THE ROYAL SOCIETY

Received 18 February 2003
Accepted 16 June 2003
Published online 10 September 2003

The development of cooperative relationships: an experiment

Gilbert Roberts* **and James S. Renwick**

School of Biology, Henry Wellcome Building, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK

Pairs of individuals frequently face situations in which they could do well if they cooperated, but each risks being exploited. The Prisoner's Dilemma is widely used for investigating such scenarios, but it is framed in terms of cooperating and defecting, whereas in reality cooperation is rarely 'all or nothing'. Recent models allowing for variable investment in cooperation indicated the success of a strategy of 'raising-the-stakes' (RTS), which invests minimally at first and then increases its investment if its partner matches it. We tested whether this strategy was adopted by subjects participating in an experiment in which they could choose how much money to give to a partner, reciprocity being encouraged by doubling donations. Subjects did increase their donations over successive rounds, both when playing against a stooge who reciprocated with the same investment, and when playing with a partner who was free to choose their investment. Subjects showed a strong tendency to match variations in their partner's investments. Cooperation was therefore achieved through a combination of initial escalation (RTS strategy) and quantitative responsiveness ('give-as-good-as-you-get' strategy). Although initial offers were higher than predicted, our results were broadly consistent with theoretical expectations.

Keywords: cooperation; Prisoner's Dilemma; reciprocal altruism; relationships

1. INTRODUCTION

Mutual aid and social exchange are fundamental to human societies (Axelrod 1984; Alexander 1987; Cosmides & Tooby 1992; Ridley 1996). However, such cooperation presents a paradox because, although reciprocal altruism can profit both partners (Trivers 1971), an altruist risks being exploited by an individual that fails to reciprocate. For this reason, a great deal of theoretical attention has been devoted to understanding the conditions under which reciprocity will be stable. Central to this work is the Prisoner's Dilemma, a game-theoretical model that encapsulates the mixed motives involved in reciprocal altruism. If the game is played only once, the rational decision is to defect, but when it is played repeatedly, cooperation becomes possible through responsive strategies such as tit-for-tat (TFT; Axelrod & Hamilton 1981).

While a great deal of theoretical interest has focused on finding variant strategies that will out-perform TFT (e.g. Nowak & Sigmund 1992), the usefulness of this approach has been questioned owing to the mixed success of attempts to apply the models to behaviours such as allogrooming in impala (*Aepyceros melampus*; Hart & Hart 1992), food sharing in female vampire bats (*Desmodus rotundus*; Wilkinson 1984), predator inspection in sticklebacks (*Gasterosteus aculeatus*; Milinski 1987), pride defence in lionesses (*Panthera leo*; Heinsohn & Packer 1995) and pecking keys for food in blue jays (*Cyanocitta cristata*; Clements & Stephens 1995). Nevertheless, direct tests of whether individuals follow specific TFT-like strategies have had more success in humans (e.g. Wedekind & Milinski 1996).

One way in which the apparent divide between theoretical and empirical work might be bridged is by making models more realistic, and one way to do this is to allow for varying investment in cooperation. The classical Prisoner's Dilemma gives individuals the choice of either cooperating or defecting. However, cooperation is rarely an 'all-or-nothing' affair (Frean 1996). In many kinds of interactions, investment in cooperation can vary. For example, lending someone £10 is clearly more generous than lending them £1. Similarly, baboons (*Papio anubis*) grooming each other might do so for several seconds or several minutes, while vampire bats might regurgitate a small or a large volume of blood. Classing all such behaviours as cooperative loses important information about how cooperative individuals are. However, if we allow for variation in cooperative investment, a new kind of cheating becomes possible because individuals may invest a little less than their partners. The concern is that this 'subtle cheating' (Trivers 1971) or 'short changing' (Roberts & Sherratt 1998) will gradually erode cooperation.

