

# Competitive altruism: from reciprocity to the handicap principle

#### Gilbert Roberts

Evolution and Behaviour Research Group, Department of Psychology, University of Newcastle upon Tyne, Newcastle upon Tyne NE17RU, UK (gilbert.roberts@ncl.ac.uk)

Current work on cooperation is focused on the theory of reciprocal altruism. However, reciprocity is just one way of getting a return on an investment in altruism and is difficult to apply to many examples. Reciprocity theory addresses how animals respond dynamically to others so as to cooperate without being exploited. I discuss how introducing differences in individual generosity together with partner choice into models of reciprocity can lead to an escalation in altruistic behaviour. Individuals may compete for the most altruistic partners and non-altruists may become ostracized. I refer to this phenomenon as competitive altruism and propose that it can represent a move away from the dynamic responsiveness of reciprocity. Altruism may be rewarded in kind, but rewards may be indirectly accrued or may not involve the return of altruism at all, for example if altruists tend to be chosen as mates. This variety makes the idea of competitive altruism relevant to behaviours which cannot be explained by reciprocity. I consider whether altruism might act as a signal of quality, as proposed by the handicap principle. I suggest that altruistic acts could make particularly effective signals because of the inherent benefits to receivers. I consider how reciprocity and competitive altruism are related and how they may be distinguished.

Keywords: cooperation; altruism; reciprocity; handicap principle; Prisoner's Dilemma

#### 1. INTRODUCTION

Altruistic behaviour fascinates evolutionary biologists because altruism, by definition, incurs a fitness cost. Why should an animal perform a costly act? For unrelated individuals, answers have focused on Trivers' (1971) theory of reciprocal altruism. However, the stability of reciprocity is problematic because altruists may be exploited by individuals which fail to reciprocate. A solution to this problem has been developed through computer simulations pioneered by Axelrod & Hamilton (1981). If animals use responsive strategies such as Tit-for-Tat, then reciprocity can become established. Yet the problem of exploitation will be so pervasive that reciprocity is only likely to become established under restricted conditions (Axelrod & Hamilton 1981). In this paper, I suggest a shift in focus from how individuals can avoid exploitation by cheats, to how individuals may compete to develop reputations as the most altruistic. This new perspective arises from incorporating individual differences together with partner choice. I go on to argue that altruism may be interpreted as a signal. I draw a parallel between the difficulty in explaining altruistic behaviour and the difficulty in explaining costly displays used in mate choice. In fact, Zahavi (see, for example, Zahavi 1977) has drawn attention to this parallel, but the handicap principle has not been widely developed as an explanation of altruism, despite its recognition as a widely applicable theory (Grafen 1990). Here, I highlight the possible application of the handicap principle to altruism, explore its relation to reciprocity, and examine how we may test the theories.

### 2. THE PROBLEM OF ALTRUISM AND THE SOLUTION OF RECIPROCITY

A fundamental problem besetting work on altruism is the difficulty in demonstrating the costs and benefits involved. Consequently, it is difficult to draw up a list of altruistic behaviours requiring explanation, let alone to actually explain them. Behaviours previously labelled altruistic (such as predator inspection visits by fish, Milinski 1987) have been re-interpreted as involving direct self-interest (Connor 1996). Nevertheless, it is reasonable to develop a theory of when we may expect to find altruism, even if the selective benefits of any particular behaviour are unclear. Arguably, examples of altruism include alarm calling, sentinel behaviour, territorial defence, allogrooming, allofeeding, cooperative foraging and restraint in contests (Dugatkin 1997).

The theory of reciprocal altruism currently dominates discussion of non-kin altruism (see, for example, Brembs 1996). It has been the subject of scores of theoretical papers based on the iterated Prisoner's Dilemma game. In this scenario, mutual cooperation pays well, but exploiting a cooperator pays best; mutual refusal to cooperate pays poorly, but cooperating and being exploited pays worst of all. Cooperation appears doomed to failure. However, if players use a Tit-for-Tat strategy (cooperating on the first move and then playing as the other player did on its last move) over repeated interactions then reciprocity can become established (Axelrod & Hamilton 1981). The key problem is in relating the elegant theory to the evidence. As yet, there is little evidence for the theory (see, for example, Wilkinson 1984), though this lack of

