Current Biology, Vol. 12, 1012–1015, June 25, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0960-9822(02)00890-4

The Long-Term Benefits of Human Generosity in Indirect Reciprocity

Claus Wedekind[1] and Victoria A. Braithwaite
Institute of Cell, Animal, and Population Biology
University of Edinburgh
West Mains Road
Edinburgh EH9 3JT
Scotland
United Kingdom

Summary

Among the theories that have been proposed to explain the evolution of altruism [1–7] are direct reciprocity [8–11] and indirect reciprocity [12–21]. The idea of the latter is that helping someone or refusing to do so has an impact on one's reputation within a group. This reputation is constantly assessed and reassessed by others and is taken into account by them in future social interactions. Generosity in indirect reciprocity can evolve if and only if it eventually leads to a net benefit in the long term. Here, we show that this key assumption is met. We let 114 students play for money in an indirect and a subsequent direct reciprocity game. We found that although being generous, i.e., giving something of value to others, had the obvious short-term costs, it paid in the long run because it builds up a reputation that is rewarded by third parties (who thereby themselves increase their reputation). A reputation of being generous also provided an advantage in the subsequent direct reciprocity game, probably because it builds up trust that can lead to more stable cooperation.

Results and Discussion

Twelve separate groups of university students each played three different sessions. The first session, a repeated simultaneous two-player Prisoner's Dilemma [10, 22] (PD; a direct reciprocity game) was a practice session only. From then on, the students played for money. The second session was an indirect reciprocity game. There are several ways to implement a reputation that corresponds with the degree of generosity [14, 15, 17, 23, 24]. We chose to use an image score [14, 18] that was graphically displayed. In the third session, the students played the PD again but with the important modification that their last image score from the indirect reciprocity session was displayed. Given that a reputation of being generous may build up trust [25], this last session was to test whether the generosity displayed in indirect reciprocity is rewarded in subsequent direct reciprocity games.

The Indirect Reciprocity Game

The proportions of giving differed between the groups (ANOVA, $F_{11,102} = 4.41$, $p < 0.0001$) but were in general high (group means range from 38% to 79%). As a consequence, the mean image score of the players increased from the first round on and reached a total average of 3.39 (SE = 0.31) at the end of the game. Whether or not the personal account was displayed did not significantly influence the players' decisions (nested ANOVA including group as a factor, effect of displaying account: $F_{1,102} = 0.31$, $p = 0.57$).

The receivers' image score had an influence on the donors' decisions: in all 12 groups, receivers who got something had on average a higher image score than receivers who got nothing (Figure 1A; repeated measures ANOVA, effect of giving or not giving: $F_{1,99} = 35.0$, $p < 0.0001$; interaction with group: $F_{11,99} = 1.9$, $p = 0.04$). The donors' decisions were also influenced by their own image score: in 11 of 12 groups, donors with a low image score were more likely to donate something and thereby improve their own image score than donors with a relatively high image score (Figure 1B; effect of giving or not giving: $F_{1,99} = 217.2$, $p < 0.0001$; interaction with group: $F_{11,99} = 2.9$, $p = 0.003$).

Overall, players with a high mean image score earned more money in the indirect reciprocity game than players with low image score (Figure 2A). This is largely because the mean payoff per group increased with the groups' average generosity ($r = 0.93$, $n = 12$, $p < 0.0001$), and both less generous (Figure 2B) and more generous players (Figure 2C) profited from this group effect. Within groups, the correlation coefficients between the players' mean image score and their final account ranged from -0.78 to $+0.52$ and was, on average, not significantly different from zero (one-sample signed rank test, $p = 0.57$).

Building up a high image score has immediate costs that were apparent in the first few rounds of the indirect reciprocity games (Figure 2D). However, the donors' tendency to reward high image scores increasingly compensated for the costs of building and maintaining these high image scores. From the tenth round on, the correlation coefficient between image score and account was positive in sign, and, in the last rounds of this session, this positive correlation was statistically significant (Figure 2D).

