Evolving the ingredients for reciprocity and spite

Marc Hauser, Katherine McAuliffe and Peter R. Blake

References	**This article cites 73 articles, 21 of which can be accessed free** http://rstb.royalsocietypublishing.org/content/364/1533/3255.full.html#ref-list-1
Rapid response	Respond to this article http://rstb.royalsocietypublishing.org/letters/submit/royptb;364/1533/3255
Subject collections	Articles on similar topics can be found in the following collections behaviour (782 articles) cognition (193 articles) evolution (1078 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

Evolving the ingredients for reciprocity and spite

Marc Hauser[1,2,*]**, Katherine McAuliffe**[2] **and Peter R. Blake**[3]

[1]*Department of Psychology,* [2]*Department of Human Evolutionary Biology, and* [3]*Graduate School of Education, Harvard University, Cambridge, MA, USA*

Darwin never provided a satisfactory account of altruism, but posed the problem beautifully in light of the logic of natural selection. Hamilton and Williams delivered the necessary satisfaction by appealing to kinship, and Trivers showed that kinship was not necessary as long as the originally altruistic act was conditionally reciprocated. From the late 1970s to the present, the kinship theories in particular have been supported by considerable empirical data and elaborated to explore a number of other social interactions such as cooperation, selfishness and punishment, giving us what is now a rich description of the nature of social relationships among organisms. There are, however, two forms of theoretically possible social interactions—reciprocity and spite—that appear absent or nearly so in non-human vertebrates, despite considerable research efforts on a wide diversity of species. We suggest that the rather weak comparative evidence for these inter-actions is predicted once we consider the requisite socioecological pressures and psychological mechanisms. That is, a consideration of ultimate demands and proximate prerequisites leads to the prediction that reciprocity and spite should be rare in non-human animals, and common in humans. In particular, reciprocity and spite evolved in humans because of adaptive demands on cooperation among unrelated individuals living in large groups, and the *integrative* capacities of inequity detection, future-oriented decision-making and inhibitory control.

Keywords: reciprocal altruism; spite; ultimate pressures; proximal constraints

1. INTRODUCTION

In *The descent of man*, Darwin (1871) pondered the evolutionary origins of altruism and self-sacrifice among humans. The puzzle, as Darwin realized, was that such behaviours pose significant costs to the individual: 'he who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature' (p.163). To solve this problem, Darwin assumed that self-sacrifice might payoff in the currency of group benefits. He thus stated, if 'a tribe including many members who ... were always ready to give aid to each other and sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection' (p.166). In other words, the costs to the individual of self-sacrifice and other altruistic behaviour could evolve if the individual's group benefited relative to other groups lacking such behaviours.

As this history has been recounted many times, here we simply reiterate the key ideas and findings in telegraphic form so as to set up the essential problems discussed in this essay. In brief, sociobiologists raised what Dawkins (1976) famously described as the problem of *subversion from within*, that is, in a group of self-sacrificial altruists, defectors immediately win as they reap the benefits without paying the costs.

Thus, group selection was attacked as, at best, a weak account of the evolution of altruistic behaviour. As an alternative, Hamilton (1964) and Williams (1966) proposed and developed a gene's eye view of altruism, arguing that self-sacrifice would evolve if the costs to the individual were offset by benefits to the individual's close kin. What drives the evolution of altruism, therefore, is a consideration of the distribution of winning genes, rather than winning individuals or groups. But what about altruistic behaviour among genetically unrelated individuals? The solution, provided by Trivers (1971), was reciprocal altruism: self-sacrifice is offset because the initial act of altruism is conditioned upon a reciprocated act of altruism in the future.

These brilliant ideas can be placed in the context of a social matrix that considers the gains (benefits) and losses (costs) of an act from the perspective of a donor and putative recipient or recipients (figure 1). Moving clockwise from the top left, *altruism* arises when the donor incurs a loss but delivers a gain to the recipient. *Spite* occurs when both the donor and the recipient incur losses, but typically, the cost to the recipient outweighs the cost to the donor. *Cooperation* arises when both donor and recipient accrue gains. Finally, *selfishness* arises when the donor gains, but the recipient loses. Needless to say, there are several important distinctions within each of these cells, but critically for our purposes, are the differences within the cooperation cell. In particular, some forms of cooperation entail joint action and mutual, simultaneous benefit, whereas others (i.e. reciprocity) entail delays.

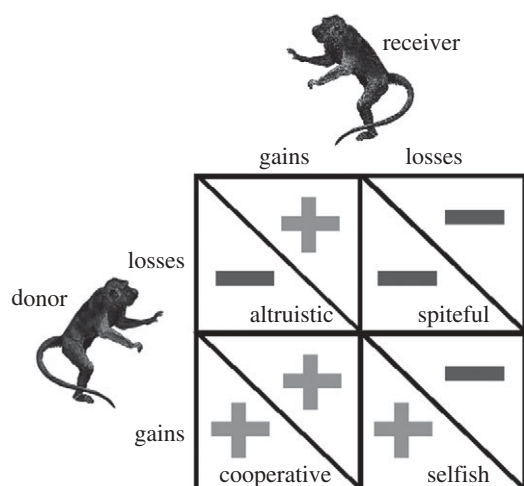* Author for correspondence (mdh102559@gmail.com).

Figure 1. A social matrix of interactions based on gains (plus) and losses (minus) to the donor and recipient.

Based on the clarity of these ideas, a torrent of empirical research soon emerged, confirming the logic of the gene's eye view of sociality. There were, however, two noticeable puzzles: though studies of insects, fishes, amphibians, birds and mammals explored the issues of reciprocity and spite, there is, at best, only a few studies that provide the necessary evidence, and many authors have concluded that there is no evidence at all (Foster *et al.* 2001; Hammerstein 2003; Hauser 2006; Jensen *et al.* 2006; Noe 2006); these authors, and the reviews that they have written, generally distinguish within- from between-species interactions, and consequently, do not address the much more significant evidence for reciprocity among, for example, cleaner fish and their hosts (Bshary & Grutter 2005, 2006). To explain this apparent phylogenetic gap, especially why we may be one of the only species to engage in reciprocity and spite with members of our own species and we turn first to a re-analysis of the logic of reciprocity and spite, focusing on both ultimate pressures and proximate requirements. We argue that accounts of the evolution of reciprocity and spite that neglect the requisite mechanisms for these social behaviours will fail. Similarly, studies that attempt to describe the underlying mechanisms without considering why such mechanisms evolved will fail as well. To understand the evolution of reciprocal altruism and spite, both proximate and ultimate factors must be considered. Based on a review of some recent work on human and non-human animals, we show that only our own species evolved under conditions that favoured reciprocal altruism and spiteful interactions, and importantly, evolved the brains to carry out such behaviours, even early in life.

2. RECIPROCAL ALTRUISM

Trivers developed his adaptationist's analysis of reciprocity (primarily direct as opposed to indirect) by laying out the structure of the interaction, targeting its economic, temporal and conditional properties. Specifically, reciprocity will evolve if:

(i)   the cost associated with helping is small relative to the benefit obtained by the recipient;

(ii)   the initial act of helping is contingent upon receiving help in the future;

(iii)   there is a time lag between the initial act of helping and the reciprocated act.

The first conditional is, fundamentally, the biological definition of altruism. The second conditional links the two altruistic acts, setting up the original altruistic act as an if-and-only-if contingency; this move establishes direct reciprocity as a selfish behaviour. The third conditional places a waiting period between the originally altruistic act and the reciprocated act.

From a conceptual and modelling perspective, the three core conditions for the evolution of reciprocal altruism are clear enough. From an empirical perspective, however, they are less clear, at least in terms of the kind of evidence that would constitute a sufficient test. For example, a considerable amount of research aimed at uncovering evidence of reciprocity in animals has focused on grooming (Seyfarth & Cheney 1984; Hart & Hart 1989; Hemelrijk & Luteijn 1998; Barrett *et al.* 1999; Schino *et al.* 2007; Gumert & Ho 2008; Schino & Aureli 2008). In some studies, analyses focus on the exchange of grooming for grooming, whereas others explore the exchange of grooming for other commodities, such as support in coalitions or opportunities for co-feeding. Several studies show that animals tend to groom most of those who groom them. That is, there is a positive correlation between the time that any given animal grooms another and the amount of time that they are groomed. Such analyses are problematic because correlations establish only an association, not the contingent nature of reciprocity. Further, while reciprocal altruism *could* be the mechanism in play, more parsimonious explanations should be employed until contingency can be demonstrated.