It is not the object of this paper to provide a detailed critique of work on reciprocity. Rather, it is my aim to ask a more fundamental question. Namely, even if the theory proved essentially sound and even if there were a number of uncontroversial examples, could reciprocity provide a general explanation for altruistic behaviours among nonkin? Reciprocity is not a catch-all term for eventually benefiting from altruism: it specifically refers to interactions within dyads, whereby a short-term cost paid by one individual leads it to receive a greater benefit from another individual. Reciprocity can involve exchange of services which an animal cannot ever perform for itself but which it can perform for others (e.g. impala, Aepyceros melampus, are unable to groom their necks themselves; Hart & Hart 1992). Or it can involve exchange of different services at the same or different times (e.g. trading eggs for sperm in simultaneous hermaphrodites, Connor 1992). Or it can involve provision of a service at one time which is returned at a later date (e.g. food sharing among vampire bats, Desmodus rotundus; Wilkinson 1984). Even if reciprocity could offer a solution for such cases, many cases do not fit into these categories. It is not clear how reciprocity could explain altruism which benefits a group. Yet if, for example, territorial defence (e.g. among lionesses, Panthera leo; Heinsohn & Packer 1995) is altruistic, a whole group benefits rather than just one individual reciprocator. It is not clear how reciprocity could explain altruism which is consistently asymmetric or even entirely one-way. Yet allopreening, for example, is two-way in some species of birds but just one-way in others (Harrison 1965). Humans often seem to be more altruistic than would be predicted on economic grounds (see, for example, Frank 1988). Furthermore, they are often altruistic to non-reciprocators (e.g. by donating to charities). An important theme of this paper is that reciprocity is just one way of getting a return on an investment in altruism. Recent work has tended towards a classification of cooperative behaviours as either mutualistic or reciprocal (see, for example, Mesterton-Gibbons & Dugatkin 1997) as if these were the only possibilities. Other mechanisms deserve attention.

I start by re-examining some of the assumptions of the current theoretical framework and suggesting a modified approach. Most computer simulations, such as those of Axelrod & Hamilton (1981) have been based on individuals which interact repeatedly with a particular partner or with all others (a round robin tournament). In many real situations, individuals can choose between potential partners, and those partners may be unequal both in strategy and in the resources which they allocate to altruism. This is not to say that partner choice (see, for example, Bull & Rice 1991; Peck 1993) and varying investment in altruism (Frean 1993) have not been considered, nor to say that models which do not include more biologically realistic complications have nothing to offer. However, the focus on the instability of reciprocity and

the conclusion that it can only be stable within a narrow range of conditions are linked to the way in which models based on the iterated Prisoner's Dilemma have been implemented. In the following, I consider the consequences of changing some of the standard assumptions.

### 3. INDIVIDUAL DIFFERENCES AND PARTNER CHOICE CAN LEAD TO AN ESCALATION OF GENEROSITY

Imagine that individuals can make a strategic decision about how much of their available resources to invest in altruism. I refer to this as their 'generosity'. Imagine also that individuals can choose to interact with particular other individuals on the basis of their experience and that they prefer generous partners. There are a number of ways of implementing the themes of varying generosity and partner choice into simulation models. Roberts (1998) and Sherratt & Roberts (1998) offer two different ways of doing this, based on an explicit consideration of the costs and benefits involved in reciprocal altruism. We find that in some conditions, simple partner-choice rules can lead to the most generous individuals partnering other generous individuals and leaving non-altruists ostracized. This can mean that the most generous also receive the most, and if this feeds back into the number of offspring produced, generous strategies can spread. The key result is a very simple one: introducing individual differences in generosity combined with partner choice can lead to competition for partners and thereby to an escalation in generosity which is rewarded by greater benefits for those which choose to or can invest more.

## 4. ALTRUISM WITHOUT DIRECT RECIPROCATION CAN BRING BENEFITS THROUGH REPUTATION-BUILDING

A natural extension of reciprocity is to consider whether altruism can bring benefits even if it is not reciprocated directly. Can it ever pay an individual A to perform an altruistic act for individual B to gain from individual C? In other words, can non-reciprocal altruism play a role in developing partnerships for reciprocal altruism? Many interactions between organisms occur within partnerships, yet partners change. Hence, I suggest that one realistic way in which to model interactions may be to have a two-stage model incorporating both an 'assessment' stage and a 'partnered' stage. In the assessment stage, animals interact with the whole population, whereas in the subsequent paired stage they interact solely with their chosen partner.

