79–90.
- Camerer, C. and Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica*, 56:1–36.
- Cason, T. N. and Mui, V.-L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology*, 42:248–265.
- Cochard, F. and Willinger, M. (2005). Fair offers in a repeated principalagent relationship with hidden actions. *Economica*, 72:225–240.
- Croson, R. (1996). Partners and strangers revisited. *Economic Letters*, 53:25–32.
- Croson, R. (2000). Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization*, 41:299–314.
- Dal Bo, P. and Fréchet, G. R. (2007). The evolution of cooperation in infinitely repeated games: Experimental evidence. Working paper 2007-7, Brown University.

- Davis, D. D. and Holt, C. A. (1993). *Experimental economics*. Princeton University Press, New Jersey.
- Dawes, R. M. and Thaler, R. H. (1988). Cooperation. *Journal of Economic Perspectives*, 2:187–197.
- Engelmann, D. and Strobel, M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Reply. *American Economic Review*, 96:1918–1923.
- Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73:2017–2030.
- Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, 13:1–25.
- Fehr, E. and Gächter, S. (2000). Fairness and retaliation: the economics of reciprocity. *Journal of Economic Perspectives*, 14:159–181.
- Fehr, E., Naef, M., and Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review*, 96:1912–1917.
- Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism—experimental evidence and new theories. In Kolm, S. C. and Ythier, J. M., editors, *Handbook on the Economics of Giving, Reciprocity and Altruism*, pages 615–691. Elsevier, Amsterdam.
- Fessler, D. M. and Haley, K. J. (2003). The strategy of affect: Emotions in human cooperation. In Hammerstein, P., editor, *Genetic and Cultural Evolution of Cooperation*, pages 7–36. MIT Press, Cambridge.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178.
- Fischbacher, U. and Gächter, S. (2006). Heterogeneous social preferences and the dynamics of free riding in public goods. Discussion paper 2006-01, CeDeX.
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71:397–404.
- Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies*, 38:1–12.
- Fudenberg, D. and Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54:533–554.
- Gächter, S. and Renner, E. (2006). The effects of (incentivized) belief elicitation in public good experiments. Discussion paper 2006-16, CeDeX.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206:169–179.
- Goeree, J. K., Holt, C. A., and Laury, S. K. (2002). Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *Journal of Public Economics*, 83:255–276.
- Huck, S., Ruchala, G., and Tyran, J.-R. (2006). Competition fosters trust. Working paper 06-22, University of Copenhagen.
- Katok, E., Sefton, M., and Yavas, A. (2002). Implementation by iterative dominance and backward induction: An experimental comparison. *Journal of Economic Theory*, 104:89–103.

- Keser, C. and van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, 102:23–39.
- Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27:245–252.
- Lacomba, J. A., Lagos, F. M., Reuben, E., and van Winden, F. (2008). After the war: An experimental study of post-conflict behavior. Working paper, University of Amsterdam.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21:153–174.
- Marwell, G. and Ames, R. E. (1981). Economists free ride, does anyone else? experiments on the provision of public goods, iv. *Journal of Public Economics*, 15:295–310.
- McKelvey, R. D. and Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica*, 60:803–836.
- Muller, L., Sefton, M., Steinberg, R., and Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary-contribution experiments. *Journal of Economic Behavior and Organization*, forthcoming.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83:1281–1302.
- Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory*, 21:1–9.
- Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopol-experiments. In Sauermann, H., editor, *Beiträge zur Experimentellen Wirtschaftsforschung*, pages 136–168. J.C.B. Mohr (Paul Siebeck), Tübingen.
- Selten, R. and Stoecker, R. (1986). End behaviour in sequences of finite prisoner's dilemma supergames. *Journal of Economic Behavior & Organisation*, 3:47–70.
- Sonnemans, J., Schram, A., and Offerman, T. (1999). Strategic behavior in public good games: when partners drift apart. *Economics Letters*, 62:35–41.
- Stahl, D. O. I. (1991). The graph of prisoners dilemma supergame payoffs as a function of the discount factor. *Games and Economic Behavior*, 3:368–84.
- Vyrastekova, J. and Onderstal, S. (2005). The trust game behind the veil of ignorance: A note on gender differences. Discussion paper 2005-96, CentER.