The difficulty of demonstrating contingency has come to light in a recent debate focused on mobbing behaviour in pied flycatchers (*Ficedula hypoleuca*). Pied flycatchers attempt to drive predators away by mobbing them and will assist neighbouring groups that initiate a mobbing response. In a series of elegant experiments, Krams *et al.* (2008) showed that pied flycatchers are more likely to assist neighbours who have assisted them in the recent past than those who have refused to assist them. The authors interpret their results as evidence of reciprocity (Krams *et al.* 2008; Wheatcroft & Price 2008). In response to this interpretation, Russell & Wright (2008) argued that although this behaviour is consistent with what would be expected in a reciprocally altruistic relationship, their experiments failed to demonstrate that the reciprocators' behaviour was contingent on an initially altruistic act. Instead, one can interpret these results as evidence for by-product mutualism (see Wheatcroft & Krams 2008 for response). This debate illustrates that simple and more common forms of cooperation should be evoked to account for seemingly reciprocal behaviour until all three components of reciprocal altruism are satisfied.

In addition to the three core conditions for the evolution of reciprocal altruism, Trivers also noted that reciprocity was most likely to evolve in highly social

species that are long-lived and have the opportunity to engage in repeated interactions with the same individuals. These added conditions were necessary because if the odds of a future encounter are low, then the contingency conditional is stripped away, allowing for the possibility that one's partner will fail to reciprocate; such failures could arise due to natural circumstances (e.g. frequent emigrations, low survival rate), making the recipient *unable* to reciprocate, or to more strategic and planned situations as when the recipient *decides to renege* because of more profitable opportunities. Building on these ideas, Nowak (2006) recently unified several models for the evolution of cooperation, arguing that direct reciprocity evolves when the probability that another interaction between altruist and recipient exceeds the cost–benefit ratio of the altruistic act. Thus, a common assumption or consideration in virtually all models of reciprocity is that individuals have the opportunity to interact with others, frequently, and with fairly dependable outcomes with respect to exchanging resources.

A wide variety of mammals and birds satisfy these life history and demographic conditions. That is, in a number of species, individuals live for many years, in relatively stable social groups and with numerous opportunities to interact. That said, a fundamental question, and one for which we have little understanding, is whether animals living in such groups are sufficiently dependent upon reciprocal interactions among genetically unrelated individuals to favour the evolution of reciprocity. Animals may have multiple opportunities to help non-kin, and to be helped by them, but selection on such relationships may be weak because most of the time animals can rely upon aid from close kin. For example, in Wilkinson's (1984) classic study of reciprocity among vampire bats (*Desmodus rotundus*), the vast majority (over 90%) of blood regurgitations arise between either mother and daughter ($r = 0.5$) or grandparent and grandchild ($r = 0.25$). The remaining cases among individuals with lower degrees of relatedness contribute little, at least in terms of the number of observations, and as in other studies of reciprocity, it is possible that these are instances of mistaken kin recognition (Coyne & Sohn 1978; Hammerstein 2003; Hauser 2006). Critically, therefore, it appears that reciprocity may be only weakly selected because in most animal societies, individuals can rely on kin-based helping to survive.

An additional evolutionary (ultimate) consideration is the type of resource exchanged, and the relative benefits of receiving it. Whatever the currency or resource type, interactants must be able to quantify it, including the relative costs of giving it up and the relative benefits of receiving it. As Whitlock *et al.* (2007) have noted, for resource sharing by means of reciprocity to evolve, the fitness value of the resource must differ between reciprocating partners, and this differential must reverse at some point in the interaction. That is, at time T, resource R is worth more to player 1 than it is to player to 2, but at time T + 1, R is worth more to player 2 than it is to player 1. Based on a series of models, Whitlock and co-workers (p. 1774) show 'that the conditions for

the evolution of resource sharing by reciprocity will become extremely difficult to satisfy. In all but a few cases, resource sharing is unlikely to evolve by reciprocity, but sharing may evolve readily via kin selection.'

In sum, several recent theoretical analyses and modelling efforts suggest that on ultimate grounds, reciprocity is unlikely to evolve and be selected for in most, if not all animal societies. Most animals societies are small, consist of a significant number of kin, and the differential in resource value among non-kin is insufficient to put pressure on reciprocal relationships. The opposite seems to be the case for most human societies, including those that appeared in our early origins.

Thus far, we have focused on ultimate considerations. The proximate prerequisites for reciprocity are no less significant and, we suggest, impose substantial constraints on its evolution and stability over time. Had these been mapped out in detail at the start, theorists may have predicted that reciprocal altruism would not evolve in animals! Trivers was, of course, sensitive to many of the mechanistic requirements for reciprocity, pointing to the importance of individual recognition, memory for prior interactions, and quantification of costs and benefits. Other mechanisms are also important, but were not considered in early writings, including the ability to delay gratification and read intentions, processes we discuss in greater detail below. There is no question that these cognitive abilities are firmly in place in adult humans. And although some of these capacities are also in play in some animals, we suggest that they are weakly integrated with each other, thus limiting the ability to both initiate and maintain stable reciprocal exchanges.

We first provide a brief, but critical description of three experiments on reciprocal altruism in animals; each provides some support for reciprocity, but also reveals the limits of this work and of the evidence to date. We then turn to a discussion of the mechanisms required to support reciprocity, showing that although the pieces are largely in place, they fail to combine in the context of initiating and maintaining a reciprocal relationship.

In a clever experiment with captive jays (*Cyanocitta cristata*), Clements & Stephens (1995) set up an operant experiment in which individuals were either playing the role of cooperator or defector. Individuals paired-off and started with a payoff matrix that simulated either mutualism or a Prisoner's Dilemma. Mutual cooperation yielded the highest payoff for each in the mutualism game, whereas defection–cooperation yielded the highest payoff for the defector in the Prisoner's Dilemma game (figure 2). Subjects played several rounds of one game before switching to the other. The results were striking and clear: subjects rapidly gravitated to cooperate–cooperate in the mutualism game, but precipitously turned to defect–defect in the Prisoner's Dilemma game. Thus, subjects were able to maximize individual payoffs in the mutualism game, but suffered relatively large costs in the Prisoner's Dilemma. Critically, in a follow-up experiment by Stephens *et al.* (2002) in
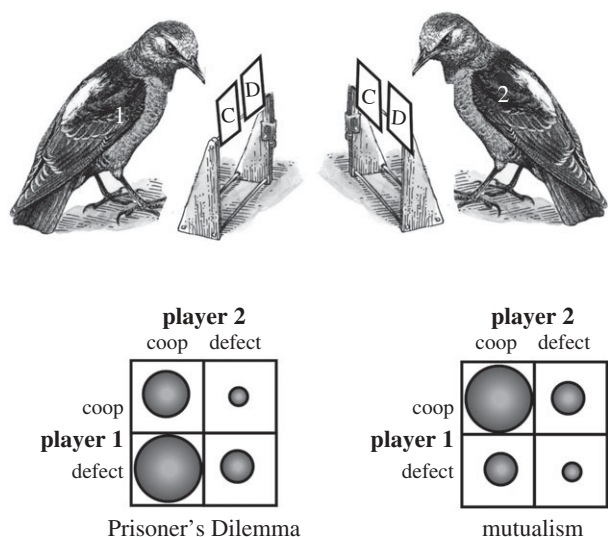
Figure 2. Cooperative games among captive jays (Clements & Stephens 1995). Each jay had access to a cooperate (C) and defect (D) button, each associated with a different payoff. Pairs of jays played several rounds of either the Prisoner's Dilemma game or the mutualism game, each associated with a different payoff matrix. In the matrices above, the size of the circle represents the relative payoffs from the perspective of player 1. Thus, in the Prisoner's Dilemma, the highest payoff to player 1 arises with defect–cooperate. For mutualism, the highest payoff to player 1 arises with cooperate–cooperate.

which individual payoffs were delayed, jays were able to solve the Prisoner's Dilemma and stabilize on cooperate–cooperate. That is, by taking away the temptation to immediately obtain the potentially largest payoff—defect (against cooperate)—the jays were able to settle on the long term, but more profitable strategy of cooperate–cooperate. As Stephens and colleagues noted, however, this capacity to solve the Prisoner's Dilemma must be placed in the context of a highly unnatural testing environment, one that would never arise in the wild. In other words, though jays have the cognitive ability to solve the Prisoner's Dilemma, they required thousands of opportunities to interact over a short period of time, as well as enforced delays for all reward payouts. Such a situation would never arise under natural conditions, perhaps for any animal.

