TI 2001-003/1
Tinbergen Institute Discussion Paper

Social Status and Group Norms: Indirect Reciprocity in a Helping Experiment

Ingrid Seinen
Arthur Schram

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Social Status and Group Norms: Indirect Reciprocity in a Helping Experiment

Ingrid Seinen* and Arthur Schram*

CREED Department of Economics and Econometrics
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands

e-mail:
* *IngridS@fee.uva.nl*
* *ArthurS@fee.uva.nl*

Manuscript, submitted

Printed April 19, 2000

Abstract

This paper provides experimental evidence showing that indirect reciprocity may important in economic decision making and in the development of group norms. We study a 'repeated helping game' with random pairing in large groups, with individuals equally divided between donors and recipients. Donors decide whether to help the individuals they are matched with against a certain cost or not to help, enduring no costs. We observe that many decision makers respond to the information we give them about former decisions of the recipients, even if they realize that this information is based on transactions with third parties. (*JEL* C92)

I Introduction

Among economists, there is a growing awareness that both trust and reciprocity play important roles in much of economic activity. Because many activities take place sequentially, where one actor endures costs before obtaining the benefits, it is often a matter of trust that the future benefits will indeed be delivered by another actor. In theory, this future delivery can be governed by formal contracts. Often, however, information asymmetries and prohibitive costs of perfect monitoring make the use of enforceable contracts impossible. One has to rely on implicit contracts (Arrow 1974), the enforcement of which is based on reciprocity — here defined as the conditional behavior to reward kind and punish hostile acts even when this is costly.

The distinguishing feature of reciprocity is that it is not based on explicit incentive schemes. In other words, the individual decision to reward or punish

acts of others is not governed by her or his direct interest. A good example of an (experimental) setup used to study implicit contracts and the role of reciprocity is the investigation of the relationship between an employer and an employee by Fehr et al. (1997, 1998). In these experiments, a wage level (and corresponding effort level) is first determined in a contract. Once this has been done, the employee has to decide on whether to supply the work effort agreed upon or to shirk. In some experiments this is followed by an opportunity for the employer to punish or reward the employee. This type of experiment has been undertaken in various institutional settings. A major conclusion is that " ... if both parties in a trade have the opportunity to reciprocate, reciprocal motivations have a robust and very powerful impact on the enforcement of contracts" (Fehr et al. 1997, p. 836).

Many other experimental studies also show that reciprocity is an important motivation guiding human behavior (e.g. Fehr and Gächter 1998). It can be either negative (punishing uncooperative actions; e.g. Güth et al. 1982) or positive (rewarding cooperative actions; e.g. Fehr and Gächter 1996). The games in which reciprocity has been observed include public goods games (Brandts and Schram 1998), the prisoners dilemma (Andreoni and Miller 1993, Cooper et al. 1996), centipede games (McKelvey and Palfrey 1992), an investment game (Berg et al. 1995) and a gift exchange game (Gächter and Falk 1997). Most experimental studies are firm in their conclusions that some type of reciprocal motivations can be observed in the laboratory. In addition, reciprocity seems to be a stable outcome, in which frequency and strength do not decline over time (Roth et al. 1991, Fehr et al. 1993, 1998, Cooper et al. 1996, Gächter and Falk 1997).

The motivation governing reciprocity can be related to other regarding preferences such as (reciprocal) fairness (e.g. Rabin 1993) or (reciprocal) altruism (Trivers 1971, Levine 1998) or to the quest for efficiency gains through cooperation (Brandts and Schram 1998). For an overview of this type of motivations, see Schram (1998). In this paper we focus on the form that reciprocity might take and on the consequences it may have for the (repeated) interaction between individuals in a group.

Once the occurrence of reciprocity has been established, there is also room for strategic reputation building by individuals. Even those who would otherwise not act cooperatively, might do so in order to increase the probability of being reciprocated. Hence, reciprocity provides an explanation for cooperative behavior of individuals, for whom it is not in their (short-term) interest to cooperate. As a consequence, it provides necessary conditions for cooperative behavior to be stable in the long run and is therefore seen as an important mechanism in the evolution of cooperation in human societies (Trivers 1971, Axelrod 1984, Ridley 1996, Binmore 1998).

To a large extent the literature has focused on *direct* reciprocity, i.e., the motivation to respond to the acts of individuals one has interacted with before. There has been much less attention for *indirect* reciprocity, where a cooperative action is reciprocated by a third actor, not involved in the original exchange. The biologist Alexander (1987) argues that indirect reciprocity plays a central role in human societies. In his view, the link between actors is made through 'reputation' or 'social status'. Individuals in society are continuously being evaluated and reassessed with respect to how 'cooperative' they are. This gives them a reputation that may be used by others when deciding on how cooperatively to act towards them. We will discuss the literature on indirect reciprocity

more extensively in the next section.

In this paper, we present experimental evidence showing that indirect reciprocity may be an important motivation governing the choices of many people in the laboratory. The hypothesis tested is that people behave cooperatively to people who were cooperative towards others. Information about choices made by someone else does not require direct interaction, but can be obtained by observation. Hence, (providing) information about former behavior of the partner in a game is a way to control levels of cooperation. We use a 'repeated helping game' with random pairing in large groups, in which individuals either decide whether or not to help the subject they are matched with or are on the receiving end of this interaction. We observe that many decision makers respond to the information we give them about the 'social status' of the person they are dealing with. This occurs even though they realize that this status is based on transactions with third parties and not on previous interactions with themselves. We use this result to study the consequences of indirect reciprocity for the dynamics of decision making in a group and the development of group norms.

The paper continues as follows. The following section discusses the theoretical literature on indirect reciprocity. Section III presents our experimental design and procedures. The results are presented in section IV, where we distinguish general results and a more detailed study of individual strategies in our experiments. The consequences for group dynamics and group norms are discussed in section V. Section VI summarizes and concludes.

II Models of indirect reciprocity

Since Axelrod ran his famous computer tournaments (Axelrod and Hamilton 1981, Axelrod 1984), the most commonly used framework for analyzing reciprocity has been the repeated Prisoner's Dilemma (PD) game. The outcome of the tournament predicted that if two players play a repeated PD game and choose simultaneously to cooperate or to defect, the best thing they could do would be to play Tit for Tat (TFT), that is, to start cooperatively in the first period and then to imitate (reciprocate) in all subsequent rounds. Theorists did not unanimously accept the analysis of Axelrod and Hamilton and showed that TFT is not always an evolutionary stable outcome.[1]

In a slightly different scenario, introduced by Trivers (1971), a model of cooperation is presented in which two players act in an *alternating* sequence of moves. The player whose turn it is to act has the decision between altruistic or non altruistic behavior. In this setup, conditional altruism is protected against exploitation, and it ensures that both partners have an incentive to cooperate. This repeated alternating PD has been studied less intensively, but it has been shown to have a cooperative evolutionary stable equilibrium (Nowak and Sigmund 1993, Leimar 1997).