Again, there is a range of ways in which such a scenario could be modelled, but some general features can be outlined. In the previous section, I considered that individuals might differ in their generosity. This had direct benefits through assortative partner choice. However, we can also imagine a scenario whereby high generosity does not have such direct benefits. It may be that an increase in generosity in the assessment stage is uneconomic in the short-term, but that the debt could be recouped in the long-term if the generous altruist was more likely to obtain a generous, profitable long-term partner.

An alternative scenario is to think of individuals as differing in 'forgiveness', where forgiveness represents the probability of behaving altruistically towards an individual which has not reciprocated in the past. Once again, if individuals use information about the altruistic behaviour of others in deciding with whom to form a long-term partnership, then it may be worth going into debt in the short term to secure a profitable, altruistic partner in the longer term.

#### 5. BEYOND RECIPROCITY: COMPETITIVE ALTRUISM

A key theme of the above scenarios is that, perhaps paradoxically, competition between unequal individuals may be an important driving force in cooperation. This is surprising because the Prisoner's Dilemma paradigm sees individuals as being better off in the short term by taking the benefits without paying the costs. For example, at any one time, a bird might do best by being allopreened without returning the service. However, I suggest that, in certain circumstances, individuals may compete to be and to be seen to be altruistic.

This shift in emphasis arises from considering the wider context in which interactions take place. Thus, allopreeners may not just be playing with each other, they may be competing against a population of potential partners. Those who allopreen can be thought of as a resource for which others compete. Individuals may still face the essential problem encapsulated in the Prisoner's Dilemma, namely that exploitation pays better than cooperation. However, where there is competition for partners, the most altruistic can preferentially interact with each other, leaving non-altruists ostracized and unable to exploit them. Thus, in effect, cheating is not a profitable option. The paradox is that to score points in the wider context it may be necessary to forego points in the narrow confines of a particular interaction defined by the Prisoner's Dilemma. If reputation is more important than shortterm gains, then altruism could persist in a Prisoner's Dilemma through competition: competition for the attentions of other altruists, competition for mates. If the competition is for higher stakes than is any particular altruistic interaction, the one who invests more in altruism may be at a competitive advantage.

This is important for two reasons. The first is that it predicts that the occurrence and nature of cooperative interactions will depend on the extent of individual differences and the potential for competition between partners. The second is that behaviour may not be as directly responsive as has previously been predicted. The link between performing an altruistic act and receiving a benefit is likely to become less direct if individuals both use information gained from watching others interacting and act so as to be observed. There would seem to be a benefit to individuals which could avoid interacting with those it has observed to be exploitative, and which could encourage cooperative individuals to preferentially interact with them. It is therefore surprising that little attention has been paid to such reputation-based behaviour, Pollock & Dugatkin (1992) being an exception. Axelrod (1984) briefly considers the effects of reputations and suggests that it is best to foster a reputation as a bully so that you will be at a lower risk of exploitation. I suggest instead that where there is competition for partners of different quality, a generous reputation could pay off in the long term. The use of reputations is particularly

plausible in small, tight social groups in which individuals will make use of information other than their own experiences. However, there are also problems in using information not based on personal experience because how individuals behave with others is not necessarily a good guide to how they will behave with you.

Altruists might recoup their costs in a number of ways. First, altruism may lead to direct reciprocation, as proposed by Trivers (1971). Second, altruism may not be reciprocated directly by the recipient, but may bring indirect economic benefits through more generous or more forgiving individuals being able to select more cooperative partners for longer-term reciprocal interactions. A potential problem with this scenario is that an individual might cheat by securing a reputation and then defecting. One answer is that this need not be a problem if the reputation is continually tested and individuals can switch partners. Another answer to this is that if altruistic behaviour serves as a signal of individual quality then reciprocation need not continue. Altruism might not be maintained by the return of altruism at all, and the altruist may benefit not through receiving altruistic acts either directly or indirectly, but in other ways, such as increasing its mating opportunities. This brings us to the handicap principle.