Carry-Over Effects to the Direct Reciprocity Games

Players who won the direct reciprocity game, i.e., who received the additional £5.00 reward, had on average a higher mean image score during the indirect reciprocity game than players who did not win (Figure 3; $F_{1,112} = 8.15$, $p = 0.005$). The analogous pattern could be observed within the groups: in 10 of the 12 groups, winners in the PD had on average a higher mean image scores than losers (one-sample signed rank test, $p = 0.009$).

In order to describe the players' strategy in the PD, we use four parameters: P_{cc}, P_{cd} (i.e., the probability of playing c after self's c and partner's d), P_{dc}, and P_{dd}. Players with a high image score played the PD differently

[1]Correspondence: c.wedekind@ed.ac.uk

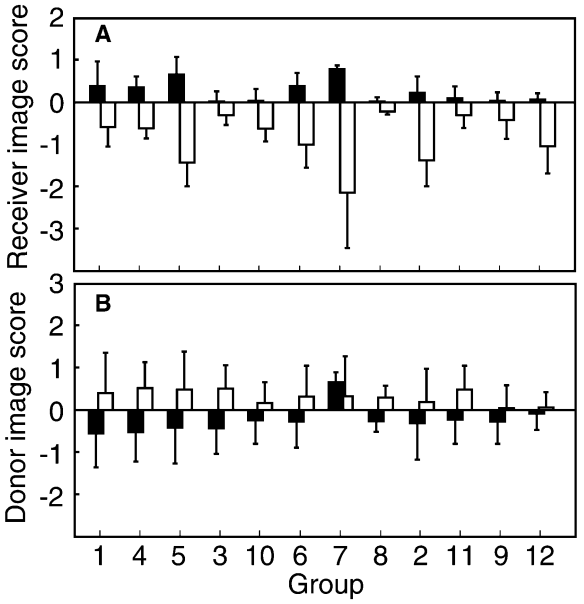

Figure 1. Influence of Image Score, i.e., Reputation, on the Decisions in the Indirect Reciprocity Game

The figure shows the image scores of (A) the receivers and of (B) the donors before the donor gives something (filled bars) or does not give something (open bars). To correct for round effects, the scores shown here are the participants' average deviations from the mean image scores of the respective group and round. The figure gives the means ± SE. The groups are ordered according to their average generosity (group 12 is highest).

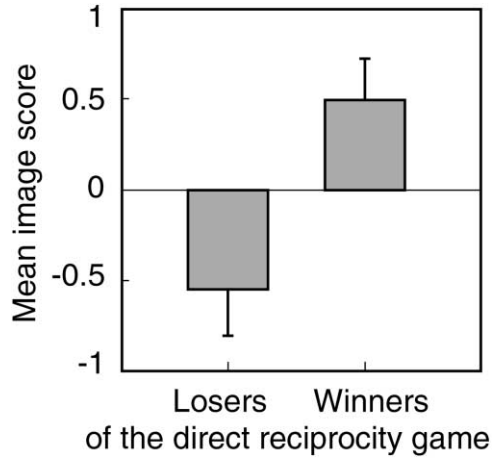

Figure 3. The Mean Image Score during the Indirect Reciprocity Game of Winners and Losers of the Subsequent Direct Reciprocity Game

The image scores plotted here are residuals (means ± SE) that correct for round effects (as in Figure 1).

than players with a low image score (multiple regression on mean image score, with P_{cc}, P_{cd}, P_{dc}, and P_{dd} as predictors: $F_{global} = 2.50$, d.f. = 4, $p = 0.047$). This was mainly because players with high image score played higher P_{cc} ($r = 0.19$, $p = 0.028$) and higher P_{dc} ($r = 0.20$, $p = 0.013$), while P_{cd} and P_{dd} did not correlate with mean image score ($r = 0.04$ and 0.15, p always > 0.05). To test whether the players' strategy in the PD depended on their own and their partner's image score from the previous indirect reciprocity game, we grouped all players as either "generous" and "nongenerous" as described in Figures 2B and 2C. Generous players played more cooperatively in the PD game if their partner was generous than if he/she was not (paired t test, $t_{63} = 3.22$, $p = 0.002$), and they achieved higher payoffs when playing with generous rather than nongenerous partners ($t_{63} = 3.21$, $p = 0.002$). The same was true for nongenerous players who were also more cooperative ($t_{48} = 2.87$,