To determine whether cooperation can be established despite the risks posed by both cheats and subtle cheats, Roberts & Sherratt (1998) investigated a model that allowed for varying investment in cooperation. The model effectively involved decomposing the Prisoner's Dilemma matrix into a reciprocal altruism game in which each individual has a decision to make about how much to invest in altruism. Each act results in a fitness cost to the altruist of u, with $u = 0$ corresponding to defection in the discrete model. The benefit to the recipient is given by ku, with $k > 1$ where both obtain a net benefit. Note that, over a symmetrical course of interactions, payoffs sum to fit the defining inequalities of the Prisoner's Dilemma, so, although investment can vary, the essential conflict between cooperation and exploitation remains. The variable-investment model requires a new strategy set that allows for quantitative responsiveness. Because the potential range of strategies is so large, Roberts & Sherratt

*Author for correspondence (gilbert.roberts@ncl.ac.uk).

(1998) confined consideration to a small subset representing some broad classes of behaviour, such as never investing ('non-altruism', NA), matching what the partner last gave ('give-as-good-as-you-get', GGG), or giving a little less than the partner ('short-changer', SC). It transpired that a strategy of 'raise the stakes' (RTS), which offers a small amount on first meeting and then, if matched, raises its investment, could readily invade a population of non-altruists and could not be invaded. Cooperative relationships could therefore develop through 'testing the water' followed by incrementally increasing investment.

The findings of Roberts & Sherratt (1998) were supported by genetic-algorithm models in which the strategies were allowed to evolve rather than being pre-determined (Sherratt & Roberts 1999*a*). However, they have been disputed by Killingback & Doebeli (1999), who argued that cooperation would tend to decline. Despite extensive efforts using a whole range of analytical and simulation methods, attempts to replicate these findings have failed (Sherratt & Roberts 1999*b*). Instead, Sherratt & Roberts (2002) have confirmed and extended their original findings.

As yet, there has been no direct experimental test of whether individuals actually use an RTS-like strategy in building up relationships. However, there has been a study on chacma baboons (*Papio cyanocephalus ursinus*) that investigated quantitative responsiveness in grooming interactions (Barrett *et al.* 2000). This study concluded that, while there was no evidence for the escalating investment predicted by RTS, there was quantitative matching as predicted by the GGG strategy. However, it was questionable whether the Roberts & Sherratt (1998) model could be applied, as the observations were necessarily snapshots of long-term relationships rather than showing how the subjects behaved on first meeting. Bshary (2002) has also applied the Roberts & Sherratt (1998) model to field experiments on the cleaner fish (*Labroides dimidiatus*) and its client reef fishes. He found that cleaners and resident clients did build up relationships, but with heavy initial investment, while there was no evidence that cleaners built up relationships with clients that had access to several cleaners. Again, as the author notes, the model may be difficult to apply to this scenario, in this case because of the asymmetries between partners with respect to their payoffs and strategic options.

In humans, there is much anecdotal evidence that people are more cooperative with others that they know well, and this is supported by recent experiments in which pairs of friends cooperated more than pairs of strangers (G. Roberts, unpublished data). Yet the strategies people use to develop from being uncooperative strangers to being cooperative friends are unknown. The aim of this study was to test whether people use an RTS-like strategy when developing cooperative relationships. We proposed that relationships should develop through increasing investment, and we tested this using a simple experimental game.

2. MATERIAL AND METHODS

We recruited 32 university students aged between 18 and 21 years who were naive with respect to the theory of reciprocal altruism. Subjects participated in a variant of the iterated Prisoner's Dilemma game that allowed for variable investment in cooperation. In each round of the game, subjects were allocated £10, which they could decide to donate to the other player in units of £1. Each £1 they donated was doubled by the experimenter. Therefore, both parties could make more money from donating more, but each risked being exploited if the other player did not reciprocate.

