PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2010) **365**, 2635–2650
doi:10.1098/rstb.2010.0146

Review

Punishment and spite, the dark side of cooperation

Keith Jensen[1,2,*]

[1]*School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*
[2]*Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

Causing harm to others would hardly seem to be relevant to cooperation, other than as a barrier to it. However, because selfish individuals will exploit cooperators, functional punishment is an effective mechanism for enforcing cooperation by deterring free-riding. Although functional punishment can shape the social behaviour of others by targeting non-cooperative behaviour, it can also intimidate others into doing almost anything. Second-party functional punishment is a self-serving behaviour at the disposal of dominant individuals who can coerce others into behaving cooperatively, but it need not do so. Third-party and altruistic functional punishment are less likely to be selfishly motivated and would seem more likely to maintain norms of cooperation in large groups. These forms of functional punishment may be an essential part of non-kin cooperation on a scale exhibited only by humans. While punitive sentiments might be the psychological force behind punitive behaviours, spiteful motives might also play an important role. Furthermore, functionally spiteful acts might not be maladaptive; reckoning gains relative to others rather than in absolute terms can lead to hyper-competitiveness, which might also be an important part of human cooperation, rather than just an ugly by-product.

Keywords: punishment; spite; cooperation

> Men are the only animals who devote themselves assiduously to making one another unhappy.
> (H. L. Mencken 1956)

1. INTRODUCTION

The importance of cooperation, and the challenge in trying to explain it, has long been a central focus of evolutionary biology. The contributions in this volume are a hallmark to the longstanding interest in a conundrum raised by Darwin (1859). Most of this work has attempted to explain prosocial behaviour, particularly altruism (or helping), since this is unlikely to evolve given the fitness costs incurred by the actor. The key feature of all of the ultimate (evolutionary) arguments (Mayr 1961; Tinbergen 1963) is that the individual who helps must benefit from doing so, either directly through net fitness gains in the helper's lifetime, or indirectly through other individuals carrying copies of the helper's genes (see Lehmann & Rousset 2010; see also West *et al.* 2007; Clutton-Brock 2009).

Economists and ecologists are typically more interested in functional explanations, namely the reasons behaviours are performed based on their immediate costs and benefits. The key difference between functional and evolutionary approaches is the time scale of the costs and benefits. In the former case, pay-offs are immediate, and in the latter, the pay-offs are accrued as lifetime fitness gains and losses. Psychologists—and recently economists—are interested in another level of analysis, proximate level explanations, namely the immediate incentives for behaviour (Mayr 1961; Tinbergen 1963). Like evolutionary biologists, social scientists also tend to view the individual as selfish, though on a motivational level. Economists have classically modelled human behaviour on assumptions of rational self-interest. Consider Smith's (1776/2007) famous invisible hand, in which every individual acting for his own good produces—as unintended by-products—benefits for others. Psychologists, as well, often regard helpful acts as being selfishly motivated, whether to consciously achieve material outcomes as in 'calculated reciprocity' (Brosnan & de Waal 2002), or unconsciously as a means to achieve psychological benefits such as the 'warm glow' that comes from helping (Andreoni 1990). Concern for the well-being of others is not necessary for prosocial behaviour (though Smith did pay special attention to these moral emotions; Smith 1759/2005). However, for an act to be prosocially motivated, it has to have as its primary goal the benefit to the recipient (Batson 1991). Any benefits to the actor, such as reputation gained, harm avoided or indirect benefits through nepotism must be incidental. Such positive other-regarding (prosocial) concerns must overcome rational, hedonistic, self-interested motivations. At both the proximate

*k.jensen@qmul.ac.uk

One contribution of 14 to a Theme Issue 'Cooperation and deception: from evolution to mechanisms'.

and ultimate levels, selfishness is an obstacle that must be overcome or manipulated for joint social ventures to work.

