THE ROYAL SOCIETY

*doi* 10.1098/rspb.2001.1809

# Cooperation through indirect reciprocity: image scoring or standing strategy?

**Manfred Milinski[1]*, Dirk Semmann[1], Theo C. M. Bakker[2] and Hans-Jürgen Krambeck[1]**

[1]*Department of Evolutionary Ecology, Max-Planck-Institute for Limnology, 24306 Plön, Germany*
[2]*Institute for Evolutionary Biology and Ecology, University of Bonn, An der Immenburg 1, 53121 Bonn, Germany*

Theorists have only recently shown that cooperation through indirect reciprocity can evolve. The first modelling approach favoured a mechanism called image scoring. Helping someone increases one's image score, whereas refusing to help reduces it. The evolutionary outcome was a discriminator image scoring strategy that helps everybody who has, for example, a positive image score. Two experimental studies with humans found results that were compatible with discriminator image scoring. However, a new analysis of other theorists, based on another population structure, has cast doubts on the evolutionary stability of strategies using the recipient's score as a sole basis for decision. The new theoretical study confirmed that a strategy aiming at 'good standing' has superior properties and easily beats image scoring. An individual loses good standing by failing to help a recipient in good standing, whereas failing to help recipients who lack good standing does not damage the standing of a potential donor (but would reduce his image score). The present empirical study with 23 groups of seven human subjects each was designed for distinguishing between the two proposed mechanisms experimentally. The results differed strongly from standing strategies, which might demand too much working memory capacity, but were compatible with image scoring or a similar strategy to a large extent. Furthermore, donors of constant 'NO players' compensated for their refusing to help these players by being more generous to others.

**Keywords:** evolution of cooperation; indirect reciprocity; image scoring; good standing

## 1. INTRODUCTION

Understanding how egoists can maximize their fitness by helping unrelated conspecifics has been a long-standing puzzle. Evolutionary theorists have developed two main concepts: direct (Trivers 1971; Axelrod & Hamilton 1981) and indirect reciprocity (Alexander 1987; Zahavi 1991). Under direct reciprocity, an individual helps someone who thereby gains more than the individual's help costs. If the help is reciprocated on the next occasion, each player has a net benefit. The problem is that, if the recipient of an altruistic act refuses to reciprocate, he would gain even more than he would have won had he reciprocated. This situation is analogous to what game theorists call the Prisoner's Dilemma. Over the last 20 years, theorists have searched for evolutionarily stable strategies for the repeated Prisoner's Dilemma and found, unexpectedly, stable cooperation under realistic conditions (Axelrod 1984; Nowak & Sigmund 1992, 1993, 1994; Milinski 1993; Frean 1994; Leimar 1997). Under indirect reciprocity, support is given to individuals who have helped other individuals. 'Give and you shall receive' has been believed to work, but theorists have only recently shown that indirect reciprocity through image scoring can evolve (Nowak & Sigmund 1998*a*,*b*; Lotem *et al*. 1999).

Nowak & Sigmund (1998*a*,*b*) showed, using computer simulations, that the strategy of helping those who have helped others could evolve to fixation in the population. For example, in one computer simulation a population comprised 100 individuals and each had two to three interactions (as a donor or as a receiver) per lifetime. At the beginning of each generation, all players had an image score of zero. Their image scores were increased by each act of helping and decreased by each act of withholding help. The strategies for helping ($k$) ranged from $k = -5$ to $k = +6$. A $k = -5$ strategy helps everybody who has an image score of at least $-5$, which means almost unconditional helping. A $k = +6$ strategy helps everybody who has an image score of at least $+6$, which means almost unconditional defecting. A $k = 0$ strategy is maximally discriminating; it helps everybody who has a zero or positive image score. This most discriminating strategy dominated the population after 150 generations.

Nowak & Sigmund's (1998*a*,*b*) predictions were tested in an experiment with human players in eight groups consisting of 10 subjects each (Wedekind & Milinski 2000). The players could repeatedly give money to others and receive from others, but they were told that they would never meet the same person with reversed roles. As in the model, the donor's cost of giving was smaller than the receiver's benefit. The players were given pseudonames and their history of giving and non-giving was displayed at each interaction. The findings were compatible with an image scoring strategy: receivers who were given money had significantly higher image scores than receivers who were not given money. Those subjects who donated rarely had a strong preference for only giving to receivers with a relatively high image score (Wedekind & Milinski 2000). Seinen & Schram (2001) found similar results in a more detailed experiment.