In the literature, reciprocity is generally understood as direct reciprocity, requiring repeated encounters between the same two individuals. However,

[1] Although TFT was the winning strategy in Axelrod's tournaments, it is not an evolutionary stable strategy (Selten and Hammerstein 1984) and not the winning strategy in all environments (Boyd and Lorberbaum 1987, Foster and Young 1990, Nowak and Sigmund 1994). However, reciprocal strategies, like TFT, perform better than 'all-defect' strategies in many theoretical and experimental environments and thus have at least some explanatory power in the evolution of cooperation.

many authors have stressed that reciprocity does not need to be restricted to two individuals (Trivers 1971, Sugden 1986, Alexander 1987, Binmore 1992).

According to Alexander, who introduced the term indirect reciprocity, individuals not only use information from their own experience, but also react to interactions they observe between other individuals within their group. *"In indirect reciprocity the return is expected from someone other than the recipient of the beneficence. This return may come from essentially any individual or collection of individuals in the group. Indirect reciprocity involves reputation and status, and results in everyone in a social group continually being assessed and reassessed by interactants, past and potential, on the basis of their interactions with others" (Alexander 1987, p. 85).* Moreover, he calls indirect reciprocity the evolutionary basis of moral systems, which prescribe cooperative behavior.

A first attempt to model reciprocity in larger groups was within small intransitive networks: individual A helps B, who helps C, who helps D, who in return helps A (Boyd and Richerson 1989). This type of indirect reciprocity has been studied experimentally in pension games by Heijden (1996). However, it still requires very strict interactions and is not based on reputation.

With the strategy *Observer Tit for Tat* Pollock and Dugatkin (1992) implement a notion of indirect reciprocity, based on reputation. They study a repeated PD game in which they allowed players to occasionally observe a co-player before starting the repeated interaction. If the future co-player was seen defecting in her or his last interaction, then Observer Tit For Tat prescribes to defect in the first round. This strategy outperformed the usual Tit for Tat and could even coexist with a subpopulation of defectors when no degree of future interaction with the current partner was presumed.

In spite of these various attempts, the first authors who fully recognized the scope of Alexander's indirect reciprocity theory for the evolution of cooperation were Nowak and Sigmund (1998a,b). In their main model, they use image scores, which are integer values in the range from -5 to 5, to describe the 'level of cooperation' of individuals. Wedekind (1998) uses Alexander's term 'social status' when referring to this score. All individuals have their own score, starting at zero, but changing with decisions made in the game.

A game is played for several rounds. In each round two individuals are randomly chosen, one recipient and one donor. The donor decides whether or not to give an amount b to the recipient at a cost $c < b$. When the donor gives to the recipient ('helps'), her or his score is increased with one point, otherwise (s)he loses one point. The decision whether or not to help is based on the score of the recipient. Every individual is assumed to have a strategy k, an integer value between -5 and 6. Donors only cooperate if the image score of the recipient is at least k. The strategy $k = -5$ thus represents unconditional cooperators, whereas the strategy $k = +6$ represents unconditional defectors.

The game is studied through simulation. At the beginning, the k-values are randomly distributed across the individuals. After 125 donor-recipient pairs have been chosen, a new generation starts, with a distribution of k-values proportional to their payoffs in the previous generation. Nowak and Sigmund find that the whole population consists of $k = 0$ after 166 generations, which is the most discriminating cooperative strategy.

Cooperative regimes also evolve well when mutations are added, when interactions are only observed by some of the group members, or when individuals also care about their own score. An important requirement is that a sufficient proportion of the individuals is conditionally cooperative. When initially the

whole population consists of only self-regarding strategies, no cooperative society can evolve. An essential part of the Nowak and Sigmund studies is that they use replicator dynamics. This means that they implicitly assume that individuals (learn to) cooperate, if this will increase their payoff. The strategies are cut-point strategies based upon the scores of the recipients and/or on their own scores. Under these assumptions they were able to show that cooperation will evolve when the strategies are at least partly based upon the strategy of the other.

In this setting, whether or not indirect reciprocity can enforce a norm of (some degree of) cooperation depends strongly on the ability of individuals to recognize the possibility to build a reputation by strategically cooperating, and on their willingness to enforce this norm by punishing defectors (at the cost of blemishing their own reputation). Note that this requires stronger assumptions than needed for direct reciprocity, because individuals do not react to what is done to themselves, but to what is done to others. They need to either identify with the earlier opponents of their counter player, or be willing to conform to a social norm of reciprocity.

It is this Nowak and Sigmund model of indirect reciprocity that we test experimentally in this paper. From earlier experiments we know that direct reciprocity leads to high cooperative levels in interactions with fixed partners. People behave nicer to people who were nice to them. The new hypothesis tested in this paper is that people also behave nicer to people who were nice to others.

III Experimental procedures and design

The experiment consisted of 6 sessions that were run at the CREED laboratory at the University of Amsterdam. Subjects voluntarily signed up after public announcement. Most of the participants were students from various faculties including economics, social sciences, law, chemistry and computer sciences. In each session 28 subjects participated simultaneously. They were randomly assigned to seats in the laboratory, which are separated by partitions. No communication was allowed. Instructions were computerized and could be read at one's own pace.[2] An obligatory quiz was used to ensure that subjects understood the instructions. When all subjects had finished reading the instructions the experiment started. The experiment itself was also computerized and consisted of at least 90 rounds. Thereafter, any additional round was started with a common knowledge probability of 90%. The duration of one session was about 5 quarters of an hour. There was a show up fee of either 20 or 30 guilders, depending on the specific parameters in the session concerned. In addition, the subjects were told that they would be paid the earnings (which could also be negative) of 20 rounds that would be randomly chosen at the end of the experiment. Therefore, at any point, subjects did not know exactly how much they had earned. This was designed as such to avoid changes in behavior as a result of income effects.[3] Subjects earned on average 35.37 guilders. At the end of the experiment subjects were asked to fill in a questionnaire concerning personal background and motivation.

In each round, pairs of subjects were chosen, consisting of one donor and

[2] An English translation of the instructions is provided in the appendix.

[3] Of course, the implicit assumption here is that the difference in show up fee does not cause an income effect. We did not find any evidence of such an income effect.

Table 1: Parameters for the three conditions tested.

Treatment	Benefit b for recipient	Cost c for donor	# of previous choices of recipient shown to donor	Show-up fee
LCI	250	50	6	2000
HCI	250	150	6	3000
HCN	250	150	0	3000

Note: The numbers referring to costs and benefits represent the payoffs (in Dutch cents) in case the donor chose yellow (help). The payoff for both the recipient and the donor was zero in case the donor chose blue (pass).

one recipient. Both the roles and the pairing were determined completely at random. Subjects were told that they would be randomly matched with anyone else of the 28 participants in the laboratory. In fact, however, the randomization was done in 2 separate groups of 14 subjects each in order to increase the number of independent replicates.