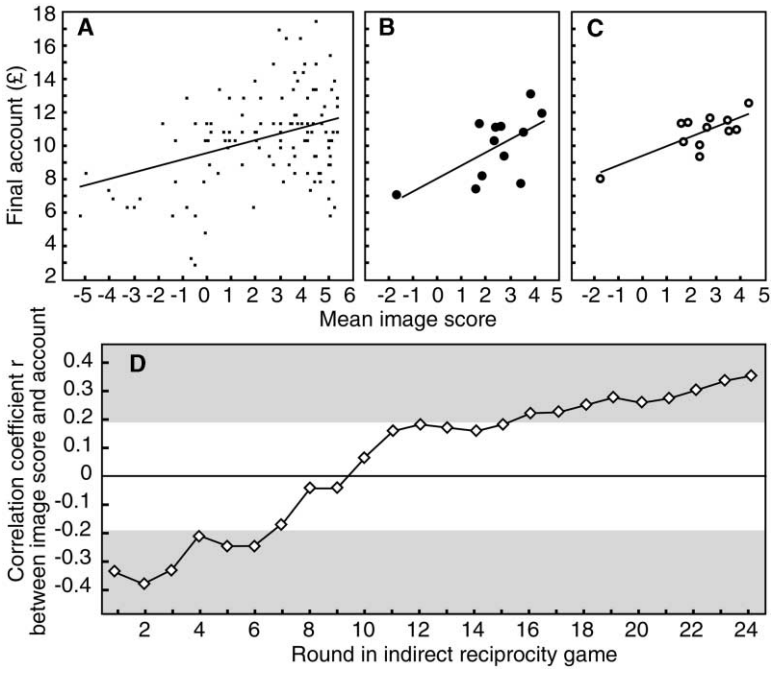

Figure 2. Costs and Benefits of Image Score within the Indirect Reciprocity Game

Regressions between the final account and (A) the mean image score of each player ($r = 0.35$, $p = 0.0001$). In (B) and (C), the regressions are shown separately for the group means (n = 12): (B) for the less generous players of the groups (i.e., mean image score < group mean; $r = 0.60$, $p = 0.04$) and (C) for the more generous players of the groups (mean image score ≥ group mean; $r = 0.76$, $p = 0.004$). (D) The Pearson's correlation coefficients r between image score and account, plotted separately for each round of the indirect reciprocity game. Correlation coefficients within the shaded area are significantly different from zero at $p < 0.05$, two-tailed.

$p = 0.006$) and achieved higher payoffs ($t_{48} = 3.93$, $p = 0.0003$) when playing with generous than with nongenerous players. As a consequence, pairs of generous players achieved a mean payoff of 3.5 (SE = 0.06) points, while pairs of nongenerous players reached a mean of only 3.0 points (SE = 0.07; $F_{1,112} = 21.5$, $p < 0.0001$).

Nonreciprocal altruism among non-kin is frequently observed in humans. Such "generalized altruism" [26] could, for example, be a cultural trait [27], or it could have evolved because it normally provides a net fitness benefit [28]. Indirect reciprocity is one of the major evolutionary concepts that could explain generous behavior. We have verified the two key predictions from the indirect reciprocity models. (1) Generosity builds up some kind of reputation that is later rewarded by third parties. (2) Building up the reputation of being generous is, on average, adaptive behavior in groups with many social interactions.

The generosity we observed in our experiments was on average higher than expected from theory [14, 15, 17] but nevertheless rewarded by third parties to an extent that eventually exceeded the costs of the donations. Such effects can only be seen in groups that have many social interactions. The two types of rewards that we observed, i.e., from indirect and from direct reciprocity games, may lead to some sort of assortative group formation with respect to willingness and ability for generous behavior.