Leimar & Hammerstein (2001) recently questioned the robustness of Nowak & Sigmund's (1998*a*,*b*) predictions using another population structure. Although they

*Author for correspondence (milinski@mpil-ploen.mpg.de).



*Received* 4 May 2001 *Accepted* 26 July 2001

concluded that indirect reciprocity could evolve in principle, they discussed serious problems with image scoring strategies. They argued that it would not be in an individual's interest to base his decisions partly or wholly on the score of a potential recipient of help. For example, if they followed the image scoring rule by refusing to help an individual with a low image score, they would decrease their own score and suffer from not being helped (see also Nowak & Sigmund 1998*a*). A rational individual should increase his own score, but ignore the score of a potential recipient. Leimar & Hammerstein (2001) found that the evolution of cooperative image scoring could only occur under restrictive conditions that consist of either a substantial influence of genetic drift or a very small cost of helping.

Leimar & Hammerstein (2001) found that Sugden's (1986) strategy of aiming for 'good standing' has superior properties. It can be an evolutionarily stable strategy and, even if not, it usually beats image scoring. In Sugden's (1986) model, everyone is initially in good standing. An individual loses good standing by failing to help a recipient in good standing, whereas failing to help recipients who lack good standing does not damage the standing of a potential donor. A crucial difference between this strategy and the image scoring strategies is, as Leimar & Hammerstein (2001) noted, that, in a population playing the standing strategy, it is in an individual's interest to react to the standing of a potential recipient. Leimar & Hammerstein (2001) showed that the strategy has additional robustness properties and that it can invade a population of image scorers. They concluded that the standing strategy appears to be a viable candidate mechanism for human cooperation based on indirect reciprocity. Leimar & Hammerstein (2001) argued that Wedekind & Milinski's (2000) experiment left open the extent to which a refusal to help an unhelpful individual is held against a potential donor, in comparison with a refusal to help a helpful individual.

Nowak & Sigmund (1998*b*) had already suggested 'that Sugden's strategy is a good approximation to how indirect reciprocity actually works in human societies: it offers, as Sugden remarks, a workable insurance principle. But as stressed in Boerlijst *et al.* (1997) in connection with Contrite Tit For Tat, strategies based on standing are prone to be affected by errors of perception. If information is incomplete, then a player observed while withholding his help may be misunderstood; he may have defected on a player with good standing, or punished someone with bad standing. An eventual error can spread. The discriminator rule is less demanding on the player's capabilities, and still works. We expect that in actual human communities, indirect reciprocity is based on more complex reckonings, and believe that this should be amenable to experimental tests' (p. 573). Leimar & Hammerstein (2001) included errors of perception in their simulations and concluded that '. . . it seems that the standing strategy also has some robustness when there are errors in perception' (p. 750). However, the error probability was smaller than that considered by Nowak & Sigmund (1998*a*) (where only 10 people watch an interaction in a population of size 20–100).

The present study was designed for distinguishing between the two proposed mechanisms of indirect reciprocity, i.e. discriminator image scoring and standing strategies, experimentally. We studied 23 groups comprising seven players each. However, one of the players had been secretly instructed to adopt a fixed strategy: he always refused to give (the 'NO player'). Donors of NO players should refuse to give anything to the NO player right from the beginning, both if they use a discriminating image scoring strategy and if they use a standing strategy and found out that the NO player had no justification for non-giving. The donors of the donors of the NO player should refuse to give to donors of NO players if they use an image scoring strategy. However, they should ignore all the acts of non-giving to the NO player if they use a standing strategy, because these acts were justified (figure 1). The probability of the donors of the NO player of receiving NO themselves is thus the test case for discriminating between image scoring and standing strategies.

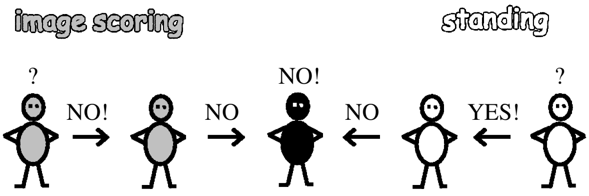


Figure 1. Experimental set-up for distinguishing between image scoring and standing strategies. See § 1 for further explanation.

Nowak & Sigmund (1998*a*) considered strategies that cooperate if the image score of an individual's opponent is at least $k$ and if the individual's own image score is less than $h$. The most frequent strategy in their simulations was $k = 0$ and $h = 4$. Since donors of the NO player have a low image score from refusing to support the NO player, their $h$-value might often be below the threshold. Therefore, they might be more generous to other players than are, for example, receivers of the NO player. If the donors of the NO player adopt an image scoring strategy then in this way they might compensate for their low image score. If donors of the NO player adopt a standing strategy they need not compensate for refused support to a NO player because these NOs would not reduce their standing.