In a study of cotton-top tamarins (*Sanguinus oedipus*), Hauser *et al.* (2003) attempted to test three properties of a reciprocal relationship: altruistic contingency, reputation tracking and distinguishing intentional from accidental outcomes. Genetically unrelated tamarins played in four different games, each requiring an actor to decide whether to pull a tool that would deliver food to either self, the other or both (figure 3). In each game, there were 24 trials per session, with each subject playing 12 alternating trials. In game 1, individual subjects played against one of the two trained stooges, one *nice* cooperative tamarin trained to pull the tool 100 per cent of the time, and one *mean* uncooperative tamarin trained to never pull the tool. In this game pulling the tool resulted in one piece of food for the recipient and no food for the actor, and thus, was considered an
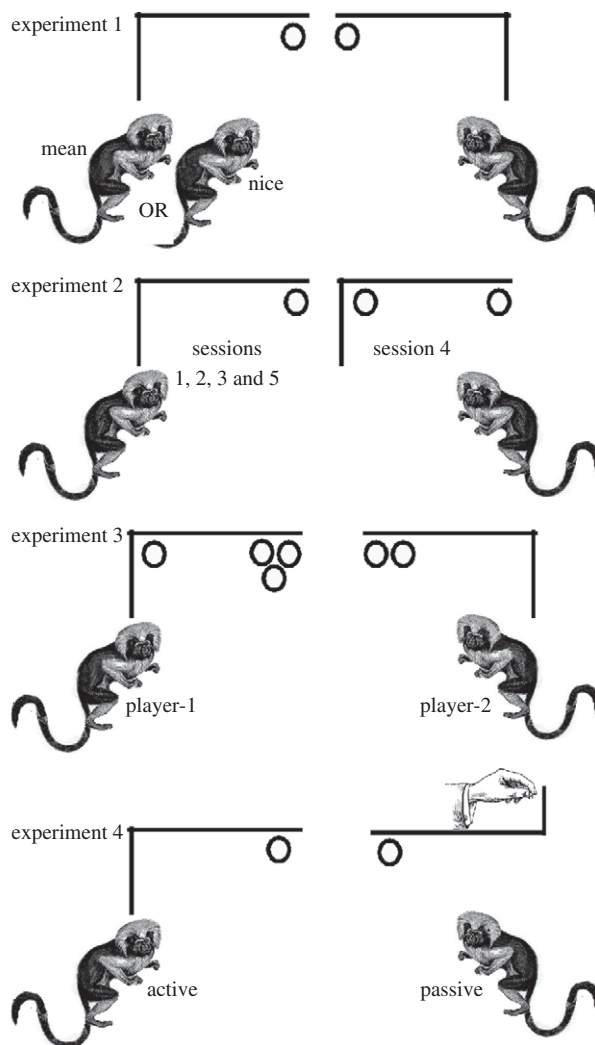
Figure 3. Four different games played by genetically unrelated cotton-top tamarins. In experiment 1, each subject played alternating sessions (24 trials, 12 trials each) with either a nice stooge (trained to pull the tool 100% of the time) or a mean stooge (trained to pull the tool 0% of the time). In experiment 2, subjects played a reciprocating altruism game (i.e. no food for actor but one piece for recipient) for the first three sessions, the fourth session as byproduct mutualism game (i.e. a piece of food for the actor and the recipient), and the final fifth session as a reciprocating altruism game. In experiment 3, one subject was assigned to the player-1 position (pulling provided one piece to the actor and three pieces to the recipient) and one to the player-2 position (pulling provided no food for the actor and two pieces to the recipient). In experiment 4, the active tamarin could pull as an altruistic act (no food for self, one for the recipient), while the passive tamarin had no opportunity to pull; instead, when the tool switched to the passive tamarin's side, the experimenter pushed the tool on 100 per cent of the trials, thus mimicking the payoff structure for the nice stooge in experiment 1.

altruistic act targeting a genetically unrelated individual. Results showed that subjects pulled significantly more often when paired in games with the nice stooge than the mean stooge. This suggests that tamarins can distinguish recipients based on their cooperative tendencies, and respond contingently. However, two criticisms immediately arise. First, identifying cooperators requires an ability to recognize the partner's

motivations—do they incur a cost in order to cooperate (altruism) or do they only cooperate when they also benefit (mutualism). Experiments 2 and 3 explored this possibility. Second, subjects may pull more when they themselves receive food, and this situation arises most when playing against the nice stooge who always delivers food. In other words, the higher rates of pulling when paired with the nice stooge is simply a reflection of the higher rates of reinforcement, a situation that could just as easily be achieved by a machine delivering food. Experiment 4 attempted to test this alternative account.

In experiment 2, two genetically unrelated tamarins played five sessions, with sessions 1, 2, 3 and 5 associated with altruistic actions (the actor's pull results in no food for self but one piece of food for the recipient) and session 4 associated with food delivery as a byproduct (the actor's pull results in one piece of food for self and one for the recipient). Here, we expected the rate of pulling to decline from sessions 1–3, then rise to 100 per cent in session 4 (i.e. the actor selfishly gains regardless of food delivered to recipient). If subjects perceive session 4 as altruistic—that is, each pull consists of an attempt by the actor to give (at some cost) food to the recipient—then actors should pull at higher rates in session 5 than in session 3; in other words, session 4 should kick start cooperation in session 5. In contrast, if subjects perceive session 4 as a case of byproduct mutualism, where food is obtained as an accidental byproduct of otherwise selfish behaviour, then pulling rates should decline or remain the same in session 5. Results showed that pulling rates in session 5 were not significantly different from session 3, supporting the byproduct mutualism hypothesis, and rejecting the more general interpretation of experiment 1 in terms of reinforcement history.

To push on the interpretation of experiment 2, experiment 3 set up different payoffs for each of the two players; although each tamarin played in both the player-1 and -2 roles across different games (i.e. different partners), each player kept their role within a game. In the player-1 position, pulling the tool brought one piece of food to the actor and three pieces to the recipient. In the player-2 position, pulling the tool brought no food to the actor and two pieces to the recipient. Thus, if both players pulled, each would obtain three pieces of food after a round of two trials. The interpretation of the results of this game hinge on player-2's analysis of player-1's pulls. If player-2 perceives player-1's pulls as selfish (and we assumed player-1 would pull on virtually 100% of the trials as the act delivers one piece of food to self), then player-2 should never pull as the receipt of three pieces of food from player-1 is an accidental byproduct. In contrast, if player-2 perceives player-1's act of pulling as intentional in some sense, driven by the goal of giving player-2 three pieces of food, then player-2 should pull as a reciprocated altruistic gesture. Results showed that player-2 virtually never pulled, supporting the first interpretation, and suggesting that tamarins make economically relevant decisions on the basis of subjects' underlying motivations as opposed to the outcomes alone. This

conclusion is supported by other studies of non-human primates in both economic decision-making contexts as well as other situations (de Waal 2000; Call *et al.* 2004; Rochat *et al.* 2007; Warneken *et al.* 2007; Buttelmann *et al.* 2008; Lakshminaryanan *et al.* 2008; Wood *et al.* 2008).