The setup we use in our experimental setting is that of a 'helping' experiment. In any given round, only the donor had to make a decision. (S)he had the choice to either 'help' the recipient at a cost c, in which case the recipient would receive a benefit $b > c$, or to 'pass', in which case both individuals received 0. After the decisions were made the recipients were informed about the decision of the donor they were matched with. They did not receive any information other than this outcome.

The choices were presented in a payoff table with yellow and blue representing the choices 'help' and 'pass', respectively. The payoff values chosen for the case where the donor chose yellow (help) are presented in table 1, below. Note that the benefit b for the recipient was always 250 cents[4], whereas the cost varied across sessions. Two values were chosen: $c = 50$ cents in the low-cost (LC) condition and $c = 150$ cents in the high-cost (HC) condition.

Besides the costs, we also varied the amount of information given to the donor. In a baseline condition without information, donors were not told anything about the previous choices of the recipient. This is referred to as the No-Information (N) condition. This baseline was only used in the HC condition. In the Information (I) condition the notion of social status was implemented in the following way. Before making their choice, donors were given information about the previous 6 decisions made by their recipient when (s)he had been appointed the role of donor. This information was summarized in 2 numbers, the number of choices for yellow (help) and the number of choices for blue (pass). No order in these choices could be inferred. All in all, we ran three treatments: LCI; HCI; and HCN (cf. table 1). Because we organized two sessions each and distinguished two groups of 14 in each session as discussed above, we have 4 (statistically independent) groups per treatment. Within each group we have 14 subjects[5], each making about 45 to 50 decisions.

Note that information about 6 decisions (on average 12 periods) reflects a limited memory, as is often used in evolutionary game theoretic models (Young

[4]This is fl. 2.50 in Dutch currency and corresponded with approximately $ 1.25.

[5]In two groups of HCN we had only 10 subjects due to no-shows.

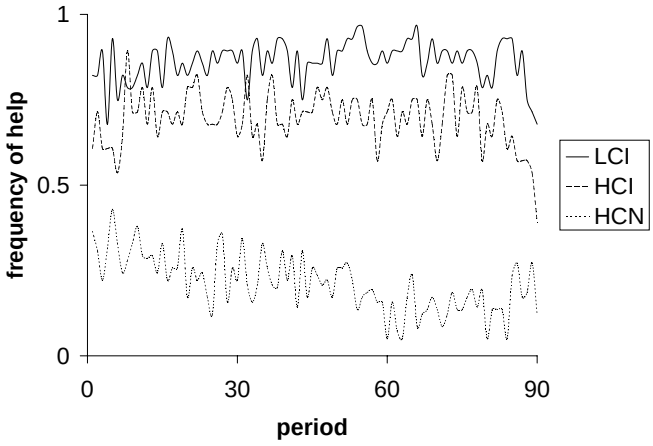


Figure 1: Average frequency of helping per treatment.

1998). This gives subjects the opportunity to 'clean' their record. We chose memory to be limited in order to decrease the influence of early periods and to keep both the impact of one single decision and the information level constant.

In all sessions individuals were given a summary of their own last 6 choices and their own results from all earlier rounds, both being displayed permanently on the screen. No summary statistics or information about the decisions of others (except the recipient) was provided.[6] In this setup, subjects can learn about what others do, because they interact with someone else every period, but they cannot infer how a donors decision is related to her or his own score.

IV Results

The results are presented in two subsections. First, we present a general overview and analysis of helpful behavior, both over periods and across treatments. In addition, we provide some data reflecting the relationship between individual choice and the social status of the receiver and donor. Then, we elaborate on this relationship by analyzing individual strategies in more depth. These strategies may depend on the social status in various ways.

IV.A General Results

We find a high level of cooperative choices ('helping') in all sessions. The percentage (in the first 90 rounds) is 86% in LCI, 70% in HCI and 22% in HCN. Figure 1 presents the frequency of helping per round (up to round 90) for the three treatments.

This figure shows various things. First of all, there is a clear order in helping frequency, the highest fraction of helpful choices being observed in LCI and the lowest in HCN. Second, LCI and HCI show some signs of an end effect, with helpful choices starting to drop around period 80. Nevertheless, *help* is still

[6] We did not provide information about the donor to the recipient in order to avoid direct reciprocity towards subjects with the same score and to minimize the information flow about strategies of others.

Table 2: Helping frequency per group.

Treatment	groups				total
LCI	0.74	0.89	0.90	0.92	0.86
HCI	0.49	0.66	0.77	0.86	0.70
HCN	0.18	0.19	0.22	0.27	0.22

Note: The numbers in the cells are the fraction of helpful choices across the first 90 periods. Groups are shown in increasing order of helping frequency from left to right.

being chosen in more than 50 % of the cases (except for the final two periods in HCI). No end effect is observed in HCN.

Even though there are differences in the average level of helpful choices across treatments, there are also important differences between groups within a treatment. Figure 2 shows the 7-period moving average of choices of each of the four groups of 14 subjects per treatment.

Figure 2 shows that various groups have their own dynamics. This is especially clear in HCI. Note that this makes it impossible to assume statistical independence of choices within a group. To undertake testing, we therefore need to summarize group statistics. Table 2 presents the helping frequency per group for each treatment.

Using the four averages per treatment as four independent observations, we conducted a Mann–Whitney test for a pairwise comparison across treatments. The results show that the differences LCI-HCN and HCI-HCN are significant at the 5 %-level and the difference LCI-HCI is significant at the 10 % level. Therefore, even with such a conservative test, we find treatment effects, where both the level of costs and the provision of information affects choices.

Next, we turn to the influence of social status of the recipients on the choices made by donors. Recall that this is reflected in the information we provide to the donor about the six previous choices of the recipient. Because the level of helpful choices in LCI shows too little variation for a fruitful analysis, we will not analyze this treatment.

Social status may matter in two ways. First of all we will look at the influence that social status of recipients has on the decision making of donors.

Figure 3 shows the fraction of helpful choices as a function of the recipient's social status in HCI (note that this is not observed in HCN). Social status may vary across 7 categories: from (0,6) (zero help and six pass) to (6,0) (six help and zero pass) in the previous six choices. It shows that subjects with a high score are almost always reciprocated whereas individuals who never choose to help are helped about 25 % of the time. Apparently it matters what the social status of the recipient is. This is a first indication of indirect reciprocity motivating subjects' behavior. Note that this motivation is not likely to be strategically aimed at direct reciprocity. Subjects believe that the probability that they will meet their partner again in any given round is 1/28. This probability is much too low to motivate any kind of decisions in the hope of receiving a direct reciprocal response from the same partner. Moreover, direct reciprocity cannot explain the difference between the helping frequency in HCN and HCI.