Since the limited working-memory capacity of the players might constrain their choice of strategy (Milinski & Wedekind 1998) such that they are unable to adopt a standing strategy, which may require a large memory capacity, we performed two experiments. In one experiment, we only provided information about the potential receivers' history of giving or non-giving ('little information'), as in previous studies (Wedekind & Milinski 2000; Seinen & Schram 2001). In the other experiment, we provided additional information about the receiver's previous receivers' history of giving or non-giving ('much information'). In this case, the donors could read whether a NO to their present receiver had been justified. We would expect that the players in the experiment with much information would have a higher probability of adopting a standing strategy than the players in the experiment with little information.

## 2. METHOD

The experiment consisted of 23 sessions at the Universities of Bonn and Hamburg. The participants were students from biology, economics, medicine and social sciences. Seven subjects participated simultaneously in each session. They were randomly assigned to a row of seats that were separated by partitions. They could see a screen on which instructions and the actual game was projected. The whole experiment was computerized.

The experiments were controlled and documented by a PC executing an application program under WINDOWS 98. Communication between the seven players and the computer was performed using boxes, each of which contained two silent switches for YES and NO and a lamp indicating an expected answer when lit. The application was written (by H.J.K.) in DELPHI 5, which is an object-orientated programming language based on Borland PASCAL and distributed by Borland Inc. (Scotts Valley, CA, USA). The program performs the following tasks: display of the introduction texts, random assignment of the players, managing the questions, recording and displaying of the answers and calculating basic data. The program keeps track of all events and stores information such as who gives, who receives and which key is pressed at what time in a log file. Thus, the whole experiment was documented completely.

Before the program was started the students were told (i) that they had to be naive with respect to the hypotheses, but would be sent full documentation of the hypotheses and results, (ii) that they had to be anonymous and, thus, would be given a pseudo-name by the program, (iii) that it was very important that they also maintained their anonymity after the game and did not talk to others (potential later participants) about the game, (iv) that the two experimenters must not know the players' identity either and would therefore sit behind blinds, and (v) that the money that they gained or lost in the game was theirs and would be paid anonymously in cash after the game. No communication was allowed.

The computerized instructions, which were presented at the pace of the slowest reader, gave the subjects the following information.

(i) That each person had a starting account of DM25 (*ca.* US$12) and could gain more money or lose some dependent on his/her and the participants' decisions.
(ii) That all decisions were anonymous.
(iii) How to use the YES or NO silent switches in front of each player.
(iv) That there were two roles that were randomly but repeatedly allocated to the players, i.e. that of a potential donor (active) and that of a potential receiver (passive).
(v) That all players would be assigned equally often to either role.
(vi) That the same two names would never meet with reversed roles, i.e. direct reciprocity was excluded.
(vii) That each player would be assigned a pseudo-name (these names were moons of our solar system, e.g. Metis, Ananke, Kallisto, Despina, Japetus, Telesto and Galatea); 'you will have a new identity, you will be, for example, Portia during this game'.
(viii) That a potential donor, say Triton, would be asked whether he would give to Portia; Triton would lose DM2.50 from his account and Portia would gain DM4 in her account if Triton decided YES, whereas Triton would not lose anything and Portia would not gain anything if Triton decided NO.
(ix) That, in the game (in groups with little information), Triton (and all other players) would be provided with information on whether Portia had given in previous rounds when she had been in the role of the potential donor (see table 1) or that (in groups with much information) additional information would be provided about the giving or non-giving of Portia's receivers when she gave or did not give (table 1).
(x) That Triton's answer (YES or NO) would be displayed for *ca.* 2s.

Table 1. Information on public screen.

(As an example, we are in round 4 of the game with much information where Triton is the potential donor and has to decide whether he gives DM2.50 from his account to Portia who is the potential receiver. The following information is projected to the screen (except for the information given in parentheses).)

donor: Triton
receiver: Portia

| round | did Portia give? | did the receivers of Portia give in preceding rounds? |
|---|---|---|
| 1 | yes | N (i.e. no in round 1) |
| 2 | no | NY (i.e. no in round 1 and yes in round 2) |
| 3 | no | YNY (i.e. yes in round 1, no in round 2 and yes in round 3) |

The information indicated in table 1 as well as the decision of the potential donor was projected on the public screen and could be seen by all of the players.