In the final study, tamarins were placed in a set up that was virtually identical to experiment 1, except that instead of the *nice* tamarin stooge, a human experimenter pushed the tool towards the tamarin, providing a reward structure that was identical to the nice stooge. That is, after every trial in which the *active* tamarin had an opportunity to altruistically give food to a *passive* tamarin recipient, the tool then changed sides and now, the experimenter pushed the tool towards the *active* tamarin, giving him or her a piece of food. Once again, if tamarins only attend to the reinforcement structure of games (i.e. only outcomes as opposed to the means by which they are attained), then the active tamarin should pull at a high rate, comparable to the rates observed in experiment 1 with the *nice* stooge. In contrast, if tamarins pay attention to the means, and recognize that the passive tamarin played no role at all in the delivery of food, then the active tamarin should rarely pull. Results showed that tamarins rarely pulled in this condition, with rates approximating to those observed in experiment 1 for the *mean* stooge.

Together, these studies suggest that tamarins are sensitive to some of the important proximal ingredients that enter into reciprocity, including altruistic contingency, reputation tracking and distinguishing the means by which outcomes are obtained. That said, when one explores the longer term pattern of cooperation observed in these experiments, it is clear that tamarins are incapable of sustaining reciprocity as even a rather brief period of defection causes the cooperative relationship to unravel. In particular, based on a game theoretic analysis of the tamarin results from the non-stooge games, it is clear that after two consecutive rounds of defection, tamarins stop pulling in the altruistic condition, and never recover the reciprocally cooperative relationship (Chen & Hauser 2005). Thus, although tamarins may have some of the cognitive prerequisites for reciprocity, these capacities appear insufficient to sustain reciprocity. Moreover, and paralleling the study on jays, the reciprocity observed among tamarins only emerges under fairly artificial conditions, including the presentation of discrete packages of food, highly predictable periods of interaction, and with individuals trained to be *pure* cooperators or defectors.

The final study of reciprocity concerns a set of experiments on captive chimpanzees (*Pan troglodytes*). In thinking about the cognitive building blocks of reciprocity, as well as the evolutionary pressures that would select for this kind of cooperation, chimpanzees would seem to be the most promising of animal species (Stevens & Hauser 2004; Melis *et al.* 2008). For example, under natural conditions, chimpanzees show significant levels of cooperation in the context of coalitions during aggressive competition (both within- and between-groups) as well as during hunting. Although chimpanzees do not live in large

groups, the fission–fusion nature of their social organization means that they cannot always rely on particular individuals for help. Perhaps as a result, recent molecular and behavioural research shows that chimpanzee cooperation occurs between kin and non-kin (Langergraber *et al.* 2007). Added on to these ultimate considerations are experiments and observations targeting proximate mechanisms. In particular, chimpanzees have the capacity for numerical quantification (Boysen & Berntson 1995; Boysen *et al.* 1996; Kawai & Matsuzawa 2000; Beran *et al.* 2008), show significant levels of delayed gratification (Evans & Beran 2007; Rosati *et al.* 2007), inequity detection (Brosnan *et al.* 2005; Brauer *et al.* 2006), prosocial helping in non-food contexts (Warneken & Tomasello 2006; Warneken *et al.* 2007), vengeance (Jensen *et al.* 2007b), discrimination of intentional and accidental actions (Call *et al.* 2004), selectively choosing previous collaborators over non-collaborators in joint cooperation tasks (Melis *et al.* 2006), and recognizing individuals by face and voice (Parr 2003).

Taking advantage of these capacities, Melis *et al.* (2008) designed an elegant series of experiments with captive chimpanzees, asking whether subjects would preferentially choose to reciprocate an altruistic action towards a previously nice and cooperative stooge over a previously mean and uncooperative stooge. Underlying these experiments was prior evidence that chimpanzees could recruit collaborators in a joint action task (i.e. a task in which two subjects must work together to attain a reward, and where defection by one leads to an overall failure such that no one attains any reward), and preferentially select the most collaborative collaborator (Melis *et al.* 2006). In experiment 1, subjects first learn that the nice stooge always provides them with access to a rope, that if jointly pulled, brings food, whereas the mean stooge never gives them access. Following this exposure phase, subjects are then presented with an opportunity to allow either the nice or mean stooge to join them at the pulling tray. In the first block of trials, one out of eight subjects picked the nice stooge, two were indifferent, and five actually picked the mean stooge. In the second block of trials, three subjects picked the nice stooge (only one with a strong preference), two were indifferent and three picked the mean stooge. Although there was a slight increase in the preference for the nice stooge over the baseline period, this effect was only just significant at the $p < 0.05$ level and with a one-tailed test. Thus, based on analyses of individual preferences, there was, at best, only weak evidence of reciprocity.

In a second series of experiments (figure 4), the nice stooge altruistically opened the door for the subject to get food, whereas the mean stooge opened the door to selfishly get food for himself. As in experiment 1, the question here was whether subjects would give the nice stooge more frequent access to the pulling tray when compared with the mean stooge. Pooling across individuals, there was no evidence that subjects opened the door more often for the nice than the mean stooge. On an individual level, only one subject out of eight showed a significant difference between stooges in the predicted direction, opening the door on every trial for the nice
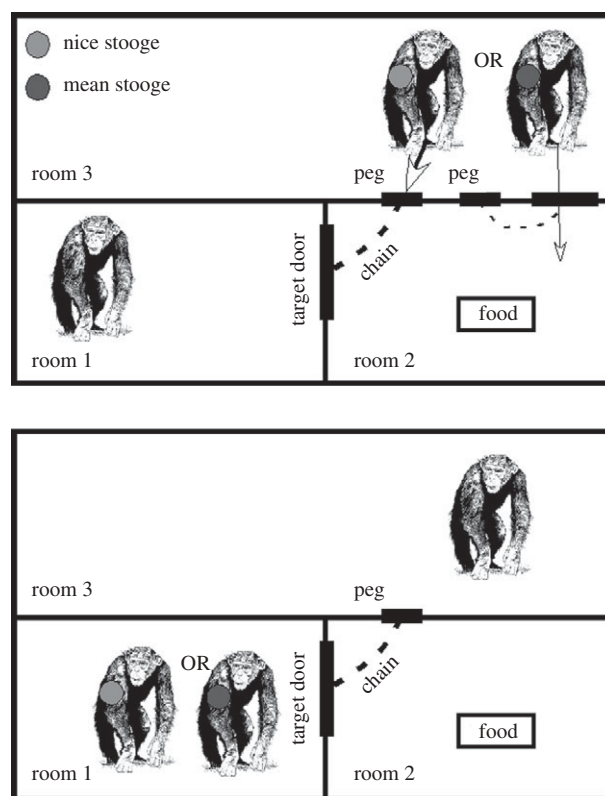

Figure 4. Experiments on reciprocity in captive chimpanzees by Melis *et al.* (2008). The top panel shows the initial set up used to establish the reputation of the nice (green dot) and mean (red dot) stooges. When the subject removed the peg for the nice stooge, this individual entered and collaborated in pulling in the tray with food. When the subject removed the peg for the mean stooge, this individual entered and consumed the food alone. The bottom panel reveals the set up for the critical test phase in which the subject has the opportunity to open the door, on separate trials, either the nice or mean stooge.

stooge and never for the mean stooge. In summary, as the authors note, this study provides only weak evidence of reciprocity in chimpanzees.

What, therefore, is missing in the toolkit of animal cognition that appears in human cognition? Why, in most studies, are animals apparently incapable of engaging in reciprocal altruism, and in those cases where there is some evidence, why are the effects weak and dependent upon quite extreme experimental setups that rarely, if ever, arise under natural conditions (de Waal 2000; Stephens *et al.* 2002; Hauser *et al.* 2003; Melis *et al.* 2008)? The research on chimpanzees reveals the fundamental contrast with humans. Here is a species with the capacity to delay gratification, quantify potential payoffs, detect inequities and punish individuals for norm violations, but these ingredients do not combine to create a system for reciprocity. By contrast, human children understand norms of reciprocity by early elementary school (De Cooke 1992) and use these norms to establish friendships (for review see Eisenberg *et al.* 2006). The ability to delay gratification (Mischel *et al.* 1989) and integrate intentions and outcomes may take longer to develop (Sutter 2007), but children appear fully capable of reciprocal altruism by about

9 years of age. Our proposal is that animals generally lack the capacity to integrate cognitive functions required for reciprocal altruism, while for humans that integration occurs as a normal part of development (see §4).