#### 6. CAN ALTRUISTIC BEHAVIOUR BE UNDERSTOOD AS A HANDICAP?

The peacock's tail has become a metaphor for the apparent over-indulgence of many animals in attributes used in mate choice. Zahavi has argued that the cost of such characters was an essential feature of an honest signalling system (Zahavi 1975). The idea was initially controversial but gained credibility with Grafen's (1990) models. Grafen (1990) showed through an evolutionarily stable strategy (ESS) model how costly, strategic signals could provide an honest indicator of quality. Zahavi (Zahavi 1977, 1995; Zahavi & Zahavi 1997) has argued that altruism can be understood as a handicap. However, while Grafen (1990) recognized the wide applicability of the handicap principle, the theory has not been widely developed in the field of altruistic behaviour and a recent review of cooperative behaviour makes no mention of it (Dugatkin 1997).

Can altruistic behaviour qualify as a handicap? To address this question I consider a behaviour which may qualify as both sexually selected and altruistic, namely allopreening by a male of a female. Advertising can be considered as the level of investment in allopreening. If the following hold, then the handicap principle can apply. Males of higher quality (where quality encompasses attributes such as time and energy available for activities such as allopreening which do not directly contribute to survival) have higher fitness. A male will be better off if a female assesses its quality as being high. Allopreening reduces fitness and is more expensive to those of lower quality. The gain in fitness to a partner from a better assessment is at least as great as for an individual with high reserves as for one with low. Given these assertions, it follows from Grafen (1990) that we can expect that at the ESS, higher quality males will allopreen more, and the costs incurred will be more than compensated for by

the effects on female choice. Thus, an individual might make a strategic investment in altruism which acts as a signal of its ability to be altruistic or of some other aspect of its quality, and which influences receiver behaviour in such a way that the fitness of the altruist is increased. The implication of this argument is that with a few reasonable assumptions we can apply the handicap principle to an altruistic act and we can use the theory developed for handicap signals in general. However, there is a particularly interesting feature of altruism as a signal which I will now address.

### 7. ALTRUISM AS A SIGNAL BENEFITING THE RECEIVER

Perhaps the most unusual feature of altruism as a signal is that the receiver actually benefits. This is in contrast to typical signaller-receiver interactions where the receiver expends energy, time or risk in assessing signals (Dawkins & Guilford 1991). These authors argued that receiver costs will impose a selection pressure away from honest signalling towards conventional signalling which is open to cheating. However, if a receiver actually benefits from an altruistic signal then this argument may not apply. Altruistic signals may be less likely to become unreliable through the effects of the receiver costs discussed by Dawkins & Guilford (1991). Indeed, there is the intriguing possibility that altruism could be used by signallers to promote attention to their signals. There is a parallel here with commercial advertising. Some advertisers promote products by offering 'free gifts'. This is their way of attracting the reluctant consumer's attention to their products. They offer an incentive to attend. If this 'free gift theory' is correct, mate-choice signals involving altruism should be particularly widespread. We can expect selection on receivers in favour of attending to altruistic signals. Thus, signals are unlikely to be arbitrary. For example, the most effective way of signalling foraging ability might be through courtship feeding, which could be seen not as an arbitrary means of strengthening pair bonds but as a signal with a direct benefit.

### 8. DISCUSSION: HOW MAY THE THEORIES BE DISTINGUISHED?

I have discussed a number of mechanisms which might explain the evolution of non-kin altruism, from direct reciprocity among individuals which differ in their generosity and which can choose partners, through reputation-based behaviour, to altruism as a signal of quality. There is a wide range of scenarios in nature and we may need a corresponding range of explanations. Seeing altruistic behaviour as a signal provides a possible link between these mechanisms, but how can we distinguish the theories?