Thereafter, the game began and lasted for 16 rounds. The subjects did not know this. Thus, we had 16 decisions from each subject as a donor and 16 possibilities of each subject receiving something. A round consisted of the seven decisions of the subjects in the group in a randomized sequence. Pairs comprising a potential donor and a potential receiver were randomly chosen by the computer program. Since direct reciprocity was excluded, each player was actually allocated three pseudo-names as a potential donor and three other pseudo-names as a potential receiver for the whole game. Thus, the NO player also had three potential donors and three potential receivers. There may be cycles (A gives B, B gives C, C gives A, etc.) that are quite short. Indirect reciprocation based on such loops has been studied theoretically by Boyd & Richerson (1989). However, the average player would be unlikely to be aware of these loops.

Sessions with much information (m) were alternated with sessions with little information (l) in such a way that time and sequence effects could not bias the results (e.g. lmml on one day and mllm on the next day). There was a student in each group who had been secretly instructed by us to always decide NO, i.e. the NO player. In order to avoid pseudo-replication, our statistical unit was the group, thereby reducing our $n$ to 12 for experiments with much information and to 11 for experiments with little information. Thus, $n$ is the number of groups in our analyses. All choices in the statistical analyses, i.e. the probabilities of giving or receiving NO, were arcsine transformed in order to meet the requirements of parametric statistics. All of the statistical probabilities given are two-tailed.

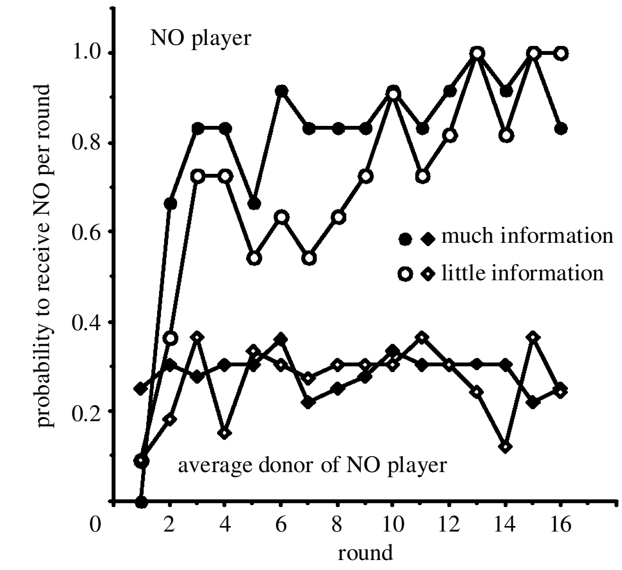


Figure 2. Probabilities of a NO player receiving NO (circles) and a donor of the NO player receiving NO (diamonds) in rounds 1–16. The means of 12 groups with much information (filled symbols) and of 11 groups with little information (open symbols) are shown. Each group has one NO player and three donors of the NO players, which are entered in the analysis as one mean value in order to avoid pseudo-replication.

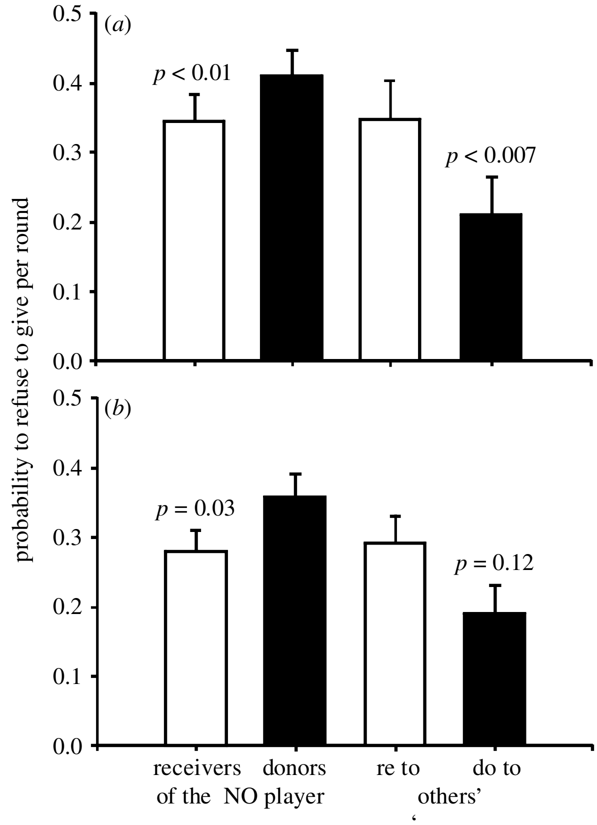


Figure 3. Mean ( + s.e.) probabilities of receivers (open bars) and donors (filled bars) of the NO player refusing to give per round. The three receivers and three donors of the NO player per group are entered as mean values, respectively, in the analysis. 'Others' are individual receivers of both receivers and donors of the NO player. 'Re to others' is the mean ( + s.e.) probability of the receivers of the NO player refusing to give to 'others' and 'do to others' is the mean ( + s.e.) probability of the donors of the NO player refusing to give to 'others'. (*a*) The mean values of 12 groups with much information and (*b*) the mean values of 11 groups with little information ($p$-values after paired $t$-tests) (see the text).