Players received written instructions explaining the choice they had of how much to donate and the possible outcomes. Although subjects could not keep the money awarded during the experiment, prize money gave the experiment greater ecological validity and acted as an incentive for subjects to focus on the task. Cooperating to achieve high scores was encouraged by giving the five highest-scoring subjects an equal chance of winning the £20 prize. Subjects interacted anonymously, their decisions on each round being communicated through the experimenter, who moved between the subjects' private booths. The experimenter's movements were such that each subject believed that they were playing with another subject, although in some games they were actually playing with the experimenter.

All subjects played the game twice, under different conditions. In the 'matched response' condition, each move by the subject was matched by an experimental stooge, whereas in the 'free response' condition, two subjects responded to each other's moves. To control for order effects, half of the subjects played under the 'matched response' condition first, the other half played under the 'free response' condition first. The 'matched response' condition began with the experimenter telling the subject that they had been randomly selected to lead. After making their move, they were told that their co-player would see the amount they gave and would then respond. In fact, each amount given was matched by an equivalent response from the experimenter playing a strategy of GGG (Roberts & Sherratt 1998). The GGG strategy was chosen for this condition, as it is the most neutral response. In the 'free response' condition, two subjects played against one another. One was arbitrarily chosen to lead and each subject's moves were communicated to the other by the experimenter. Subjects were unaware, in advance, of how many iterations of the game would be played. We therefore avoided the problem that the rational decision is to defect on the last move, and therefore on preceding moves. We analysed only the first nine rounds; all games continued for longer than this. After subjects had completed both conditions, they were interviewed and debriefed.

The data analysed here are in the form of the amount given by each subject in each of nine rounds. Amounts given were divided by 10 to give proportions, then, because of their truncated distribution, arcsine transformed before all statistical analyses.

3. RESULTS

There appeared to be a general trend for cooperative investments to increase with round number (figure 1), but there also appeared to be effects of the different experimental conditions and of the order in which they were experienced. Taking the 'matched response' condition first, although order did not have an effect on the mean investment across the nine rounds ($F_{1,30} = 0.817$, $p = 0.373$), there was a significant effect on the first move ($F_{1,30} = 0.858$, $p = 0.006$). Similarly, for the 'free response' condition, there was no effect of order on mean