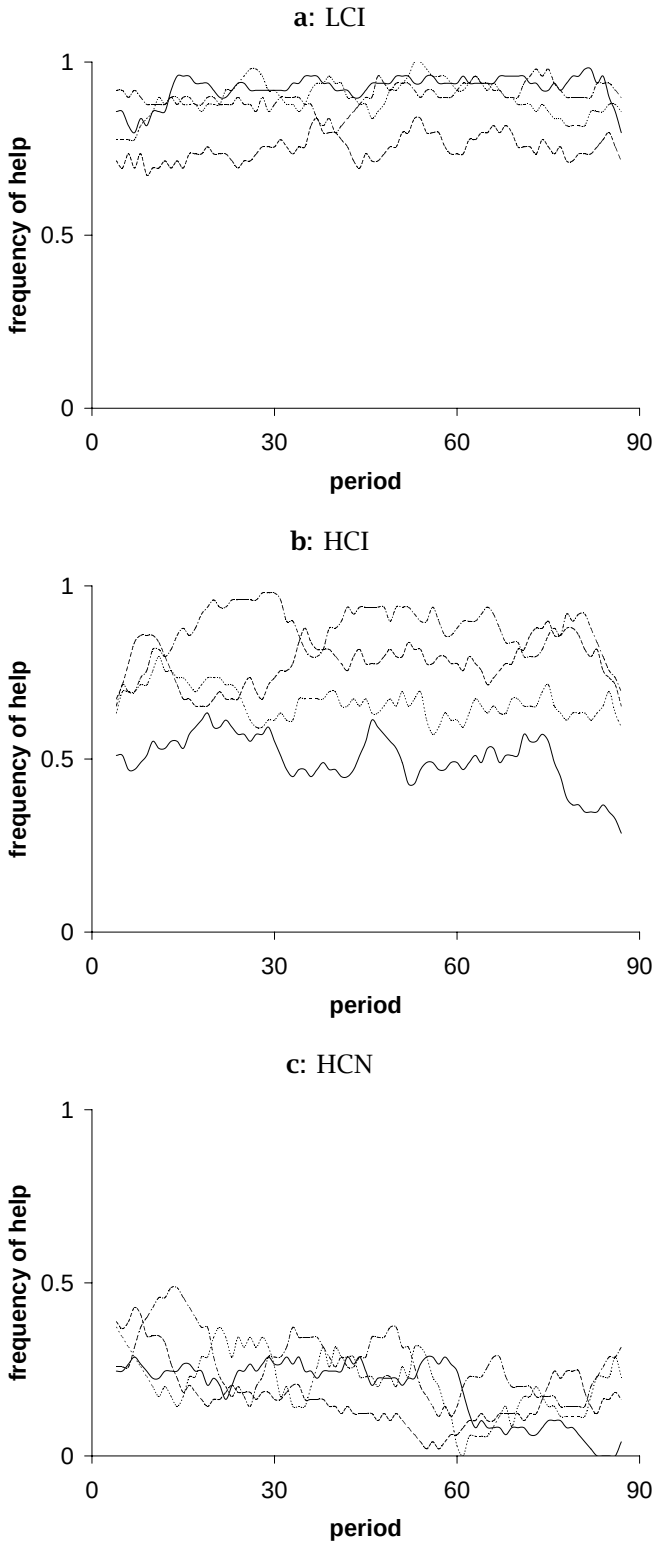


Figure 2: Moving average of helping frequency per group.

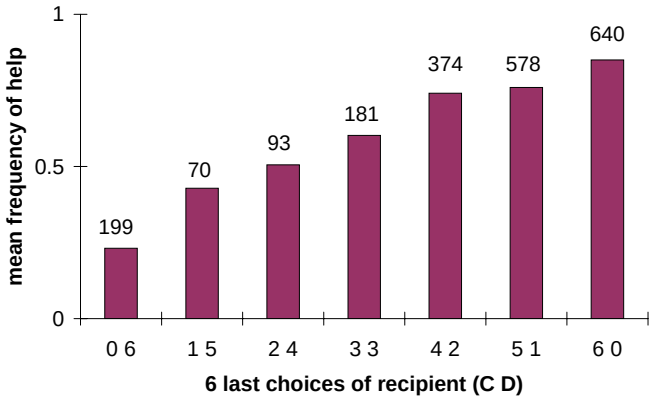


Figure 3: Helping frequency based on previous 6 decisions of recipients up to period 90 of all periods before which both recipient and donor made at least 6 decisions. Number of observations per category are shown above the respective bars.

Given the numbers underlying figure 3, one can estimate the expected return for various levels of social status. It turns out that the highest return (85 cents per round) is to be expected from keeping a score of 4 out of 6 and the lowest (58 cents per round) from keeping a score of 0 (unconditional defection).

To test the influence of indirect reciprocity, we conduct a Page test (see Siegel and Castellan 1988) for ordered alternatives on the rankings from 1 to 7 for each of 7 categories in the 4 groups in HCI. The rankings are based on the mean number of helpful choices towards recipients with a score 0 to 6.[7] The relation between the score of the recipient and the number of helpful choices is significant at the 0.001 level.

If social status of the recipient is important in deciding whether to help or not, then donors might consider the effect that their own score will have on their future donors when they will be in the role of recipient themselves. For HCN this should not be the case, because the donor knows that this information will not be passed on to others. This is reflected in the data presented in table 2. The average helping frequency level is much higher in case the information is passed on. Figure 4 shows the fraction of helpful choices as a function of the own social status. Social status may again vary from (0,6) (zero help and 6 pass) to (6,0) (six help and zero pass) in the previous 6 choices. The figure for HCN shows a uniformly increasing frequency of help. This simply reflects some stability in individual choices: people who chose *help* more often in the previous 6 opportunities are more likely to do so now.

The own social status plays a much more important role in the decisions of donors in HCI than in HCN. Besides from the total frequency of helping, this is also apparent from figure 4. Especially for low levels of own social status, subjects are much more likely to help if they know that their score will be passed on.

From the numbers underlying figure 4 one can derive the development of helping behavior during the experiment. For example, if the frequency of help-

[7]The null hypothesis is that the average rank in each of the categories are the same. The alternative hypothesis is that the average rank increases across category 1 to 7 (all differences are inequalities and at least 1 difference between 2 successive categories is a strict inequality).