## 3. RESULTS

The groups with much information needed significantly more time (mean + s.e. = 23.75 + 2.23 min) for the actual game than did the groups with little information (16.27 + 0.81 min) (d.f. = 21, $t = 3.11$ and $p = 0.005$), thereby suggesting that the players took notice of the additional information provided. NO players received NOs with increasing probability during the 16 rounds. This probability increased more quickly with much information than with little information (figure 2) (much information and little information for rounds 1–8, 0.7 + 0.15 and 0.54 + 0.14, respectively and unpaired $t$-test, d.f. = 21, $t = 2.80$ and $p = 0.01$ and much information and little information for rounds 9–16, 0.91 + 0.12 and 0.88 + 0.10, respectively and unpaired $t$-test, d.f. = 21, $t = 1.03$ and $p = 0.31$). This suggests that donors of the NO player discovered more quickly that the NOs of the NO player were not justified with much information. There was a trend of donors of the NO player to receive more NOs during the first eight rounds with much information than with little information (figure 2) (much information and little information for rounds 1–8, 0.29 + 0.04 and 0.22 + 0.04, respectively and unpaired $t$-test, d.f. = 21, $t = 1.09$ and $p = 0.29$ and much information and little information for rounds 9–16, 0.29 + 0.05 and 0.28 + 0.04, respectively and unpaired $t$-test, d.f. = 21, $t = 0.17$ and $p = 0.86$).

### (a) *Donors of the NO player compensate for NOs to the NO player by fewer NOs to 'others'*

The donors of the NO player had a significantly higher probability of giving a NO response than the receivers of the NO player in both games with much information (figure 3*a*) (paired $t$-tests, $n = 12$ pairs, $t = 3.16$ and $p < 0.01$) and games with little information (figure 3*b*) (paired $t$-tests, $n = 11$ pairs, $t = 2.51$ and $p = 0.03$). In order to avoid pseudo-replication the three receivers and the three donors of the NO player per group are entered as mean values in the analysis, respectively. For the analysis we formed a new group which we called 'others'. Players in the group 'others' were not the NO player and they had to receive from donors of the NO player as well as from receivers of the NO player. For example, if an 'others' player, e.g. Metis, was three times a receiver of a donor of the NO player, e.g. Ananke, and four times a receiver of a receiver of the NO player, e.g. Kallisto, the mean probabilities of receiving NO in these two kinds of interaction were combined with the respective values from other 'others' players in order to provide the two respective mean values of that group. In the example, Metis might have been a donor of the NO player herself. Donors of the NO player gave significantly fewer NO responses to 'others' than did receivers of the NO player in games with much information (figure 3*a*) (paired $t$-tests, $n = 12$ pairs, $t = 3.32$ and $p < 0.007$), whereas this result was not significant in games with little information (figure 3*b*) (paired $t$-tests, $n = 11$ pairs, $t = 1.69$ and $p = 0.12$).

**(b) *Testing for image scoring***

The measured probabilities of donors of the NO player receiving NO per round of the 12 groups with much information and the 11 groups with little information are compared with the values that are expected if the players adopt an image scoring strategy. The expected value is determined by calculating the probability of receiving NO of the average (of each group) receiver of the NO player as if they have given the number of NOs that the average (of each group) donor of the NO player has given (including those NOs given to the NO player) (i.e. expected percentage of NOs for a donor of the NO player/percentage of NOs given by a donor of the NO player = percentage of NOs received by a non-donor of the NO player/percentage of NOs given by a non-donor of the NO player). The NOs that the receivers of the NO player received from the NO player were not taken into account because they had been set by default. Donors of the NO player tended to receive fewer NOs than expected from image scoring in games with much information (figure 4*a*) (paired *t*-tests, $n = 12$ pairs, $t = 1.59$ and $p = 0.14$). There was no such trend in games with little information ($n = 11$ pairs, $t = 0.14$ and $p = 0.89$). The difference between the differences (between the measured and expected values for much and little information) would be evidence for a standing strategy if it were significant (unpaired *t*-test, d.f. = 21, $t = 0.73$ and $p = 0.48$).