What ultimate pressures might encourage a cognitive system that enables reciprocal altruism to evolve? One possibility is that over the course of human evolution, the gradual expansion of small kin groups into large stable groups of unrelated individuals led to the evolution of reciprocity, and subsequently, strong demands on the capacity to detect and punish cheaters. As Boyd and his colleagues (Boyd & Richerson 1992; Boyd *et al.* 2003) have noted, stable cooperation requires not only punishment of cheaters but punishment of those who do not punish cheaters. And because punishment plays such a critical role in human social interaction, it appears to *spill over* into other forms of social behaviour, including spiteful actions (see §3) that are often accompanied by feeling good about another's misfortunes, an emotion that only the German language has assigned to a single word: *Schadenfreude*. Thus, when we punish others for what we see as a wrongdoing, including cases where we incur a personal cost for imposing such punitive measures, our actions are personally rewarding as evidenced by the fact that the reward areas of the brain are significantly activated (de Quervain *et al.* 2004).

Thus far, there is little evidence that chimpanzees, or any other species, directly punish (i.e. as opposed to indirect punishment in the form of ostracism, which may be present in animals) individuals who fail to cooperate, and in fact, explicit experimental and observational evidence that they do not (Heinsohn & Packer 1995; Wilson *et al.* 2001; Jensen *et al.* 2007*b*). Thus, for example, although chimpanzees will move out in groups during border patrols and when confronted by territorial intruders, there is no cost to individuals that lag behind (Wilson *et al.* 2001). What makes the absence of punitive action in these cooperative contexts of interest is that punitive behaviour arises in other, non-cooperative situations (Jensen *et al.* 2007*b*). For example, chimpanzees will take food away from another chimpanzee who has taken food from them in the recent past. This shows that chimpanzees can engage in a form of 'vengeful punishment' when a norm violation has arisen (e.g. theft of property), but do not tap this capacity in the context of cooperation.

3. SPITE

As Gardner & West (2004) noted in the most recent review of the literature, spite is the 'relatively neglected ugly sister of altruism' (p.1195). Although there has certainly been less research on spite than on other forms of social interaction (see figure 1 and the matrix of social interactions), we suggest that the lack of research has perhaps less to do with neglect than with the ultimate and proximate considerations required for spite to evolve, and the confusions in the literature concerning the criterion for demonstrating spite (Foster *et al.* 2001; West *et al.* 2007). Part of

this problem stems from the fact that different disciplines have approached spite from different angles. Evolutionarily oriented researchers have focused on the ultimate conditions for the evolution of spite, targeting the significance of genetic relatedness and fitness consequences. By contrast, much of the more psychologically oriented research has been done by behavioural economists targeting the proximate processes that underlie fairness and the detection of inequities (Kirchsteiger 1994; Fehr & Fischbacher 2005). Thus, to distinguish these two approaches to spite, we call the evolutionary view *genetic spite* and the proximate approach *psychological spite*.

Following on the heels of his conceptual analysis of the evolution of altruism by means of kinship, Hamilton (1970, 1971) argued that genetic spite—an action that imposes a significant cost on another with either no fitness costs or a relatively small cost to the actor—can evolve as long as actors and agents are *negatively related*. Negative relatedness arises when $r < 0$, meaning that the odds that actor and agent share genes in common is less likely than with a randomly selected individual from the population. Selection can therefore favour genetic spite among negatively related individuals because it reduces the frequency of competitive genes in the population. Wilson (1975) considered an alternative route to the evolution of genetic spite by considering the potential role of a third-party, non-interacting observer. As defined, if observer O is related to actor A, and if the benefit to O from A's spiteful action outweighs the costs to A and the recipient R of such spite, then genetic spite can evolve. Interestingly, though both Hamilton and Wilson considered the evolution of genetic spite as possible, both argued that it was unlikely to represent a significant form of social interaction in animals. For Hamilton in particular, the prediction that genetic spite would be relatively rare was based on the supposition that socially living animals (i) should infrequently find themselves in a situation of negative relatedness; (ii) most likely lack the requisite fine-grained kin recognition mechanisms; and (iii) typically engage in actions with significant costs as opposed to no or little cost, as the theory demanded.

During the first 30 years or so of empirical exploration post-Hamilton's analysis, several examples of genetic spite were put forward in the animal literature. Most, if not all, however, could be better explained as cases of selfishness or as a form of punishment where the actor's personal fitness ultimately benefits from the spiteful interaction (Clutton-Brock & Parker 1995). That is, though the actor's behaviour indeed imposed a cost on the recipient, the actor typically gained from this action, either immediately or after some delay. For example, in several species of birds and primates, adults aggressively harass and even attack juveniles, or copulating couples. Though such attacks are clearly associated with a cost in terms of energy loss to the attacker, and presumably even greater costs to the juveniles and mating couples, the attacker reaps competitive gains (Foster *et al.* 2001; Gardner & West 2004).

In behavioural economic circles, the problem of psychological spite has focused more on the relevant

proximate ingredients, and in particular, on notions of fairness with respect to payoffs. Thus, as Fehr & Fischbacher (2005) note, 'A spiteful or envious person *always* values the economic payoff of relevant reference agents negatively. The person is, therefore, willing to decrease the economic payoff of a reference agent at a personal cost to himself... irrespective of the payoff distribution and irrespective of the reference agent's fair or unfair behaviour.' On this view, there is, contrary to the evolutionary approach, no burden to show negative relatedness among the players. However, there is a need to provide evidence that an agent imposes a cost on another by incurring a personal cost, and with such evidence, explain the individual utility of spiteful actions, starting with assessing the cognitive factors involved.

Two recent studies by Jensen *et al.* (2007*a,b*) set out to provide a test of psychological spite in chimpanzees. In the first study, they set up a task that was designed to mimic important details of the payoff structure of an ultimatum game. In particular, in each of the four conditions, the donor had the opportunity to choose between two payoff distributions and the recipient had the option of accepting or rejecting the offer; acceptance led to the donor and recipient gaining access to the associated reward, whereas rejection led to no food for either donor or recipient. The four option pairings were always anchored against eight for the donor and two for the recipient: (1) 5 versus 5; (2) 2 versus 8; (3) 8 versus 2; and (4) 10 versus 0. Thus, condition-1 provided a fair option, condition-2 a hyper-fair option, condition-3 no option, and condition-4 a hyper-selfish option. There were two central results. Donors typically proposed the option that provided them with the highest returns, and critically, recipients accepted all non-zero offers. That is, and in contrast with human subjects similarly tested, chimpanzees did not act spitefully, rejecting unfair offers at a personal cost, and an even greater cost to the donors. However, the chimpanzee responders did draw the line at offers of zero in condition 4, refusing to endorse the division when they had nothing to gain.

In the second task, briefly described above in the reciprocity section, Jensen and colleagues set up a collapsible table between an actor and either an empty room or another chimpanzee. In each condition, the actor had access to a rope that allowed him or her to collapse the table, dumping any item on top of the table onto the floor. In the two critical conditions, the actor either faced another chimpanzee on the other side of the table with food or an empty room. If chimpanzees are psychologically spiteful, then they should actively dump the table (i.e. pull the rope) when the other individual has access to food and they don't; they should pull the rope more often in this social condition than when there is inaccessible food on the table, but the adjacent room is empty. In contrast, if chimpanzees are simply frustrated by their own inability to access the food, then they should pull the rope equally frequently in the social and non-social conditions. Results strongly confirmed the latter hypothesis: chimpanzees pulled equally frequently in the social and non-social conditions,

revealing that frustration, but not psychological spite, drives their behaviour. Thus, and as noted in the previous section, chimpanzees are capable of vengeance, but not psychological spite.

For psychological spite to evolve, subjects must be willing to incur a direct cost in order to decrease the welfare of another and even in cases where there is no cost to the self, subjects must accept the possibility that the recipient will retaliate. Recipients will not take an attack sitting down so to speak. Thus, psychological spite is a risky behaviour that requires the computation of current and future costs and benefits, capacities that seem to be in play at some level, in both apes and corvids (Mulcahy & Call 2006; Raby *et al.* 2007). Additionally, psychological spite requires some degree of inhibitory control. Using the example of chimpanzees above, actors would have to inhibit their frustration when food is out of reach and only collapse the tray when a peer enjoys the advantage of access to the food. Similarly, in the ultimate game, chimpanzees would have to resist the temptation of a small reward in order to reject a disadvantageous outcome. Though these inhibitory and computational capacities are, to some extent, present in chimpanzees, they once again seem weakly integrated into a single system.