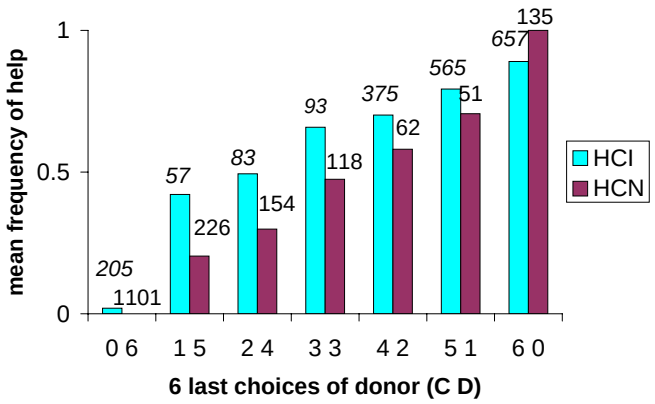


Figure 4: Helping frequency based on previous 6 decisions of donor in periods up to 90 before which both recipient and donor made at least 6 decisions. Number of observations per category are shown above the respective bars.

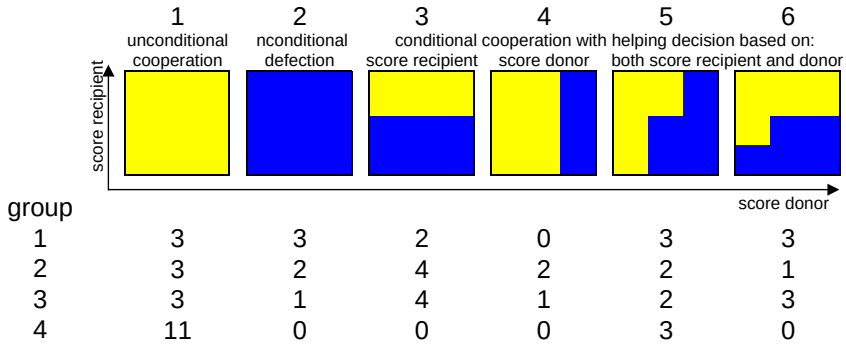
ing with a (1,5) score is higher than 1/6, then there is a drift towards higher frequencies of helping. Figure 4 reflects a drift towards lower frequencies of helping for HCN, except by those already choosing to help for 100% (6,0). HCI shows a tendency towards helping behavior in 4 or 5 out of 6 periods. Recall that we concluded from the analysis of figure 3 that the highest payoffs could be expected from a social status of 4, so the drift in case information is passed on is to the score that maximizes the expected gains.

IV.B Individual strategies

In this section, we study individual strategies in more detail. We will allow these strategies to depend on the own image score and/or the image score of the recipient[8]. Again, we focus the analysis on the HCI sessions, because LCI does not provide enough variation in strategies and HCN does not provide donors with information about the score of the recipient. We describe strategies using a figure where the number of own previous helpful choices (out of 6) is given on the horizontal axis and the number of previous helpful choices by the other is given on the vertical axis. Figure 5 shows the strategies we distinguish.

In this figure, light gray areas refer to a helpful strategy and dark gray areas to a 'pass'. Hence, strategy 1 shows a strategy of unconditional cooperation and strategy 2 describes unconditional defection. Strategy 3 reflects a cut point strategy of cooperation if and only if the other has helped enough recipients in the past, whereas strategy 4 reflects a cut point with respect to the own social status (keeping one's own score at some specified level). Finally, strategies 5 and 6 use both the own score and the score of the recipient. In strategy 5, the donor ensures a certain own score but the willingness to help beyond that depends on the score of the recipient. In strategy 6, the donor never helps a recipient whose score is too low, but is willing to help others, depending on the scores of both. The numbers in figure 5 refer to the number of participants in the various HCI groups using the respective strategy. This was simply deter-

[8]An obvious naive strategy would be to consider the effect of the outcome in the previous round. However, this is only found to be significant in 4 individuals and therefore ignored in this analysis.



group						
1	3	3	2	0	3	3
2	3	2	4	2	2	1
3	3	1	4	1	2	3
4	11	0	0	0	3	0

Figure 5: Individual Strategies: On the x-axis the own social status is shown, on the vertical axis the social status of the recipient. Light grey denotes helping, dark grey denotes passing. Strategy 5 also includes all strategies with a minimum own score, but no maximum own score (choose *help* when own score *or* recipient's score are at least k_i). Strategy 6 also includes all strategies for which no score of the recipient is high enough to ensure help (choose *help* when own score not higher then k_d *and* score recipient is at least k_r).

mined by the number of choices that could be explained by the strategy in the first 60 periods, excluding the periods in which less than 6 decisions had been made. If two strategies could explain an equal number of choices, the more 'simple' one (i.e. with the least parameters) was chosen.[9] The strategies estimated involve unconditional cooperation (36%); unconditional defection (11%); conditional cooperation based on the recipient's social status (18%); conditional cooperation based on the own social status (5%); conditional cooperation primarily based on the own social status, but also on the recipient's social status (18%) and conditional cooperation primarily based on the recipient's social status, but also on the own social status (12%). These individual strategies were used to predict individual choices in periods 61–80. Table 3 presents the results of these exercises.

The number of choices that can be explained with one of the cut point strategies (strategy 3–6) is significantly higher than the number of choices explained by choosing *help* or *pass* only (strategy 1 or 2). The total improvement relative to this benchmark is 11%. [10]

Table 3 shows that the strategies can predict out of sample quite accurately. On average, the choice made is predicted 84% of the time in periods 61–80. We used predictions of a random strategy as a benchmark. A random strategy predicts to always help when the observed frequency of *help* in periods before 61 $p \geq 0.5$, otherwise it predicts to always choose *pass* (strategy 1 and 2). The total improvement from the benchmark to the strategy analysis is 5%.

Unfortunately, the analysis of the individual strategies could not always provide us with the exact scoring rules individuals used. Different scores could explain their behavior equally well, mainly because of the fact that not all indi-

[9] Strategies 1 and 2 are considered to be more simple than 3 and 4; 3 and 4 are considered to be more simple than 5 and 6.

[10] To determine the extent to which this improvement can simply attributed to increasing the number of parameters, we need a randomization procedure to determine the number of choices that can be explained by adding three parameters to divide the strategy space into 6 sections. Reshuffling *help* and *pass* per individual across periods and calculating best strategies 10000 times shows an increase of 4% at the 0.001 level. Hence the increase in predictive power we observe is not simply a consequence of adding parameters.