**(c) *Testing for standing strategies***

The measured probabilities of donors of the NO player receiving NO per round of the 12 groups with much information and the 11 groups with little information are compared with the values that are expected if the players adopt a standing strategy. The expected value is determined by calculating the probability of receiving NO from the average receiver of the NO player (of each group) as though they have given the number of NOs that the average donor of the NO player (of each group) has given (treating the NOs given to the NO player as YES). The NOs that the receivers of the NO player received from the NO player were not taken into account because they had been set by default. Donors of the NO player received significantly more NOs than expected from standing in both games with much and little information (figure 4*b*) (paired *t*-tests, $n = 12$ pairs, $t = 8.79$ and $p = 0.0001$ and $n = 11$ pairs, $t = 4.94$ and $p = 0.0006$, respectively).

An alternative way of calculating expected values for standing would simply be to exclude these rounds when computing the percentage of NOs given by the donor of the NO player instead of treating NOs given to the NO player as YES, since the player should neither lose nor gain from giving NO in these rounds according to the standing strategy (O. Leimar, personal communication). The new predicted values of 0.19 + 0.020 for games with much information and 0.14 + 0.043 for games with little information were still significantly different from the measured values that are shown in figure 4*b* ($p = 0.0003$ and $p = 0.043$, respectively).

**(d) *Pay-offs***

The winners were the donors of the NO player. They earned DM 47.06 + 1.30 in games with much information and DM 46.50 + 1.47 in games with little information.

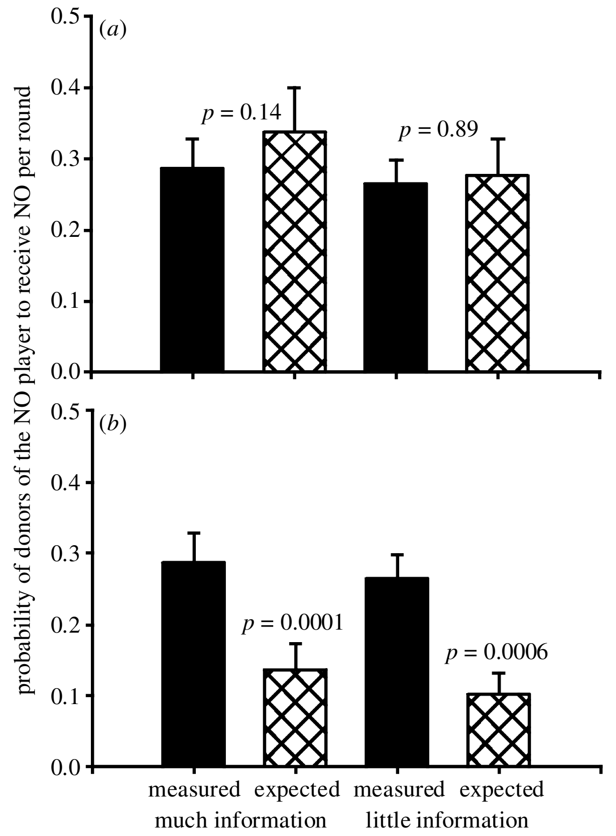


Figure 4. Mean (+ s.e.) probabilities of donors of NO players receiving NO per round of the 12 groups with much information and the 11 groups with little information. (*a*) Expectation for image scoring: measured probabilities (filled bars) are compared with expected probabilities (hatched bars). (*b*) Expectation for standing: measured probabilities (filled bars) are compared with expected probabilities (hatched bars) (*p*-values after paired *t*-tests) (see the text).

The NO players earned less (DM 37.67 + 1.76 and 43.82 + 1.77, respectively) and the receivers of the NO player least (DM 28.69 + 0.93 and 29.83 + 1.19, respectively). The last group suffered from always receiving NO from the NO player.

In summary, it can be concluded that the players took notice of the additional information provided because the groups with much information needed significantly more time for the actual game than did the groups with little information. Donors of the NO player had a significantly higher probability of refusing help than other players in both games with much and little information. They ceased helping more quickly in games with much information. Donors of NO players compensated for their refusal to help these players by being more generous to others. Donors of the NO player did not receive significantly fewer NOs than expected from image scoring. However, they received significantly more NOs than expected from standing in both games with much and little information.