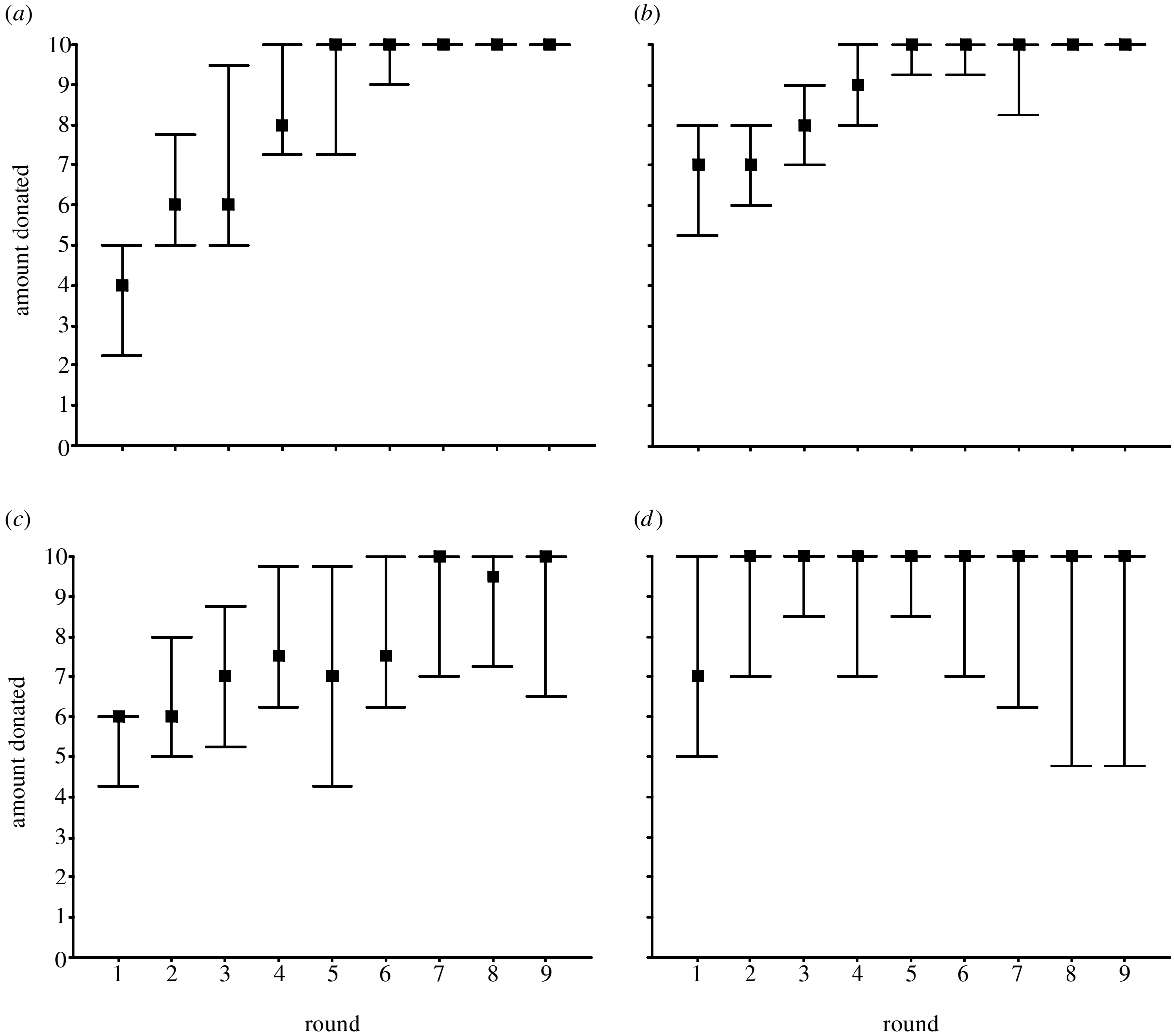

Figure 1. The amounts given in each round in the 'matched response' condition for those taking this condition (*a*) first and (*b*) second; and in the 'free response' condition, for those taking this condition (*c*) first and (*d*) second. Data are plotted as medians across 32 subjects with 25% and 75% quartiles.

investment ($F_{1,30} = 1.982$, $p = 0.169$), but there was a significant effect on the first move ($F_{1,30} = 9.689$, $p = 0.004$). Data are therefore analysed separately for the 'naive' (those playing their first condition) and 'experienced' (those playing their second condition) players.

Considering the 'matched response' condition first, we used repeated-measures analyses of variance (ANOVA) to analyse the donations of each of the subjects in each of the rounds. For naive players, there was a significant effect of round number on amount donated ($F_{8,120} = 34.891$, $p < 0.0005$). Examining the within-subjects contrasts revealed significant linear ($F_{1,15} = 112.45$, $p < 0.0005$) and quadratic ($F_{1,15} = 18.759$, $p < 0.001$) effects of round, confirming that there was a directional trend in the mean values across rounds. Similarly, experienced players showed a significant effect of round number on amount donated ($F_{8,120} = 19.708$, $p < 0.0005$), with both linear ($F_{1,15} = 75.469$, $p < 0.0005$) and quadratic ($F_{1,15} = 14.202$, $p = 0.002$) trends in the mean amount donated across rounds.

In the case of the 'free response' trials, subjects were tested in pairs whose investments cannot be considered as independent, so analyses were based on mean values for each of the 16 pairs. Again considering naive subjects first, repeated-measures ANOVA showed a significant effect of round ($F_{2.564,\ 17.948} = 3.943$, $p = 0.03$; note that, because Mauchly's test of sphericity gave a significant result, this uses the more conservative Greenhouse–Geisser method). Within-subject contrasts revealed a significant linear trend in the mean values given across rounds ($F_{1,7} = 7.53$, $p = 0.029$). However, there was no effect of round in the case of experienced players ($F_{1.482,\ 10.374} = 0.650$, $p = 0.498$, using the Greenhouse–Geisser method).