In contrast to other species, humans clearly engage in spiteful actions, and certainly engage in behaviour that is costly to self and even more costly to others (Trivers 1971; Fehr & Gachter 2002; Fehr & Henrich 2003; Henrich *et al.* 2006). Again, experiments in behavioural economics provide important details. As noted above, in the ultimatum game chimpanzees do not reject inequity—as long as responders receive some reward, they endorse unequal allocations of food. However, human versions of this game became an important test case for economic theory precisely because responders are willing to reject non-zero offers. Review of the many variations of ultimatum games leads to the conclusion that responders reject offers of 20 per cent or less of the stake about half of the time (Camerer 2003). The high rejection rates in the game constitute a form of punishment (as opposed to psychological spite which is not reactive to a norm transgression)—proposers have intentionally violated some norm of fairness which responders are willing to punish at a cost to themselves. However, responders continue to reject low offers even when intentions have been removed. When unequal offers are generated by a computer (Blount 1995) or by a roll of dice (Falk *et al.* 2008), over 60 per cent of participants still reject offers that favour the other player.

Notably, responses to inequity appear to be shaped by culture. Henrich *et al.* (2005) found wide variation in levels of rejection in ultimatum games across small-scale societies. Their study showed that while individuals in all societies demonstrate a sensitivity to fairness, the perception of what constitutes a fair offer is shaped by cultural norms. For example, in two societies in particular—the Au and Gnau of Papua New Guinea—responders even rejected very generous offers. While the authors explain this phenomenon in terms of cultural practices such as competitive gift-giving, what remains to be seen is what aspects of

human cognition underlie the willingness to reject any non-zero offer, prior to the influence of culture.

Research on children's development provides an opportunity to assess the emergence of different cognitive capacities in relation to behavioural outcomes. Economic experiments with children have shown that 9-year-olds are willing to reject unequal rewards in ultimatum games even when the proposer has no choice in the allocation. Notably, children of this age reject inequity both when they receive less than a peer and, to a lesser extent, when they receive more (Sutter 2007). This latter result differs markedly from adults of the same Western culture who were quite willing to accept generous offers (only 3% rejected). To assess the developmental origins of children's aversion to inequity, we designed a series of competitive games for young children between the ages of 4 and 8 (Blake *et al.* submitted). In these experiments, one individual always played the role of donor, and the other the role of receiver. Subjects were unfamiliar and genetically unrelated. For a given game, two children sat across from each other with the test apparatus placed in between them. The donor had access to two levers that controlled the action of two plates, one proximally associated with the donor, the other with the receiver. If the donor pulled the green lever, the plates tipped up and caused rewards (candy) to roll, respectively, towards the donor and recipient. If the donor pulled the red lever, the plates tipped down and caused the rewards to roll into a bowl and disappear. Thus, pulling the green lever was associated with accepting the distribution of rewards, whereas pulling the red lever was associated with rejecting the rewards. Each pair of children played a total of 12 trials. Within a session, there were always six trials with an equitable distribution of one candy for the donor and one for the recipient. The other six trials were either set up as an *advantageous inequity* game (pulling the green lever brought four candies to the donor and only one candy to the recipient) or a *disadvantageous inequity* game (pulling the green lever brought one candy to the donor and four pieces to the recipient). In both the equity and inequity games, pulling the red lever caused all of the candies to roll out of reach, thus creating a round of no rewards for either child. Rejecting the rewards in the advantageous condition represents an example of self-sacrifice, whereas rejecting the rewards in the disadvantageous condition represents a case of psychological spite.

Pilot data collected thus far, and restricted to the equity and disadvantageous inequity conditions, indicate that children of all ages accepted the majority of the equity trials, thus demonstrating that one piece of candy was sufficiently motivating to continue distributing an equal reward to a peer. In the disadvantageous inequity condition, even the youngest children demonstrated a willingness to reject one piece of candy in order to prevent their peer from getting four pieces. Though preliminary, these results are surprising given that sacrificing any candy requires inhibitory control which develops slowly with age (Davidson *et al.* 2006) making sacrificial rejections much more likely for older children.

Placing these results in the context of the animal work discussed above, some striking patterns appear.

First, although young children tend towards self-interest they soon demonstrate a willingness to inhibit their desire for candy in order to prevent a peer from getting more than them—a form of psychological spite (see also Fehr *et al.* 2008). Notably, rejections in the disadvantageous condition were in reaction to the relative outcomes alone as opposed to being in response to actions of the peer. Thus, by the functional definitions in the payoff matrix (figure 1), even 4-year-olds engage in psychological spite.

Paralleling our comments in the section on reciprocity, it appears that psychological spite evolved in humans both because the requisite cognitive capacities were in place and because there was significant selection on punishing norm violators, even at personal cost; whether genetic spite also exists is unclear, and as in animal studies, may be difficult to demonstrate due to the challenges of proving negative relatedness and of demonstrating spite in isolation of a larger punishment interaction. But given the evidence for psychological spite, humans clearly evolved the ability to detect inequities, control immediate desires, foresee the virtues of norm following, and gain the personal, emotional rewards that come from seeing another punished. Though some animals have some of these capacities, only humans evolved a brain that integrates them into one system, and enables spiteful behaviour. On an ultimate level, we speculate that psychologically spiteful behaviour evolved in humans as a byproduct of selection on punishment. In particular, and as previously noted, punishment was favoured in human evolution because of the increasing significance of social norms, the increase in group size, and in particular, the relative increase in cooperation between genetically unrelated and unfamiliar others. These factors placed intense selective pressure on the capacity to detect and discourage defection. Though some animals punish others in certain restrictive contexts, punishment neither emerges as a means to enforce cooperation, nor does it appear necessary. Thus, the lack of punishment in a cooperative context, and the lack of psychological spite makes sense, if our account of the evolution of human spite as a byproduct of punishment is correct.

4. CONCLUSIONS

Darwin knew that altruism, and morality more generally, represent genuine puzzles in light of his theory of natural selection. Adopting the gene's eye view, these altruistic actions no longer represent a challenge to Darwin's logic. The costs of altruism are either neutralized by kinship or by the prospects of a reciprocally altruistic relationship. As well as this theoretical perspective works, it largely fails to account for the virtual absence among social vertebrates of reciprocal altruism and spite. In this essay, we have argued that consideration of proximate and ultimate factors leads to the prediction that reciprocity and spite (both genetic and psychological) should be rare or absent in non-human animal populations. In particular, we have suggested that non-human animals do not live in the kinds of societies that would create

strong pressure on individuals to require reciprocity to obtain help given that the density of kin is high, and thus, the probability of interacting with them is high as well; in other words, individuals can rely on their kin in times of need. Similar, ultimate level arguments apply to spite. That is, in most cases where an individual does something to impose costs on another, the underlying motivation for such behaviour is selfish. Given that there are many other ways other than spite to increase one's relative fitness, the costs of spite will very rarely be favoured.

Even if selection favoured reciprocity and spite, the mechanisms required to support these social interactions are, we have argued, largely absent in non-human animals. More specifically, even in cases where some of the relevant psychological components are in place, what is missing are the interfaces between these components. Thus, to support reciprocity, animals must be able to quantify the costs and benefits, time the returns, delay gratification, assess reputation, compute the contingencies and punish cheaters. Though many animals have the requisite skills of quantification, as well as the capacity to wait for beneficial returns and punish those who violate social norms, these capacities are not integrated into one functional system that can subserve reciprocal interactions. Similarly, though animals can sometimes behave in a purely prosocial manner (helping another at a cost and without expecting reciprocal returns; e.g. Warneken & Tomasello 2006; Warneken *et al.* 2007), and even though they can inhibit some actions with the goal of waiting for a more preferable outcome (Evans & Beran 2007; Rosati *et al.* 2007), they are not able to integrate these abilities in the service of not only detecting inequities, but acting upon them in such a way as to deliver a severe cost to another, while incurring a personal cost—psychological spite.