## 4. DISCUSSION

There is now substantial experimental evidence for cooperation through indirect reciprocity in groups of human players (Wedekind & Milinski 2000; Seinen & Schram 2001; this study). The mechanism by which indirect reciprocity is achieved is a matter of debate among theorists: Nowak & Sigmund (1998*a*,*b*) proposed image scoring as a first candidate, whereas Leimar & Hammerstein (2001) concluded that image scoring can only evolve under restricted conditions and is usually beaten by Sugden's (1986) standing strategies. It seems to us that Leimar & Hammerstein's (2001) predictions of standing strategies hold as long as human players are unconstrained by errors in perception (Boerlijst *et al.* 1997) or their usually limited working memory capacity (Baddeley 1987; Milinski & Wedekind 1998). In adopting a standing strategy, a large amount of second- if not third- and fourth-order information about the history of the social interactions of many potential receivers of help has to be stored and used. For example, in order to decide whether a potential receiver is in 'good' standing after he had given help once and refused help three times, one needs to know whether his last three receivers had been in good or bad standing which implies that they must also know whether their receivers had been in good or bad standing, which implies.... If constraints placed by our limited memory capacity are inevitable, a less demanding strategy such as discriminator image scoring may have evolved.

The present study was designed for distinguishing experimentally between image scoring and standing strategies. In order to facilitate the use of standing strategies in 12 groups of players, we provided each potential donor with first- and second-order information about their actual potential receiver. Both this information and the decision of the potential donors were projected onto a big screen that could be seen by all the players. The donor could thus 'read' whether a previous NO of their present receiver had been justified. This information should have facilitated the use of standing strategies, whereas it is not needed for image scoring. We provided only first-order information in the other 11 groups, as did previous studies (Wedekind & Milinski 2000; Seinen & Schram 2001). Obviously, the players took notice of the second-order information because the 16 rounds of the game lasted for *ca*. 50% longer than those with only first-order information provided.

The trick in distinguishing image scoring from standing was to have a secretly predetermined player in each group of seven players who always decided NO (the NO player). The NO players themselves received NO in most cases, which is compatible with either mechanism. The donors of NO players should thus have a very low image score and rarely receive help from strategies based on image scoring. However, if the players use a standing strategy the donors of NO players should receive a lot of help because their standing is not reduced as a result of refusing to help the NO player, who is in bad standing. We found a strong deviation from the expectation generated by standing. The donors of NO players received significantly more NOs than expected in both groups with much and little information provided. On the other hand, we found no significant deviation from the expectation generated by image scoring. The donors of NO players suffered almost as much from their NOs as if these NOs had not been justified. Our results thus support the image scoring idea or a strategy that appears to be similar to image scoring in our experiments. Standing strategies might be too demanding to be realized with respect to memory capacity.

There may be a simpler way of adopting standing strategies (A. Lotem, personal communication): players have to keep track of the current standing status of other players and need to update this status in each round based on the preceding round and, after updating, forget all previous rounds. This strategy requires only first-order information, but it may be prone to errors of perception. The information that we provided might actually have hindered the use of this simple standing strategy. If so, indirect reciprocity should have disappeared. The fact that we found strong indirect reciprocity suggests that our players used a different strategy.

If the image score of the recipient is important in deciding whether to help or not, then a donor might consider the effect that their own score will have on future donors. Seinen & Schram's (2001) study included an experiment in which donors were not told anything about the previous choices of the recipient. Seinen & Schram (2001) found the lowest probability of helping under this 'no-information condition' and concluded that 'subjects are much more likely to help if they know that their score is passed on' (p. 11). This can be regarded as support for an image scoring-like strategy.

Interestingly, the donors of NO players in the present study seemed to compensate somewhat for the many NOs given to the NO player by giving significantly fewer NOs to other players than those received from other players. This compensation would not be necessary if the players used a standing strategy because the NOs given to the NO player would not reduce their standing. The compensation may have helped the donors of the NO player to achieve a reasonable pay-off. The losers of the game were the players who always received NO from the NO player.

There are two hints pointing to a standing strategy.

(i) When much information was provided, the NO player's rate of receiving NOs increased more quickly during the first rounds than when only little information was provided. With much information, the donors of the NO player could already 'read' that the NO player was in bad standing in the second round. With little information, they were perhaps not sure about the standing of the NO player and offered more help than the NO player deserved.

(ii) With much information provided, we found a trend that donors of the NO player received fewer NOs than were expected from image scoring, whereas there was no such trend in groups that obtained only little information.