Table 3: Predictive power of strategies

Group	Explained choices period < 61	Predicted choices periods 61–80	Baseline predictions
1	0.92	0.83	0.63
2	0.90	0.81	0.85
3	0.89	0.85	0.81
4	0.93	0.88	0.86
Total	0.92	0.84	0.79

Note: The third column presents the fraction of correct predictions for periods 61–80 based on the calibrated strategies on periods ≤60 . The fourth columns shows the fraction of correct predictions based on the two baseline strategies: if the probability to play help $p \geq 0.5$ the prediction for period 61–80 is to always choose help, if $p < 0.5$, the prediction is 'always pass'.

viduals were confronted with every score.[11] For the 23 subjects (in four groups) estimated to use strategy 3, 5 or 6, we estimate the average cut point for the recipient's status to lie in the interval {(1.7–2.4), (2.1–2.7), (2.3–3.2) and (4.0–5.0)} respectively, (groups are ordered in number of helpful choices). For the predictions of choices in period 61–80 we use the maximum of this interval. The scoring rules of the own social status show less ambiguity. For the cut point for the own social status of the 18 subjects with estimated strategy 4, 5 or 6, we find {3.5–3.8, 4, 4.8, 5.3}.[12] Note that both cut points seem to increase with the level of observed cooperation in the group. This is elaborated in section V.

V Group dynamics and group norms

In the previous paragraph we analyzed individual strategies and showed that many strategies can be characterized by minimum score rules for the recipient and/or the donor. The minimum score for the recipient can be seen as a norm that (s)he has to satisfy. We saw that the estimated norms differ across groups. In this paragraph we will concentrate on the development of these norms and their relation to group composition. We will argue that individual norms are at least partly determined by group composition, developing similarly within, but distinct across groups. This leads to the emergence of 'group norms'.

The possibilities for learning are quite limited in our set-up. Experimenting with ones score is possible, but time consuming and the outcome will be different across subjects, because every individual is matched with a different set of others. Subjects do not know how often others are helped, and recipients do not know what type of donor (high status or low status) they are matched with. Therefore they cannot imitate behavioral strategies of (successful) others, the

[11] Consider an individual who always helped recipients with a score of 4 or higher and never helped recipients with a score of 1 or lower. If (s)he never met a subject with a score of 2 or 3, his or her cut point can be 2, 3 or 4.

[12] For two individuals with strategy 6 the own minimum score is not well defined, because the interval between the two cut points of the recipient's score is very small. No observations in this interval were made in rounds 61–80.

only thing they can copy is the score of other recipients or some measure (e.g. mean or modus) of the subpopulation of recipients they are matched with.

From the information they receive about the scores of recipients, subjects can estimate a distribution of scores within their group. They will learn what score can be considered to be low and what score can be considered to be high within their group. If they are nicer to subjects with a high social status than to subjects with a low status (e.g., if they use a strategy type 3, 5 or 6), their own score will be affected by the relative frequency of encounters with these subgroups. Taking this into account, subjects that use a norm for the recipient's social status can determine a maximum norm k_i they can use for this status, that will allow them to keep their own score $s_i \geq k_i$. In other words, subjects using strategies 3, 5 or 6 can learn to adjust their cut point in a way that their own score adheres to it. It is this type of learning, finding a norm that is consistent with the own social status, that we think is important in synchronizing norms within a group. Our data provide some evidence that this type of learning is taking place.

We mentioned earlier that various groups within the HCI treatment have their own dynamics (figure 2) and vary in the average frequency of helping (table 2). To check whether these differences across groups exist from the beginning or develop during the experiment, we use two tests. For all periods aggregated in groups of 10, we test whether frequency of helping differs across groups and whether the number of helping subjects is different across groups.

Using a χ^2 test we find significant differences in the frequency of *help* across groups in all 10-round blocks, though the difference is only marginally significant in the first block of 10 periods (first 10 periods $\chi^2 = 7.13, p = 0.068$, later periods $\chi^2 > 14, p \leq 0.003$). When comparing the helping frequency per individual during 10 periods using a Kruskal–Wallis test, groups appear to be different from the third 10-period block on (first 20 periods $p > 0.2$; later periods $p < 0.05$).

The difference between the two tests is an indication that differences across groups in early periods (< 20) are concentrated in a few individuals, whereas in later periods a larger group of individuals differ in their frequency of help. This could imply that the interaction between subjects changes individual behavior and therefore influences dynamics within groups.

Our results from the individual strategy analysis are in agreement with these findings. Of the 56 subjects in HCI, 23 were found to use a strategy based on the score of the recipient (strategy 3, 5 or 6). These subjects behave less helpful towards subjects with a low status then towards subjects with a higher status. As a consequence, in a group with more defectors (strategy 2), these conditional cooperators will have to punish more frequently and therefore obtain a lower social status themselves. We might therefore expect a negative correlation between the helping frequency of unconditional cooperators in a group and the number of unconditional defectors (strategy 2) and vice versa, a positive correlation between helping frequency and the number of unconditional cooperators (strategy 1). This is indeed the case: For the 4 groups in order of helping frequency, the fraction of helpful choices by conditional cooperators (strategies 3, 5 and 6) are {0.59, 0.74, 0.79, 0.89}, the numbers of unconditional defectors were found to be: $n_1 = 3, n_2 = 2, n_3 = 1, n_4 = 0$. Recall that the norms estimated in the previous section were also shown to be increasing in the frequency of help.

Table 4: Donors reported norm k_r, compared with social status s_r of recipients they were matched with.

periods	$s_r > k_r + 1$	$s_r = k_r + 1$	$s_r = k_r$	$s_r = k_r - 1$	$s_r < k_r - 1$
	a: frequency of help				
≤ 20	82	77	76	90	53
21–40	90	84	88	74	29
41–70	88	90	85	59	22
71–90	87	86	64	50	14
	b: fraction recipients				
Total	36	28	17	7	13

To study the relationship between behavior and group norms in more detail we use data from the post-experimental questionnaire in which 68% reported having used a minimum score rule for the recipient.[13] The average cut points $\bar{k}$ reported in the four groups are: $\bar{k}_1 = 2.4, \bar{k}_2 = 3.4, \bar{k}_3 = 3.8, \bar{k}_4 = 4.3$. First note that these cut points are similar to the cut points found in the individual strategy analysis, where $\bar{k}_1 = 2.4, \bar{k}_2 = 2.7, \bar{k}_3 = 3.2, \bar{k}_4 = 5$, as reported in section IV.B.[14]

Second, it is noteworthy that 70% of the conditional cooperators report using the mode cut point in their group: $\hat{k}_1 = 2, \hat{k}_2 = 3, \hat{k}_3 = 4, \hat{k}_4 = 4$. Finally, we once again see that there are large differences across groups. A Kruskal–Wallis test shows that groups are significantly different at the 0.01 level. (From individual comparisons ($\alpha = 0.05$) we can conclude that $\bar{k}_4 > \bar{k}_1$; $\bar{k}_4 > \bar{k}_2$ and $\bar{k}_3 > \bar{k}_1$).

Norms are consistent with the donor's own status. A lower own status corresponds with a lower norm. Until round 90 subjects can keep their own status above the norm they use for recipients in 88% of the time. In case they had used a higher norm than the reported one, this would have led to an average own status below that norm in all sessions. In that case, on average, conditional cooperators would punish each other for punishing defectors, which would lead to a cascade of passes until nobody helps anymore.