While most attention has been paid to prosocial acts themselves, and the benefits that must accrue to the helper, antisocial acts are surprisingly important for cooperation. As will be discussed here and elsewhere in this volume (Brosnan *et al.* 2010*b*; Gächter *et al.* 2010; Melis & Semmann 2010), harm, and the threat of it, can be powerful inducements for cooperation. Functionally, punishment—also referred to as negative reciprocity, coercion, harassment and return-benefits spite—is likely to be important for maintaining cooperation. From the ultimate perspective, punishing a non-cooperator is immediately costly for the actor as well as the target, but if the actor receives net fitness benefits as a result, then the punishment is ultimately selfish. From a functional perspective, economists, for instance, note that people will continue to punish even when others benefit and they alone bear the cost. Such 'altruistic' and third-party punishment has garnered recent theoretical interest and has been suggested to be essential to uniquely human cooperation (e.g. Fehr & Fischbacher 2003) or at least very rare in other species (Leimar & Hammerstein 2010; Melis & Semmann 2010). More puzzling phenomena are spiteful acts in which the actor experiences a net fitness loss. However, a Hamiltonian view shows that inclusive fitness makes a costly self-sacrifice beneficial to individuals sharing genes with the actor; spite, then, can evolve because it indirectly works as a form of altruism (Gardner & West 2006; West & Gardner 2010). Spite without inclusive fitness benefits, by definition, cannot evolve; however, spiteful acts might produce relative gains for the actor and therefore be evolutionarily selfish. I will suggest that spiteful competition allows humans to compete on scales not seen in other animals, and that this hyper-competitiveness is as essential to human cooperation as ultrasociality and hyper-cooperativeness (Richerson & Boyd 1998, 2005; Hill *et al.* 2009).

A full understanding of punitive and spiteful behaviours, as well as prosocial acts, will come from an appreciation of the cognitive mechanisms underlying them. The psychological motivations behind punishment are puzzling and difficult to elucidate. Do punishers have as a goal the benefits received by others, namely prosocial preferences? Is the goal to reform subsequent behaviour of the target? Alternatively, is the goal more abstract, such as achieving cooperative norms? Perhaps the motives behind punitive acts are antisocial, having the suffering of the target as the primary goal with any positive effects being unintended by-products. It may be the case that antisocial preferences are unique psychological mechanisms that allow for hyper-competitiveness. Aversion to inequity and other fairness concerns, stemming from a propensity for social comparison, along with sentiments such as *schadenfreude*—pleasure in the misfortunes of others—and motivations to see others suffer losses as goals unto themselves can fuel hyper-competitive behaviour. Whether hyper-competitiveness is a real phenomenon that may be unique to humans remains to be shown, but it appears to be the case that altruism's evil twin might be more than undesired baggage.

2. TERMS

Before discussing punishment, spite, cooperation, altruism, helping and so on, it is important to be clear about the use of the terms. There is considerable disagreement about usage, largely because evolutionary biologists, ecologists, economists, psychologists and the lay public tend to use the same terms, but with subtle differences in connotation. For instance, altruism was coined by Auguste Comte in the nineteenth century and is defined by the Oxford English Dictionary as 'devotion to the welfare of others, regard for others, as a principle of action; opposed to egoism or selfishness'. Spite—defined as '(1) an action arising from, or displaying, hostile or malignant feeling; outrage, injury, harm; insult, reproach; (2) a strong feeling of contempt, hatred or ill-will; intense grudge or desire to injure; rancorous or envious malice'—has an even more venerated history, with a written record dating back to at least the fourteenth century.