If altruism is competitive then the conditions in which we should expect to find altruism and the nature of the behaviour we find may differ from the predictions of reciprocity. Reciprocity is about dynamically responsive behaviour, in which altruistic acts are rewarded and non-altruism punished. Competitive altruism simply predicts that on average a highly altruistic individual should receive more. Thus, both reciprocity and competitive

altruism predict a correlation across individuals between acts performed and acts received, but in reciprocity this derives from the prediction of a Tit-for-Tat, responsive strategy whereas in competitive altruism it derives from assortative partner choice. Even if behaviour is still responsive, it may be more flexible and less dependent on the probability of future interactions with a particular partner. Individuals may perform acts of altruism to enhance their reputations, so there may be a looser link between giving and receiving than in direct reciprocity. In impala (Hart & Hart 1992), grooming does seem to be dynamically reciprocal, with virtually every move being matched. In comparison, allopreening sequences of breeding pairs and neighbouring guillemots Uria aalge (G. Roberts, unpublished data) do not show strict Tit-for-Tat responsiveness. It therefore seems reasonable to accept reciprocity as a good explanation in impala, but the number of examples of such dynamic responsiveness is limited.

In allopreening, interactions are often two-way, with continual reversal of the roles of altruist and recipient, but in some species allopreening is just one-way (Harrison 1965). If we are to explain two-way allopreening as reciprocal altruism then, unless the reciprocation is in a different currency, we must look for a different class of explanation for one-way allopreening. An appealing feature of competitive altruism is that it can provide an explanation for altruism whether or not it is reciprocated. Competitive altruists are not necessarily afraid of defection because altruism may not be maintained by the return of altruism at all. How can altruism be favoured without reciprocation? If altruism is a signal, the benefits of the altruistic act to the recipient may be largely incidental. The benefits to the altruist may come not through reciprocity, but as a result of influencing receiver behaviour in some other way. In allogrooming impala and allofeeding vampire bats, the benefits appear to lie very much in receiving an altruistic act. In these cases, reciprocity provides a coherent explanation, but such cases seem to be rare. It is worth considering whether allogrooming among primates and allopreening among birds may be maintained by selective benefits other than reciprocation. In these cases, the altruist may benefit through forming alliances and partnerships.

Is this just a trade of altruism for mating opportunities? Consider a bird allopreening its mate or a male offering a nuptial gift to a female. There is the potential for a Prisoner's Dilemma here in that the recipient might defect by receiving the benefit and giving nothing in return, whether in the same currency or in terms of mating opportunities. However, if the gift is a signal of quality, then there may not be an incentive to defect by not mating. If one of the benefits of an altruistic reputation is improved mating success then, in practice, altruism may have as much to do with sexual selection as with the economics of reciprocity. The level of altruistic behaviour found may relate more to the potential for competition between prospective partners than to the simple economics of, for example, grooming. Furthermore, for reciprocal altruism to result in mutual gain, the benefits to the recipient must be greater than the costs to the altruist. However, if altruism is not maintained by reciprocity, this need not be so. It need only be the case that the

If a whole group gains from an altruistic act, who is to return the favour? Although there have been attempts to explain cooperation in groups (see, for example, Boyd & Richerson 1988), there is no reciprocity-based solution to the 'tragedy of the commons' and this remains a key problem for reciprocity as a general solution to the problem of altruism. Competitive altruism may provide an answer. There may be advantages in developing an altruistic reputation and for group members to respect a greater access to resources of altruists. An individual challenging the access to the resources of an altruist, perhaps by expelling that individual from the group, would pay the costs of losing that individual's contribution to, for example, group defence.

Some of the most striking examples of non-kin-based altruism occur in humans. Explanations have focused on reciprocity as a particularly powerful force in humans (see, for example, Ridley 1996). But is reciprocity an adequate explanation? Take that classic example of reciprocity among humans, the live-and-let-live system of trench warfare in World War I (Axelrod 1984). While reciprocity may provide a valid explanation, it is odd that the most cited example of altruism in war is of altruism towards the enemy. Surely there are also examples of altruism towards one's countrymen. Frank (1988) argues that humans often behave in ways which are difficult to explain in economic terms. Reciprocity cannot explain why individuals should put their own lives at risk or give to non-reciprocators, such as charities. Perhaps giving to charity enhances one's reputation, which in turn has indirect benefits. If altruism, even heroism, is a quality used in mate selection, then we may expect competition to be, and to be seen to be, altruistic. If only the highest quality individuals can afford, strategically, to undertake heroic acts, then perhaps life-threatening heroism could be the ultimate handicap.