Next, we study the development of behavior over time. To do so, we analyzed the data for different time intervals separately. In all 4 groups we find the same pattern: 79–84% of all choices can be explained by the reported cut points between rounds 20 and 90 (68% in earlier rounds). Distinguishing recipient scores above and below the cut point, subjects help in 88% of the time to recipients with a social status s_r equal to or higher than their cut point k_r (79% in earlier rounds). When matched with recipients with $s_r < k_r$, donors help in 31% of the cases, (69% in earlier rounds).

For the subjects with a reported cut point k_r, table 4 provides data about the relation between k_r and the scores s_r of the recipients they were matched

[13] To test for changes in behavior we are forced to use the cut points reported in the questionnaire as opposed to the ones estimated from the choices. One cannot test behavior using a categorization based on that behavior.

[14] We use the right border of the estimated interval, because this is more closely related to the question in the questionnaire, in which we asked subjects to give the minimum score someone should have to get help.

with. In table 4b we show the percentage of recipients with s_r lower than, equal to or higher than k_r. It shows that a large majority (80 %) of these subjects is matched with recipients having a score greater than or equal to the cut point. Table 4a gives the percentage of helpful choices the donor made when matched with recipients with various s_r's.

Table 4 shows a number of interesting patterns. First of all, in the first 20 periods, the reported cut point does not predict choices very well. The highest frequency of *help* is found for the situation where the recipients have a score that is just below the cut point, i.e., when the cut point would predict that the donor should not help. This inconsistency in the first rounds reflects our earlier observation that cut points are adapted early on in order to allow the own status to fit the own norm. Second, after period 20, levels of helping are increasing in the social score of the recipient (with only two exceptions). Hence, the norm qualitatively describes the donors' choices. Third, after period 70 the frequency of help declines for any social status of the recipient. At this stage of the experiment, subjects are less inclined to help in general. Reciprocation is losing some of its motivational impact.

Even in periods 21–70, the cut points are far from perfect. Especially the high frequencies of helping when the recipient has a score that is 1 below the reported norm might seem surprising. This can be explained by considering the strategies distinguished in the previous section. Many of the subjects reporting a cut point for the recipient's status were found to (also) use a cut point for their own score. This might be especially relevant in cases where the recipient's status was 'close' to the norm. In that case, the own score might become the dominant concern, as in strategy 6. In future experiments, we hope to use a design that allows us to discover more subtleties in individual strategies.

From these observations, we conclude that the reported cut point most closely refers to the norm used in periods 21–70. In earlier periods, cut points have not been established and in later periods they play a less important role.

VI Summary and conclusions

In this paper we have presented clear evidence that indirect reciprocity plays an important role in our experimental setting. This setting was based on recent theoretical literature stressing the importance that indirect reciprocity may have in large groups. In the (experimental) economic literature, the focus has been on direct reciprocity. There is an abundance of evidence showing the existence of behavior that might be interpreted as direct reciprocity. This paper has shown that indirect reciprocity may also be an important individual motivation. This motivation may lead to stable cooperative regimes, in which the majority of the population helps.

Indirect reciprocity in our case shows up in the donor's sensitivity to the recipient's social status. It is therefore rational for individuals to build up a good reputation. This is indeed observed in our data: the own social status is important for the decision whether or not to help, when this status is passed on to future donors. Both when estimating strategies and when analyzing responses to the questionnaire, we observed that many subjects used cut points with respect to their own and the recipient's status. In the strategy analysis we distinguished 6 strategies: always help, never help, a cut point rule based on social status of recipient, a cut point rule based on own social status and two

combinations of these cut point rules (see figure 5). We found that 53% of the subjects used a cut point strategy. From these subjects 48 %-point based their decision on the social status of recipients and 35 %-point on their own social status. The fraction of subjects that based their decision on the social status of recipients was high enough to make investment in the own status worthwhile. That is, given the cut points used, the expected payoffs of keeping the own score high enough was larger than the expected payoff from never helping.

However, the cut points differ across groups. It follows that a good reputation in one group does not have to be good enough to be helped in another group. These cut points or norms develop in an early stage of our experiment and are related to the different composition of strategy types across groups. We argued that conditional cooperators pass more often when they encounter defectors and therefore lower their own social status. The norms that subjects use for recipients are related to their own social status as well. Norms are lower or equal to the own social status in 88% of the time. Consequently, subjects lower the norm they use for the recipient when they cannot obtain that status themselves. This explains why social status within groups differ less then across groups.

This study is meant to be a starting point in the understanding of how cooperative regimes can evolve in large societies. The economic consequences of indirect reciprocity and endogenous norms are obvious and warrant further research in this area. We intend to continue this research by studying the development of norms related to the donor's and the recipient's status in more detail. By applying a strategy method and restricting information flows, we hope to gather more information about the way in which norms are adjusted to observed behavior within the group.

Acknowledgements
We thank Arno Riedl and Joep Sonnemans for inspiring discussions about the experimental design as well as for useful suggestions for the paper and Sido Mylius for critically reading the manuscript.

References

Alexander, R.D. *The Biology of Moral Systems*. Aldine de Gruyter, New York, 1987.

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

Arrow, K. *The Limits of Organization*. Norton, New York, 1974.

Axelrod, R. *The Evolution of Cooperation*. Basic Books, New York, 1984.

Axelrod, R. and Hamilton, W.D. "The evolution of cooperation." *Science*, 211, pp. 1390–1396, 1981.

Berg, J.; Dickhaut, J. and McCabe, K. "Trust, reciprocity and social history." *Games and Economic Behavior*, 10, pp. 122–142, 1995.

Binmore, K.G. *Fun and Games: a Text on Game Theory*. Heath and Co., Lexington, MA, 1992.

Binmore, K.G. *Game Theory and the Social Contract, II: Just Playing*. MIT Press, Cambridge, MA, 1998.

Boyd, R. and Lorberbaum, J.P. "No pure strategy is evolutionarily stable in the repeated prisoner's dilemma." *Nature*, 327, pp. 58–59, 1987.

Boyd, R. and Riche rson, P.J. "The evolution of indirect reciprocity." *Social Networks*, 11, pp. 213–236, 1989.

Brandts, J. and Schram, A. "Cooperative gains and noise in public goods experiments; applying the contribution function approach." Forthcoming in the *Journal of Public Economics*, 1998.

Cooper, R.; DeJong, D.V.; Forsythe, R. and Ross, T.W. "Cooperation without reputation: Experimental evidence from prisoner's dilemma games." *Games and Economic Behavior*, 12, pp. 187–218, 1996.

Fehr, E. and Gächter, S. "Cooperation and punishment — an experimental analysis of norm formation and norm enforcement." Forthcoming in the *American Economic Review*, 1996.