Hamilton's (1964) uses of the terms are logical, but narrow. Altruism, to an evolutionary biologist, is an act that is detrimental to the actor's fitness but produces a fitness benefit for another individual. As for spite, he did not use the term 'spiteful behaviour' until 1970, referring initially to costly imposition of fitness costs on others as 'counter-selected' (Hamilton 1964). Some economists, sociologists and psychologists (behaviourists, also called learning theorists, whom are adamantly non-mentalistic) take a functional approach, focusing on the immediate consequences for the actor or recipient. Using the same terms, social scientists would arrive at a similar table (see table 1 in Brosnan *et al.* 2010*b*; Bshary & Bergmüller 2008). A proximate approach tries to determine the mechanisms underlying the behaviour, not just accounting costs and benefits. One such cognitive (mentalistic) proximate approach used the same sort of cost–benefit matrix as Hamilton did, classifying four types of social, fortunes-of-others emotions based on their negative and positive effects (Ortony *et al.* 1988). In this classification, sadness and suffering are negative; happiness and pleasure are positive (table 1). Clearly, there is bound to be confusion over the use of the terms.

The rigorous definitions provided by Hamilton have done much to clarify thinking about the evolution of social behaviour, but his borrowing of commonly used words has contributed to confusion across disciplines. Furthermore, even within biology, the same term can have different meanings, depending on whether one is referring to ultimate causes, phenomenological descriptions, functional explanations or proximate mechanisms. I hope to avoid confusion by using the terms as is standard in their respective disciplines and adding the adjectives appropriate to their specialist usage. I will use the adjective 'evolutionary' to refer to ultimate, fitness-based uses (i.e. evolutionary altruism, evolutionary spite; see also West &

Table 1. Social concern matrix. Adapted from Ortony *et al.* (1988).

	individual B positive feelings	individual B negative feelings
individual A positive feelings	symhedonia (+, +)	schadenfreude (+, −)
individual A negative feelings	jealousy (−, +)	empathy (−, −)

Gardner 2010). For functional explanations, I will preface the terms with 'functional'; economists, as well as ecologists, will usually use the terms in the same way, so I do not distinguish between them. Finally, for proximate mechanisms, I will use 'psychological' as an adjective rather than 'proximate' since there can be proximate explanations that are not psychological (e.g. hormonal and environmental), whereas I will focus on psychological mechanisms.[1] Specifically, I will concentrate on intentions and motivations (see also Hauser *et al.* 2009). This terminology is a departure from the useful approach advocated by the editors of this volume (see §1 and table 1 in Brosnan *et al.* 2010*b*; Bshary & Bergmüller 2008). However, because I move back and forth from evolutionary to functional to psychological levels of explanation, the latter of which is not included in the Bshary and Bergmüller taxonomy, the simple approach I will use will hopefully generate the least amount of confusion.

A final point to consider is the relationship between functional descriptions and ultimate explanations. For any trait to be selected for, it has to confer direct or indirect fitness benefits to the actor (e.g. West *et al.* 2007). Indirect fitness benefits are those that go to individuals carrying copies of the actor's genes; because the actor does not experience the benefits—such as forfeiting reproduction for the benefit of others and imposing fitness costs on others at a personal fitness cost—these can be labelled as evolutionary altruism and evolutionary spite, respectively. On the other hand, any behaviours that result in net fitness benefits for the actor in its lifetime are, in a strict evolutionary sense, selfish. This is true whether the pay-offs are immediate, as in mutualistic interactions and symbioses, or delayed as in direct reciprocity (what Trivers (1971) called 'reciprocal altruism'), indirect reciprocity (e.g. reputation) or negative reciprocity (punishment, sanctions, etc.). However, there is a difference between behaviours that produce immediate pay-offs versus those with temporal delays. In the latter case, pay-offs are not inevitable; there are more opportunities for free-riding, cheating, defecting and so on, all of which generates adaptive challenges distinct from simultaneous pay-offs (Clutton-Brock 2009). Costs paid may not be returned, and this can select for psychological traits such as individual recognition, cheater detection, account keeping, punitive strategies, moral emotions and so on that are not required when pay-offs are immediate (Trivers 1971; Brosnan *et al.* 2010*b*). For this reason, I take a functional approach when describing behaviours.