It is this capacity—the ability to create interfaces between different psychological processes—that is perhaps the hallmark of our uniquely human cognition (Hauser 2009). Though human thought, like animal thought, is built from modular processes, we have the distinctive capacity to integrate the outputs from these processes to create novel representations, or more generally, novel solutions to old and new problems. Thus, we have the capacity to quantify how much money we borrowed (number system), tag the cardinality with a linguistic symbol (language system), take the money and exchange it for some food (economic system), realize that we could have bought it for half the price at a nearby store (number, economics and language) and then experience outrage (emotional system) because the store owner ripped us off (moral system).

In sum, what enabled humans to engage in reciprocity and spite was not only a particular set of socioecological conditions that favoured such interactions, but also, a critical set of psychological mechanisms that made them possible in the first place. Thus, proximate and ultimate factors partnered up at some point in human evolution, paving the way to a species that could follow the golden rule and just as easily spite their neighbour.

REFERENCES

Barrett, L., Henzi, S., Weingrill, T., Lycett, J. & HIll, R. 1999 Market forces predict grooming reciprocity in female baboons. *Proc. R. Soc. Lond. B* **266**, 665–670.

Beran, M., Evans, T. & Harris, E. 2008 When in doubt, chimpanzees rely on estimates of past reward amounts. *Proc. R. Soc. B* **276**, 309–314. (doi:10.1098/rspb.2008.1027)

Blake, P. R., McAuliffe, K. & Hauser, M. D. Submitted. The development of inequity aversion and self-sacrifice in children.

Blount, S. 1995 When social outcomes aren't fair: the effect of causal attributions on preferences. *Organ. Behav. Human Decision Process.* **63**, 131–144. (doi:10.1006/obhd.1995.1068)

Boyd, R. & Richerson, P. J. 1992 Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethol. Sociobiol.* **113**, 171–195.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535. (doi:10.1073/pnas.0630443100)

Boysen, S. T. & Berntson, G. G. 1995 Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **21**, 82–86.

Boysen, S. T., Berntson, G. G., Hannan, M. B. & Cacioppo, J. T. 1996 Quantity-based inference and symbolic representations in chimpanzees (*Pan troglodytes*). *J. Exp. Psychol.: Anim. Behav. Process.* **22**, 76–86. (doi:10.1037/0097-7403.22.1.76)

Brauer, J., Call, J. & Tomasello, M. 2006 Are apes really inequity averse? *Proc. R. Soc. B* **273**, 3123–3128. (doi:10.1098/rspb.2006.3693)

Brosnan, S., Schiff, H. C. & de Waal, F. B. 2005 Tolerance for inequity may increase with social closeness in chimpanzees. *Proc. R. Soc. B* **272**, 253–258. (doi:10.1098/rspb.2004.2947)

Bshary, R. & Grutter, A. 2005 Punishment and partner switching cause cooperatve behaviour in a cleaning mutualism. *Biol. Lett.* **1**, 396–399. (doi:10.1098/rsbl.2005.0344)

Bshary, R. & Grutter, A. 2006 Image scoring and cooperation in a cleaner fish mutualism. *Nature* **441**, 975–978. (doi:10.1038/nature04755)

Buttelmann, D., Carpenter, M., Call, J. & Tomasello, M. 2008 Rational tool use and tool choice in human infants and great apes. *Child Dev.* **79**, 609–626. (doi:10.1111/j.1467-8624.2008.01146.x)

Call, J., Hare, B., Carpenter, M. & Tomasello, M. 2004 'Unwilling' versus 'unable': chimpanzees' understanding of human intentional action. *Dev. Sci.* **7**, 488–498. (doi:10.1111/j.1467-7687.2004.00368.x)

Camerer, C. F. 2003 *Behavioral game theory: experiments in strategic interactions.* Princeton, NJ: Princeton University Press.

Chen, M. K. & Hauser, M. 2005 Modeling reciprocation and cooperation in primates: evidence for a punishing strategy. *J. Theoret. Biol.* **235**, 5–12. (doi:10.1016/j.jtbi.2004.12.015)

Clements, K. C. & Stephens, D. W. 1995 Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. *Anim. Behav.* **50**, 527–535. (doi:10.1006/anbe.1995.0267)

Clutton-Brock, T. H. & Parker, G. A. 1995 Punishment in animal societies. *Nature* **373**, 209–216. (doi:10.1038/373209a0)

Coyne, J. A. & Sohn, J. J. 1978 Interspecific brood care in fishes: reciprocal altruism or mistaken identity? *Am. Nat.* **112**, 447–450. (doi:10.1086/283287)

Darwin, C. 1871 *The descent of man and selection in relation to sex.* London, UK: John Murray.

Davidson, M. C., Amso, D., Anderson, L. C. & Diamond, A. 2006 Development of cognitive control executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* **44**, 2037–2078. (doi:10.1016/j.neuropsychologia.2006.02.006)

Dawkins, R. 1976 *The selfish gene*. Oxford, UK: Oxford University Press.

De Cooke, P. A. 1992 Children's understanding of indebtedness as a feature of reciprocal help exchanges between peers. *Dev. Psychol.* **28**, 948–954. (doi:10.1037/0012-1649.28.5.948)

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schelhammer, M., Schnyder, U., Buck, A. & Fehr, E. 2004 The neural basis of altruistic punishment. *Science* **305**, 1254–1258. (doi:10.1126/science.1100735)

de Waal, F. B. M. 2000 Attitudinal reciprocity in food sharing among brown capuchin monkeys. *Anim. Behav.* **60**, 253–261. (doi:10.1006/anbe.2000.1471)

Eisenberg, N., Fabes, R. A. & Spinrad, T. L. 2006 Prosocial development. In *Handbook of child psychology*, vol. 3 (ed. N. Eisenberg), pp. 646–718. New York, USA: Wiley.

Evans, T. A. & Beran, M. J. 2007 Chimpanzees use self-distraction to cope with impulsivity. *Biol. Lett.* **3**, 599–602. (doi:10.1098/rsbl.2007.0399)

Falk, A., Fehr, E. & Fischbacher, U. 2008 Testing theories of fairness—intentions matter. *Games Econ. Behav.* **62**, 287–303. (doi:10.1016/j.geb.2007.06.001)

Fehr, E. & Fischbacher, U. 2005 The economics of strong reciprocity. In *Moral sentiments and material interests* (eds H. Gintis, S. Bowles, R. T. Boyd & E. Fehr), pp. 151–192. Cambridge, USA: MIT Press.

Fehr, E. & Gachter, S. 2002 Altruistic punishment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)

Fehr, E. & Henrich, J. 2003 Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In *The genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 55–82. Cambridge, USA: MIT Press.

Fehr, E., Bernhard, H. & Rockenbach, B. 2008 Egalitarianism in young children. *Nature* **454**, 1079–1083. (doi:10.1038/nature07155)

Foster, K. R., Wenseleers, T. & Ratnieks, F. L. 2001 Spite: Hamilton's unproven theory. *Ann. Zool. Fennici* **38**, 229–238.

Gardner, A. & West, S. A. 2004 Spite and the scale of competition. *J. Evol. Biol.* **17**, 1195–1203. (doi:10.1111/j.1420-9101.2004.00775.x)

Gumert, M. D. & Ho, M. R. 2008 The trade balance of grooming and its coordination of reciprocation and tolerance in Indonesian long-tailed macaques (*Macaca fascicularis*). *Primates* **49**, 176–185. (doi:10.1007/s10329-008-0089-y)

Hamilton, W. D. 1964 The genetical evolution of social behavior. *J. Theoret. Biol.* **7**, 1–52. (doi:10.1016/0022-5193(64)90038-4)

Hamilton, W. D. 1970 Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220. (doi:10.1038/2281218a0)

Hamilton, W. D. 1971 Selection of selfish and altruistic behaviour in some extreme models. In *Man and beast: comparative social behavior* (eds J. F. Eisenberg & W. S. Dillon), pp. 57–91. Washington, USA: Smithsonian Press.

Hammerstein, P. 2003 Why is reciprocity so rare in social animals? A protestant appeal. In *Genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 83–94. Cambridge, USA: MIT Press.