Fehr, E. and Gächter S. "Reciprocity and economics: The economic implications of *Homo Reciprocans*." *European Economic Review*, 42(3–5):845–859, 1998.

Fehr, E.; Gächter, S. and Kirchsteiger, G. "Reciprocity as a contract enforcement device: experimental evidence." *Econometrica*, 65, pp. 833–860, 1997.

Fehr, E.; Kirchler, E.; Weichbold, A. and Gächter S. "When social norms overpower competition — gift exchange in experimental labour markets." *Journal of Labor Economics*, 16(2), pp. 324–351, 1998.

Fehr, E.; Kirchsteiger, G. and Riedl, A. "Does fairness prevent market clearing? An experimental investigation." *Quarterly Journal of Economics*, 108, pp. 437–460, 1993.

Foster, D. and Young, H.P. "Stochastic evolutionary game dynamics." *Theoretical Population Biology*, 38, pp. 219–232, 1990.

Gächter, S. and Falk, A. "Reputation or reciprocity?" Working paper, University of Zürich, 1997.

Güth, W.; Schmittberger, R. and Schwarze, B. "An experimental analysis of ultimatum bargaining." *Journal of Economic Behavior and Organization*, 3(3), pp. 367–388, 1982.

van der Heijden, E.C.M. *Altruism, Fairness and Public Pensions*. PhD thesis, Tilburg University, The Netherlands, 1996. Chapter 6.

Leimar, O. "Repeated games: a state space approach." *Journal of Theoretical Biology*, 184, pp. 471–498, 1997.

Levine, D. "Modeling altruism and spitefulness in experiments." *Review of Economic Dynamics*, 1, pp. 593–622, 1998.

McKelvey, R.D. and Palfrey, T.R. "An experimental study of the centipede game." *Econometrica*, 60(4), pp. 803–836, 1992.

Nowak, M.A. and Sigmund, K. "The alternating prisoner's dilemma." *Journal of Theoretical Biology*, 184, pp. 471–489, 1993.

Nowak, M.A. and Sigmund, K. "A strategy of win-stay, loose-shift that outperforms tit-for-tat in the prisoner's dilemma game." *Nature*, 364, pp. 56–57, 1994.

Nowak, M.A. and Sigmund, K. "The dynamics of indirect reciprocity." *Journal of Theoretical Biology*, 194, pp. 561–574, 1998a.

Nowak, M.A. and Sigmund, K. "Evolution of indirect reciprocity by image scoring." *Nature*, 393, pp. 573–577, 1998b.

Pollock, G.B. and Dugatkin, L.A. "Reciprocity and the evolution of reputation." *Journal of Theoretical Biology*, 159, pp. 25–37, 1992.

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

Ridley, M. *The Origins of Virtue: Human Instincts and the Evolution of Cooperation.* Penguin Books Ltd, London, 1996.

Roth, A.; Prasnikar, V.; Okuno-Fujiwara, M. and Zamir, S. "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo." *American Economic Review*, 81, pp. 1068–1095, 1991.

Schram, A. "Sorting out the seeking: The economics of individual motivations." Forthcoming in *Public Choice*, 1998.

Selten, R. and Hammerstein, P. "Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of "tit for tat"." *Behavioral Brain Sciences*, 7, pp. 115–116, 1984.

Siegel, S. and Castellan, N.J., Jr. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill Book Company, Singapore, 2^{nd} edition, 1988.

Sugden, R. *The Economics of Rights, Cooperation and Welfare.* Blackwell, Oxford, 1986.

Trivers, R. "The evolution of reciprocal alruism." *Quarterly Review of Biology*, 46, pp. 35–57, 1971.

Wedekind, C. "Give and ye shall be recognized." *Science*, 280(5372), pp. 2070–2071, 1998.

Young, P.H. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton University Press, Princeton, New Jersey, 1998.

Appendix

Instructions experiment

What follows is a translation of a sample of the instructions of the experiment.[15]

Introduction

Welcome to this experiment in decision making. In this experiment you will earn money. At the start of the experiment you will receive 30 guilders. During the experiment you may win or lose money dependent on your own decisions and the decisions of other participants. At the end, your final earnings will be paid to you privately. Your decisions are anonymous. They will not be attached to your name. [...] You are not allowed to speak with other participants or to communicate in any other way. If you want to ask a question, please raise your hand.

[*A short introduction on how to use the computer followed, the instructions continued with the description of the experiment.*]

Design experiment

This experiment consists of at least 90 rounds. From round 90 on a next round will start with a probability of 90%. After the experiment has finished, the computer will choose 20 rounds randomly. Only the earnings from these 20 rounds will be paid to you.

At the beginning of every round the participants will be randomly divided into pairs. The probability to form a pair with a specific other participant is the same for all participants in every round. However, the probability to form a pair twice in a row with the same participant is very small.

One of the two participants will play role A, the other role B. Which role you play is also determined randomly in every round. Only when appointed role A you will have to make a decision, if you have role B, you will not have to do anything that round.

Choices

Being participant A you will have 2 alternatives, which are shown schematically in the table above.

When you choose 'yellow', the participant with role B that you are matched with will get 250 cents and you will loose 150 cents. If you choose 'blue' neither of you will gain or lose money in that round.

	Alternatives	
	yellow	blue
Earnings for A (in cents)	−150	0
Earnings for B (in cents)	250	0

[15] The data and the complete translation of the instructions will be sent by the authors on request.

Information

Before you are asked to make a choice, you will receive information about what the participant you are matched with in this round has chosen in earlier rounds. You will see a summary of the at most 6 most recent decisions that participant B made, when he or she was appointed role A in earlier rounds. This information looks as follows:

Participant B chose as follows in earlier rounds when in role A:
... times yellow ... times blue.

The total number of choices equals at most 6. In case participant B has never been in role A, zeros will appear at the dots in the scheme shown above. Only participant A will get this information, participant B will only see how much (s)he earned in this round.

So, if you are appointed role B, the participant A that you are matched with will also get to see your last 6 decisions when you were in role A. The information about you will not change in the periods in which you have role B, as you will not make any choices then.

At this point subjects had to answer 5 questions testing understanding. The instructions continued with a description of the windows that would appear during the experiment:

*On the left part of each computer-screen a permanent box was shown, with the round number, the own score (a summary of the last 6 choices, consisting of a blue and a yellow number) and an overview of all outcomes so far. The colors of the numbers shown corresponded with the choice (*yellow *or* blue*), the choice-maker was coded by a letter 'Y' (you) and 'O' (other).*

At the beginning of each round a window popped up with a message about which role was appointed to the subject. The A-participant first got a window at the top right with information about former choices of participant B and after a few seconds another window at the bottom right in which (s)he had to make her choice. Every decision had to be confirmed. The B-participant only received a message concerning the outcome of that round.