Indirect reciprocity could be a kind of social glue that keeps individuals together in a cooperative network. Alexander [12] argues that this is the basis of moral systems. A system of social norms would decide how an individual's behavior is translated into reputation that may then be rewarded in any kind of currency [13]. Generous players may not be aware of the fact that their generosity may be self-interested and strategic [12], but this is a necessary process for generosity in indirect reciprocity to be evolutionary stable [14–17].

Experimental Procedures

Participants and Experimental Setup

The subjects were biology students, mainly first year undergraduates (53% females), who were asked to voluntarily sign in for experimental dates. They had never heard about indirect reciprocity in their courses at the University of Edinburgh. We tested them in groups of nine or ten. They were told that they would play anonymously and that their total earnings would be paid out in a way that would not reveal their identification number (ID) to us or to their colleagues [18].

Each subject chose a plug to connect an opaque box to an impenetrable tangle of cables, chose a seat within opaque partitions that separated the players from each other, and placed their hands in their box in which they could secretly push two different buttons. These buttons were connected by the tangle of cables to a switchboard handled by one operator. The switchboard was used to connect players one at a time to a red and a green lamp that were visible to everyone. To reveal a choice, a player pushed a button before the operator at the switchboard connected that player to the lamps. One of these two lamps then lit up.

In order to learn their ID, each player drew a piece of paper with a unique sequence of four colors (red and/or green) from a pot and read it in secret. The operator announced an ID number (between one and ten) and switched the respective connection to the lamps on and off four times in a row as the players pressed out their sequence. Each player realized their ID when they saw their code sequence flashed out by the light display.

The details of the game were displayed on a projector screen visible to all players. The projector was connected to a computer run by the other operator. The rules of the game were explained at the beginning of each session (the written instructions are available on request from the authors). No information was given about the total number of interactions that would be played.

The Rules of the Games

In the first session, the participants played the PD [10, 22] with the choices to cooperate (green) or defect (red). The computer randomly chose a pair of players under the constraint that each participant played two games with a different partner each time. The others observed and waited to be chosen later. To make the game simultaneous, the lights were covered during this session so that only the operators could see them, with choices being displayed on the projector screen after both players had decided whether to cooperate or defect. Only the last pair of choices, i.e., the last interaction, was displayed. The number of interactions per game was a random number drawn between 2 and 15 (inclusive). The payoff matrix was the following: if both players cooperated, they both got four points; if both defected, they both got two points; if one player cooperated and the other player defected, the cooperator got one point, and the defector got six points. After each game, the mean payoff of both players was displayed. This practice session got the students used to the PD [10, 22] and to our experimental setup. Thereafter, each player received a new ID.

The second session was an indirect reciprocity game as described in [18], with some modifications: each player received a starting account of £3.00. A pair of players was chosen: one in the donor role, the other in the receiver role. Players were told that the same pair would never play in the reversed role, so no direct reciprocity was possible. The donor was asked to decide whether he/she would give something to the receiver (green light) or not (red light). After this single interaction, a new pair of players was chosen. The cost of giving was £0.50, the benefit of receiving was £1.00. We donated the difference.

We used an image score as suggested in [14] to implement reputation: giving something increased the image score of the donor by one point, not giving decreased it by one point. This was graphically displayed with an arrow that wandered from an initial image score of 0 to a minimum of −6 or a maximum of 6 (these limits were arbitrarily chosen; they correspond to [14] and improved the vividness of the graphical display during the experiment). Both player's histories of giving or not were displayed with these arrows before each interaction. We played 24 rounds per group. Each player played once per round as donor and twice per two rounds as receiver. To examine the potential impact of information about the players' current accounts [29, 30], the accounts were displayed through the game for seven groups but were only displayed at the end of the session for the remaining five groups.

In the third session, each player played six PD games like those in the first session. During these games, both the players' final image scores from the second session were displayed on the screen. The players with the five highest mean payoffs per group received £5.00 each, in addition to their earnings from the second session.

Acknowledgments

We thank the students for participating; B. Chan, T. Little, L. Mitchell, and G. Steedman for help; and N. Colegrave, S. Gandon, S. Mitchell, S. Nee, A. Read, A. Rivero, M. Walker, and the referees for comments and/or discussion. The John D. and Catherine T. MacArthur Foundation (Research Network on Norms and Preferences) provided financial support. C.W. is supported by the Swiss National Science Foundation, V.A.B. by the Royal Society and the Biotechnology and Biological Science Research Council.