Reciprocity is a very appealing theory, perhaps because it provides such a straightforward solution to the evolutionary biologist's question: If the animal is paying a cost, what is it getting back? The ultimate test must be whether a theory is successful in explaining animal behaviour, but while reciprocity has helped to understand some systems, it does not offer a satisfactory solution in many other cases. In many of these, direct self-interest may actually be the most parsimonious explanation. In this paper I have offered some ideas which might contribute to the development of alternative theories of how the short-term fitness costs of altruism might be recouped. It is the cost of altruism which presents the problem and which led us into the Prisoner's Dilemma. But there is a theory for which cost is not a problem. In the handicap principle, costs are inherent, and it therefore deserves consideration as a possible explanation of the apparent squandering of fitness involved in altruism.

I thank Tom Sherratt, Marion Petrie, John Lazarus, the Behaviour and Evolution Research Group and two referees for

helpful discussions and comments. I am supported by a Lord Adams Fellowship.

#### **REFERENCES**

- 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.
- Boyd, R, & Richerson, P. 1988 The evolution of reciprocity in sizeable groups. J. Theor. Biol. 132, 337-356.
- Brembs, B. 1996 Chaos, cheating and cooperation: potential solutions to the Prisoner's Dilemma. Oikos 76, 14-24.
- Bull, J. & Rice, W. 1991 Distinguishing mechanisms for the evolution of cooperation. J. Theor. Biol. 149, 63-74.
- Clements, R. & Stephens, D. C. 1995 Testing models of non-kin cooperation—mutualism and the prisoner's dilemma. Anim. Behav. 50, 527-535.
- Connor, R. C. 1992 Egg-trading in simultaneous hermaphrodites: an alternative to Tit for Tat. J. Evol. Biol. 5,
- Connor, R. C. 1996 Partner preferences in by-product mutualisms and the case of predator inspection in fish. Anim. Behav. **51**. 451-454.
- Dawkins, M. S. & Guilford, T. 1991 The corruption of honest signalling. Anim. Behav. 41, 865–874.
- Dugatkin, L. A. 1997 Cooperation among animals. New York: Oxford. Frank, R. H. 1988 Passions within reason. New York: Norton.
- Frean, M. R. 1994 The Prisoner's Dilemma without synchrony. Proc. R. Soc. Lond. B 257, 75-79.
- Grafen, A. J. 1990 Biological signals as handicaps. J. Theor. Biol. **144**, 517–546.
- Harrison, C. J. O. 1965 Allopreening as agonistic behaviour. Behaviour 24, 161-209.
- 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.
- Mesterton-Gibbons, M. & Dugatkin, L. A. 1997 Cooperation and the Prisoner's Dilemma: towards testable models of mutualism versus reciprocity. Anim. Behav. 54, 551-557.
- Milinski, M. 1987 Tit-for-Tat in sticklebacks and the evolution of cooperation. *Nature* **325**, 433–435.
- Peck, J. R. 1993 Friendship and the evolution of co-operation. J. Theor. Biol. 162, 195-228.
- Pollock, G. & Dugatkin, L. A. 1992 Reciprocity and the emergence of reputation. J. Theor. Biol. 159, 25-37.
- Ridley, M. 1996 The origins of virtue. London: Viking.
- Sherratt, T. N. & Roberts, G. 1998 The evolution of generosity and choosiness in co-operative exchanges. J. Theor. Biol. (In the press.)
- Trivers, R. L. 1971 The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35-57.
- Wilkinson, G. S. 1984 Reciprocal food sharing in the vampire bat. Nature 308, 181-184.
- Zahavi, A. 1975 Mate selection—a selection for a handicap. J. Theor. Biol. 53, 205-214.
- Zahavi, A. 1977 Reliability in communication systems and the evolution of altruism. In Evolutionary ecology (ed. B. Stonehouse & C. M. Perrins), pp. 253–259. London: Macmillan Press.
- Zahavi, A. 1995 Altruism as a handicap—the limitations of kin selection and reciprocity. J. Avian Biol. 26, 1-3.
- Zahavi, A. & Zahavi, A. 1997 The handicap principle. New York: Oxford University Press.