Hart, B. L. & Hart, L. A. 1989 Reciprocal allogrooming in impala, *Aepyceros melampus*. *Anim. Behav.* **44**, 1073–1084. (doi:10.1016/S0003-3472(05)80319-7)

Hauser, M. D. 2006 *Moral minds: how nature designed our sense of right and wrong*. New York, USA: Ecco/Harper Collins.

Hauser, M. D. 2009 The possibility of impossible cultures. *Nature* **460**, 190–196. (doi:10.1038/460190a)

Hauser, M. D., Chen, M. K., Chen, F. & Chuang, E. 2003 Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proc. R. Soc. Lond. B* **270**, 2363–2370. (doi:10.1098/rspb.2003.2509)

Heinsohn, R. & Packer, C. 1995 Complex cooperative strategies in group-territorial African lions. *Science* **269**, 1260–1262. (doi:10.1126/science.7652573)

Hemelrijk, C. K. & Luteijn, M. 1998 Philopatry, male presence and grooming reciprocation among female primates: a comparative perspective. *Behav. Ecol. Sociobiol.* **42**, 207–215. (doi:10.1007/s002650050432)

Henrich, J. *et al.* 2005 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–855.

Henrich, J. *et al.* 2006 Costly punishment across human societies. *Science* **312**, 1767–1770. (doi:10.1126/science.1127333)

Jensen, K., Hare, B., Call, J. & Tomasello, M. 2006 What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proc. R. Soc. B* **273**, 1013–1021. (doi:10.1098/rspb.2005.3417)

Jensen, K., Call, J. & Tomasello, M. 2007a Chimpanzees are rational maximizers in an ultimatum game. *Science* **318**, 107–109. (doi:10.1126/science.1145850)

Jensen, K., Call, J. & Tomasello, M. 2007b Chimpanzees are vengeful but not spiteful. *Proc. Natl Acad. Sci.* **104**, 13046–13050. (doi:10.1073/pnas.0705555104)

Kawai, N. & Matsuzawa, T. 2000 Numerical memory span in a chimpanzee. *Nature* **403**, 39–40. (doi:10.1038/47405)

Kirchsteiger, G. 1994 The role of envy in ultimatum games. *J. Econ. Behav. Org.* **25**, 373–389. (doi:10.1016/0167-2681(94)90106-6)

Krams, I., Krama, T., Igaune, K. & Mand, R. 2008 Experimental evidence of reciprocal altruism in the pied flycatcher. *Behav. Ecol. Sociobiol.* **62**, 599–605. (doi:10.1007/s00265-007-0484-1)

Lakshminaryanan, V., Chen, M. & Santos, L. 2008 Endowment effect in capuchin monkeys. *Phil. Trans. R. Soc. B* **363**, 3837–3844. (doi:10.1098/rstb.2008.0149)

Langergraber, K. E., Mitani, J. C. & Vigilant, L. 2007 The limited impact of kinship on cooperation in wild chimpanzees. *Proc. Natl Acad. Sci. USA* **104**, 7786–7790. (doi:10.1073/pnas.0611449104)

Melis, A., Hare, B. & Tomasello, M. 2006 Chimpanzees recruit the best collaborators. *Science* **311**, 1297–1300. (doi:10.1126/science.1123007)

Melis, A. P., Hare, B. & Tomasello, M. 2008 Do chimpanzees reciprocate received favours? *Anim. Behav.* **76**, 951–962. (doi:10.1016/j.anbehav.2008.05.014)

Mischel, W., Shoda, Y. & Rodriguez, M. L. 1989 Delay of gratification in children. *Science* **244**, 933–938. (doi:10.1126/science.2658056)

Mulcahy, N. & Call, J. 2006 Apes save tools for future use. *Science* **312**, 1038–1040. (doi:10.1126/science.1125456)

Noe, R. 2006 Cooperation experiments: coordination through communication versus acting apart together. *Anim. Behav.* **71**, 1–18. (doi:10.1016/j.anbehav.2005.03.037)

Nowak, M. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)

Parr, L. A. 2003 The discrimination of faces and their emotional content by chimpanzees (*Pan troglodytes*). *Ann. NY Acad. Sci.* **1000**, 56–78. (doi:10.1196/annals.1280.005)

Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. 2007 Planning for the future by western scrub-jays. *Nature* **445**, 919–921. (doi:10.1038/nature05575)

Rochat, M., Serra, E., Fadiga, L. & Gallese, V. 2007 The evolution of social cognition: goal familiarity shapes monkeys' action understanding. *Curr. Biol.* **18**, 227–232. (doi:10.1016/j.cub.2007.12.021)

Rosati, A. G., Stevens, J. R., Hare, B. & Hauser, M. 2007 The evolutionary origins of human patience, temporal preferences in chimpanzees, bonobos and human adults. *Curr. Biol.* **17**, 1663–1668. (doi:10.1016/j.cub.2007.08.033)

Russell, A. F. & Wright, J. 2008 Avian mobbing: byproduce mutualism not reciprocal altrusim. *Trends Ecol. Evol.* **24**, 3–5. (doi:10.1016/j.tree.2008.09.003)

Schino, G. & Aureli, F. 2008 Grooming reciprocation among female primates: a meta-analysis. *Biol. Lett.* **4**, 9–11. (doi:10.1098/rsbl.2007.0506)

Schino, G., di Sorrentino, E. P. & Tiddi, B. 2007 Grooming and coalitions in Japanese macaques (*Macaca fuscata*): partner choice and the time frame reciprocation. *J. Comp. Psychol.* **121**, 181–188. (doi:10.1037/0735-7036.121.2.181)

Seyfarth, R. M. & Cheney, D. L. 1984 Grooming alliances and reciprocal altruism in vervet monkeys. *Nature* **308**, 541–543. (doi:10.1038/308541a0)

Stephens, D. W., McLinn, C. M. & Stevens, J. R. 2002 Discounting and reciprocity in an iterated Prisoner's Dilemma. *Science* **298**, 2216–2218. (doi:10.1126/science.1078498)

Stevens, J. R. & Hauser, M. D. 2004 Why be nice? Psychological constraints on the evolution of cooperation. *Trends Cogn. Sci.* **8**, 60–65. (doi:10.1016/j.tics.2003.12.003)

Sutter, M. 2007 Outcomes versus intentions: on the nature of fair behavior and its development with age. *J. Econ. Psychol.* **28**, 69–78. (doi:10.1016/j.joep.2006.09.001)

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

Warneken, F. & Tomasello, M. 2006 Altruistic helping in human infants and young chimpanzees. *Science* **311**, 1301–1303. (doi:10.1126/science.1121448)

Warneken, F., Hare, B., Melis, A., Hanus, D. & Tomasello, M. 2007 Spontaneous altruism by chimpanzees and young children. *PLoS Biol.* **5**, 5–10.

West, S., Griffin, A. & Gardner, A. 2007 Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432. (doi:10.1111/j.1420-9101.2006.01258.x)

Wheatcroft, D. J. & Krams, I. 2008 Response to Russell and Wright: avian mobbing. *Trends Ecol. Evol.* **24**, 5–6. (doi:10.1016/j.tree.2008.09.002)

Wheatcroft, D. J. & Price, T. D. 2008 Reciprocal cooperation in avian mobbing: playing nice pays. *Trends Ecol. Evol.* **23**, 416–419. (doi:10.1016/j.tree.2008.04.011)

Whitlock, M., Davis, B. & Yeaman, S. 2007 The costs and benefits of resource sharing: reciprocity requires resource heterogeneity. *J. Evol. Biol.* **20**, 1772–1782. (doi:10.1111/j.1420-9101.2007.01387.x)

Wilkinson, G. S. 1984 Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184. (doi:10.1038/308181a0)

Williams, G. C. 1966 *Adaptation and natural selection.* Princeton, NJ: Princeton University Press.

Wilson, E. O. 1975 *Sociobiology: a new synthesis.* Cambridge, MA: Harvard University Press.

Wilson, M. L., Hauser, M. D. & Wrangham, R. W. 2001 Does participation in cooperative intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Anim. Behav.* **61**, 1203–1216. (doi:10.1006/anbe.2000.1706)

Wood, J. N., Glynn, D. & Hauser, M. D. 2008 Rhesus monkeys' understanding of actions and goals. *Social Neurosci.* **3**, 60–68. (doi:10.1080/17470910701563442)