Received: March 14, 2002
Revised: April 22, 2002
Accepted: April 22, 2002
Published: June 25, 2002

References

1. Hamilton, W.D. (1963). The evolution of altruistic behaviour. Am. Nat. *97*, 354–356.
2. Wilson, D.S. (1980). The Natural Selection of Populations and Communities (Menlo Park, CA: Benjamin-Cummings Press).
3. Simon, H.A. (1990). A mechanism for social selection and successful altruism. Science *250*, 1665–1668.
4. Wilson, D.S., and Dugatkin, L.A. (1997). Group selection and assortative interactions. Am. Nat. *149*, 336–351.
5. Riolo, R.L., Cohen, M.D., and Axelrod, R. (2001). Evolution of cooperation without reciprocity. Nature *414*, 441–443.
6. Sigmund, K., Hauert, C., and Nowak, M.A. (2001). Reward and punishment. Proc. Natl. Acad. Sci. USA *98*, 10757–10762.
7. Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. Nature *415*, 137–140.
8. Trivers, R. (1971). The evolution of reciprocal altruism. Q. Rev. Biol. *46*, 35–57.
9. Axelrod, R., and Hamilton, W.D. (1981). The evolution of cooperation. Science *211*, 1390–1396.
10. Axelrod, R. (1984). The Evolution of Cooperation (New York: Basic Books).
11. Nowak, M.A., May, R.M., and Sigmund, K. (1995). The arithmetics of mutual help. Sci. Am. *272*, 76–81.
12. Alexander, R.D. (1987). The Biology of Moral Systems (New York: Aldine de Gruyter).
13. Zahavi, A. (1995). Altruism as a handicap—the limitations of kin selection and reciprocity. J. Avian Biol. *26*, 1–3.
14. Nowak, M.A., and Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. Nature *393*, 573–577.
15. Nowak, M.A., and Sigmund, K. (1998). The dynamics of indirect reciprocity. J. Theor. Biol. *194*, 561–574.
16. Lotem, A., Fishman, M.A., and Stone, L. (1999). Evolution of cooperation between individuals. Nature *400*, 226–227.
17. Leimar, O., and Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. Proc. R. Soc. Lond. B Biol. Sci. *268*, 745–753.
18. Wedekind, C., and Milinski, M. (2000). Cooperation through image scoring in humans. Science *288*, 850–852.
19. Milinski, M., Semmann, D., Bakker, T.C.M., and Krambeck, H.-J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy. Proc. R. Soc. Lond. B Biol. Sci. *268*, 2495–2501.
20. Seinen, I., and Schramm, A. (2001). Social Status and Group Norms: Indirect Reciprocity in a Helping Experiment (Amsterdam: Tinbergen Institute).
21. Milinski, M., Semmann, D., and Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. Nature *415*, 424–426.
22. Wedekind, C., and Milinski, M. (1996). Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat. Proc. Natl. Acad. Sci. USA *93*, 2686–2689.
23. Sugden, R. (1986). The Economics of Rights, Co-Operation and Welfare (Oxford, UK: Basil Blackwell).
24. Boyd, R., and Richerson, P.J. (1989). The evolution of indirect reciprocity. Soc. Networks *11*, 213–236.
25. Roberts, G., and Sherratt, T.N. (1998). Development of cooperative relationships through increasing investment. Nature *394*, 175–179.
26. Trivers, R. (1985). Social Evolution (Menlo Park, CA: Benjamin Cummings).
27. Dawkins, R. (1976). The Selfish Gene (Oxford: Oxford University Press).
28. Sigmund, K., and Hauert, C. (2002). Altruism. Curr. Biol. *12*, R270–R272.
29. Kazantzis, N., and Sutton, R. (2000). Examining the motivations for generosity. Science *290*, 454–455.
30. Wedekind, C., and Milinski, M. (2000). Examining the motivations for generosity—Reply. Science *290*, 455–455.