In the 'free response' condition, we can ask how the amount given depends both on the round number and on the amount given by the partner. Taking each pair separately, we used a multiple regression with the amounts given by the follower as the dependent variable and the round number and amounts given by the leader as independent variables. Seven of the pairs had to be excluded because all, or all but one, of the amounts given were 10. Out of the remaining pairs, there were two pairs with a significant effect of round (*Beta* = 0.618, $p = 0.039$ and *Beta* = 0.858, $p = 0.010$), two with a significant effect of what the leader gave (*Beta* = 0.813, $p = 0.005$ and *Beta* = 0.938, $p < 0.0005$) and one pair with a significant effect of both (*Beta* = 0.321, $p = 0.018$ and *Beta* = 0.697, $p < 0.0005$, respectively). All but one of these significant

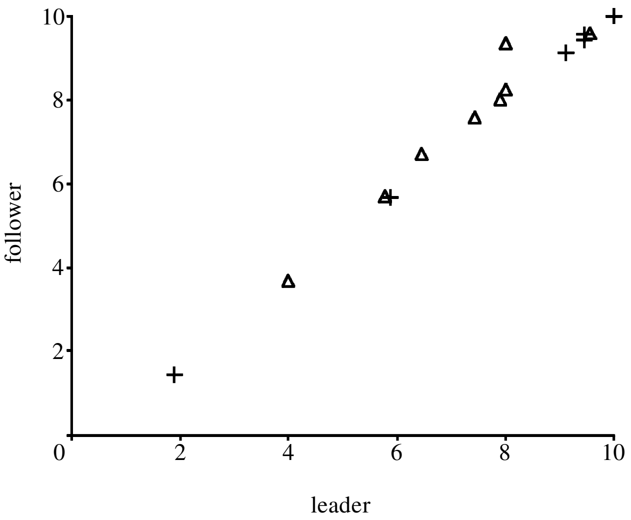

Figure 2. Donations by the leader (i.e. the individual making the first move), plotted against donations by the follower, for the 16 pairs in the 'free response' condition. Values plotted are means across all moves for that individual in rounds 1–9, plotted as triangles for those taking this condition first and crosses for those taking it second.

effects were found in the naive as opposed to the experienced pairs.

This interdependence of the amount donated and the amount received is clear when plotting the mean amount donated by the leader across the nine rounds against the mean amount donated by the follower (figure 2). Regressing the follower's donations on the leader's gives an adjusted r^2 of 0.975 ($p < 0.0005$). Furthermore, the follower's donations were in direct proportion to those of the leader: the constant was non-significant and the regression coefficient was consistent with unity ($b = 1.032 \pm \text{s.e.} = 0.043$).

4. DISCUSSION

Our experiment has provided evidence to support the prediction of Roberts & Sherratt (1998) that investment in cooperation will increase over the course of an interaction. This prediction arose from modelling work suggesting that such a strategy allows individuals to take advantage of cooperative opportunities while minimizing the risk of being exploited. Were individuals to perform a highly altruistic act without prior experience of their partner, they would be making a 'leap of faith' and laying themselves open to exploitation. Following the RTS strategy allows individuals to 'test the water', investing a little on first meeting and then incrementally increasing investment only if it is reciprocated. Such an increase was demonstrated most clearly when subjects played against a stooge who matched their donations, a procedure that allowed us to record a subject's behaviour with minimal interference. As their investments were returned, subjects appeared to gain in confidence and invest more heavily. The fact that this effect was also found in subjects that had already had experience of playing the other condition suggests that it is not the result of novelty or a lack of understanding of the game. In the 'free response' condition, a tendency to increase donations over time was also found, but only in those subjects playing for the first time. It may be that those with experience of playing the 'matched response' condition may have learned to invest more to get more in return.