Similar results were independently found by Bolton *et al.* (2002): in one treatment, players were also provided with second-order information (the mover knows not only whether the receiver last played 'give' or 'keep', but also knows whether the receiver last played 'give' or 'keep' with someone who last played 'give' or 'keep') for one previous round (not for all rounds as in the present study). They found that facing a cooperative receiver (with a history 'give'/'give' or 'give'/'keep', meaning that the receiver gave to a giver or gave to a keeper, respectively) increased, and facing a selfish receiver (with history 'keep'/'give' or 'keep'/'keep') decreased the probability that the mover gives, which agrees with image scoring. Note that 'keep'/'keep' is justified compared with 'keep'/'give'; these situations were treated marginally differently (two-tailed $p = 0.07$) (G. Bolton, personal communication), which again is a hint pointing to a standing strategy.

It seems as though our human players were not prepared to adopt a standing strategy because image scoring also predominated when we provided them with the information that is needed for adopting a standing strategy. Even if our test situation was unnatural in many respects that prevented us from finding standing, it is unlikely that humans only use image scoring in unnatural situations. However, image scoring was not likely to evolve in Leimar & Hammerstein's (2001) models. Under certain conditions, Leimar & Hammerstein's (2001) results are in good agreement with Nowak & Sigmund's (1998*a*) findings. The difference between these two theoretical studies comprises the population parameters and the rate of possible 'errors of perception'. These parameters should be measured in human populations that are similar to those of our predecessors.

We thank the students from the Universities of Bonn and Hamburg for participating in our experiment, J. Ganzhorn, M. Gewecke, J. Krink, J. Parzefall, K.-P. Sauer and H. Strutz for logistic help at Hamburg and Bonn, O. Leimar and P. Hammerstein and G. E. Bolton for showing us their unpublished studies and G. E. Bolton, P. Hammerstein, O. Leimar, A. Lotem, I. Seinen, K. Sigmund and a referee for helpful comments on our manuscript.

## REFERENCES

Alexander, R. D. 1987 *The biology of moral systems*. New York: De Gruyter.

Axelrod, R. 1984 *The evolution of cooperation*. New York: Basic Books.

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

Baddeley, A. 1987 *Working memory.* Oxford, UK: Clarendon Press.

Boerlijst, M. C., Nowak, M. A. & Sigmund, K. 1997 The logic of contrition. *J. Theor. Biol.* **185**, 281–293.

Bolton, G. E., Katok, E. & Ockenfels, A. 2002 What is in a reputation? Indirect reciprocity in an image scoring game. *Econometrica*. (Submitted.)

Boyd, R. & Richerson, P. J. 1989 The evolution of indirect reciprocity. *Soc. Networks* **11**, 213–236.

Frean, M. R. 1994 The Prisoner's Dilemma without synchrony. *Proc. R. Soc. Lond.* B **257**, 75–79.

Leimar, O. 1997 Repeated games: a state space approach. *J. Theor. Biol.* **184**, 471–498.

Leimar, O. & Hammerstein, P. 2001 Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond.* B **268**, 745–753. (DOI 10.1098/rspb.2000.1573.)

Lotem, A., Fishman, M. A. & Stone, L. 1999 Evolution of cooperation between individuals. *Nature* **400**, 226–227.

Milinski, M. 1993 Cooperation wins and stays. *Nature* **364**, 12–13.

Milinski, M. & Wedekind, C. 1998 Working memory constrains human cooperation in the Prisoner's Dilemma. *Proc. Natl Acad. Sci. USA* **95**, 13 755–13 758.

Nowak, M. A. & Sigmund, K. 1992 Tit for tat in heterogeneous populations. *Nature* **355**, 250–253.

Nowak, M. & Sigmund, K. 1993 A strategy of win–stay, lose–shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58.

Nowak, M. A. & Sigmund, K. 1994 The alternating Prisoner's Dilemma. *J. Theor. Biol.* **168**, 219–226.

Nowak, M. A. & Sigmund, K. 1998*a* Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577.

Nowak, M. A. & Sigmund, K. 1998*b* The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561–574.

Seinen, I. & Schram, A. 2001 *Social status and group norms: indirect reciprocity in a helping experiment*. Amsterdam: Tinbergen Institute.

Sugden, R. 1986 *The economics of rights, co-operation and welfare*. Oxford, UK: Blackwell.

Trivers, R. L. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.

Wedekind, C. & Milinski, M. 2000 Cooperation through image scoring in humans. *Science* **288**, 850–852.

Zahavi, A. 1991 Arabian babblers: the quest for social status in a cooperative breeder. In *Cooperative breeding in birds; long term studies in behaviour and ecology* (ed. P. B. Stacy & W. D. Koenig), pp. 105–130. Cambridge University Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.