Subjects in the 'free response' condition demonstrated quantitative responsiveness both in terms of dynamically matching what their partner had invested on the previous move and in terms of the overall correlation between amounts given and received. This behaviour is closer to the strategy of GGG (Roberts & Sherratt 1998) than it is to RTS, which would predict investing more than one's partner. Interestingly, this result fits well with the findings of Barrett *et al.* (2000) on grooming in chacma baboons. It is also in line with the predictions of Sherratt & Roberts (2001) who suggested that the RTS strategy is likely to evolve into a simpler variable-investment form of GGG that matched its partner's investment. Clearly, the RTS strategy is simplistic in assuming a linear escalation throughout the course of an interaction, and it may be that a strategy combining properties of RTS and GGG and allowing for nonlinear escalation would outperform RTS.

It is notable that initial investment levels were above those predicted by the original RTS strategy. Although subjects donated less than the maximum available on their first move, this was considerably greater than the minimal amount required to 'test the water'. This may be because subjects perceived a low risk of being exploited. RTS can be expected to predominate over strategies that do not show an initial caution only when there are cheats, subtle cheats or indeed any individuals that are at the time unwilling or unable to reciprocate adequately (Roberts & Sherratt 1998; Sherratt & Roberts 2001). As Sherratt & Roberts (2002) found, it may be that investing a minimal amount initially is favoured only in a non-cooperative environment because, when funds for investment are replenished on every move, overcaution means losing out on potential gains. The two conditions in our experiment provided environments of differing cooperativeness: whereas in the 'matched response' condition, subjects would necessarily do better by investing maximally in every round, in the more natural 'free response' condition, those that invested maximally from the start did not always find their generosity reciprocated. This was reflected in the fact that the five individuals that led in the 'free response' condition by offering £9 or £10 did not profit any more than the 11 individuals that led by offering between £2 and £6 ($t_{14} = 1.68$, $p = 0.114$). However, a full analysis of the extent to which different individuals followed different strategies, what those strategies were and whether strategies varied in their profitability would require a larger sample size.

The high levels of cooperation exhibited by our subjects are consistent with the results of a number of other studies that have found people to be more cooperative than one would predict (Buss 1999). There are a number of reasons why this might be the case. First, all subjects volunteered to take part in the study, an act that in itself suggests a relatively cooperative nature. Furthermore, although the subjects were naive with regard to the hypotheses under test, they were aware that they were taking part in an experiment on cooperation, and this may have encouraged cooperative behaviour. There are also a number of reasons for believing that subjects' incentives to cooperate might lie outside the context of the experiment (Colman 1995). One important factor is that subjects may perceive that being cooperative would enhance their reputation (Roberts 1998). Even though the subjects interacted anonymously, they might have wished to create a positive image with the experimenter, perhaps gaining social acceptance by complying with socially desirable norms and acting more cooperatively than they would have done had the experimental design been truly anonymous (Turner 1972; Hogg & Abrams 1999). There is also increasing evidence from studies in anthropology and evolutionary psychology that individuals may be cooperative to enhance their social standing (Smith & Bliege Bird 2000; Sosis 2000).

We have seen how the RTS strategy can emerge when individuals are given a choice of how much to invest in cooperation. Our results should therefore help us to understand the development of cooperative relationships. They may also be compared with analogous results from studies of contributions to public goods, where it has been suggested that incremental strategies of commitment may increase the production of public goods in later rounds (Kurzban *et al.* 2001). The prediction that reciprocal relationships will develop from small beginnings is also consistent with the observation of Axelrod (1984) that in the 'live-and-let-live' system of trench warfare 'restraint undertaken in certain hours could be extended to longer hours'. More generally, the escalation of altruism shown by RTS-like strategies appears to mirror the tendency of people to form friendships and to act preferentially towards friends.

The authors thank John Lazarus and Tom Sherratt for helpful comments.

REFERENCES

Alexander, R. D. 1987 *The biology of moral systems*. New York: Aldine 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.
Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E. & Hill, R. A. 2000 Female baboons do not raise the stakes but they give as good as they get. *Anim. Behav.* **59**, 763–770.
Bshary, R. 2002 Building up relationships in asymmetric co-operation games between the cleaner wrasse *Labroides dimidiatus* and client reef fish. *Behav. Ecol. Sociobiol.* **52**, 365–371.
Buss, D. M. 1999 *Evolutionary psychology—the new science of the mind*. Boston, MA: Allyn & Bacon.
Clements, R. & Stephens, D. C. 1995 Testing models of non-kin cooperation—mutualism and the prisoner's dilemma. *Anim. Behav.* **50**, 527–535.
Colman, A. 1995 *Game theory and its applications in the social and biological sciences*. Boston, MA: Butterworth-Heinemann.
Cosmides, L. & Tooby, J. 1992 Cognitive adaptations for social exchange. In *The adapted mind* (ed. J. Barkow, L. Cosmides & J. Tooby), pp. 163–228. New York: Oxford University Press.
Frean, M. 1996 The evolution of degrees of cooperation. *J. Theor. Biol.* **182**, 549–559.
Hart, B. L. & Hart, L. A. 1992 Reciprocal allogrooming in impala, *Aepyceros melampus*. *Anim. Behav.* **44**, 1073–1083.
Heinsohn, R. & Packer, C. 1995 Complex cooperative strategies in group-territorial African lions. *Science* **269**, 1260–1262.
Hogg, M. A. & Abrams, D. 1999 *Social identifications: a social psychology of intergroup relations and group processes*. London: Routledge.
Killingback, T. & Doebeli, M. 1999 'Raise the stakes' evolves into a defector. *Nature* **400**, 518.
Kurzban, R., McCabe, K., Smith, V. L. & Wilson, B. J. 2001 Incremental commitment and reciprocity in a real-time public goods game. *Personality Social Psychol. Bull.* **27**, 1662–1673.
Milinski, M. 1987 Tit for tat and the evolution of cooperation in sticklebacks. *Nature* **325**, 433–443.
Nowak, M. & Sigmund, K. 1992 Tit for tat in heterogeneous populations. *Nature* **355**, 250–253.
Ridley, M. 1996 *The origins of virtue*. London: Viking.
Roberts, G. 1998 Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond.* B **265**, 427–431. (DOI 10.1098/rspb.1998.0312.)
Roberts, G. & Sherratt, T. N. 1998 Development of cooperative relationships through increasing investment. *Nature* **394**, 175–179.
Sherratt, T. N. & Roberts, G. 1999*a* The evolution of quantitatively responsive cooperative trade. *J. Theor. Biol.* **200**, 419–426.
Sherratt, T. N. & Roberts, G. 1999*b* 'Raise the stakes' evolves into a defector—reply. *Nature* **400**, 518.
Sherratt, T. N. & Roberts, G. 2001 The role of phenotypic defectors in stabilizing reciprocal altruism. *Behav. Ecol.* **12**, 313–317.
Sherratt, T. N. & Roberts, G. 2002 The stability of cooperation involving variable investment. *J. Theor. Biol.* **215**, 47–56.
Smith, E. A. & Bliege Bird, R. L. 2000 Turtle hunting and tombstone opening: public generosity as costly signalling. *Evol. Hum. Behav.* **21**, 245–261.
Sosis, R. 2000 Costly signalling and torch fishing on Ifaluk atoll. *Evol. Hum. Behav.* **21**, 223–244.
Trivers, R. L. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.
Turner, J. C. 1972 *Social influence*. Milton Keynes: Open University Press.
Wedekind, C. & 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**, 2682–2689.
Wilkinson, G. S. 1984 Reciprocal food sharing in vampire bats. *Nature* **308**, 181–184.