3. PUNISHMENT

(a) *Functional second-party punishment*

The Oxford English Dictionary's definition of punishment is 'the infliction of a penalty or sanction in retribution for an offence or transgression; (also) that which is inflicted as a penalty; a sanction imposed to ensure the application and enforcement of a law'. A functional definition of punishment used by biologists differs from the standard English usage somewhat by focusing on costs to the punisher as well as the target, and by excluding institutions and norms such as laws. The functional definition is the costly imposition of costs on another individual that result in delayed benefits for the punisher (Clutton-Brock & Parker 1995). There are two important features of this definition. First is that the punisher has to benefit as a result of its actions. For instance, retaliatory aggression that does not produce some future benefit is not adaptive and therefore not likely to evolve. The second feature is that it is costly at the time it is performed; the benefits are delayed. This is to distinguish punishment in the functional sense from harassment, aggression, dominance displays and other behaviours that produce immediate benefits for the actor. As pointed out above, an evolutionary perspective does not distinguish between delayed and immediate benefits, but the distinction is important since the behaviours themselves, the consequences and the psychological causes can be quite different. Functional punishment can be thought of as return-benefits functional spite in the same way that direct reciprocity is considered as return-benefits functional altruism (Trivers 1985); the point is that the actor suffers an immediate cost that, on average, should result in fitness gains. Specifically, the future benefits are social dominance, cheater and parasite deterrence, offspring and sexual partner discipline or coercion, and the enforcement of cooperation (Clutton-Brock & Parker 1995). The last of these is the most relevant for the discussion here. The predominant view of functional punishment is that it is negatively reciprocal—an eye for an eye, a tooth for a tooth (Clutton-Brock & Parker 1995). However, this need not be the case. Aggression, for instance, can be used to maintain dominance regardless of the actions taken by the targets of aggression; random acts of aggression can be very effective in maintaining subordination (Silk 2002). The same can be true for all the forms of functional punishment. Animals can harm others to coerce them into changing their subsequent behaviour so that they gain personal fitness (Clutton-Brock & Parker 1995; Gardner & West 2004*a*). It is a way of shaping the social environment through force or through withholding benefits.

In practice, though, it is difficult to rule out immediate benefits that can arise from acts of aggression and avoidance, making it difficult to distinguish functional punishment from more obviously selfish behaviours such as harassment. An analogy with an inanimate species will highlight this point. A rose thorn causes pain to an animal trying to eat the

flower, and this causes the animal to withdraw. However, the thorn is probably not under selection pressure to cause animals to subsequently avoid that particular flower or roses more generally, but for the immediate benefit of not being eaten. This is the sense in which learning theorists (behaviourists or operant conditioning psychologists) use the term: operant (functional) punishment is any stimulus or removal of a stimulus that contingently decreases the frequency of a behaviour's occurrence (e.g. Seymour *et al.* 2007). Operant punishment, strictly speaking, should be no more efficacious than operant reinforcement in modifying behaviour, though in reality operant punishment can be a more effective learning mechanism (e.g. Yerkes 1907/2005). Similarly, to an economist, functional punishment is an incentive, and it can be more effective than rewards at maintaining cooperation (Andreoni *et al.* 2003). In this sense, rose thorns punish the eating of roses. From an evolutionary perspective, delayed benefits, as well as benefits to others, may only be by-products of immediately selfish strategies (Jensen & Tomasello in press).

There are a few examples of how punishment can function to maintain cooperative behaviour, at least from the perspective of the actor. Coral-reef fish (*Paragobiodon xanthosomus*), for instance, will suppress their own reproduction (social queuing) to avoid eviction by dominants; social stability results from the threat of functional punishment (Wong *et al.* 2007). As another example, reef fish will chase away cleaner fish (*Labroides dimidiatus*) that nibble off the client's mucus rather than the less-preferred ectoparasites; this functional punishment does diminish cheating, as was demonstrated experimentally (Bshary & Grutter 2005). In cooperatively breeding animals like meerkats (*Suricata suricatta*) and superb fairy wrens (*Malurus cyaneus*), dominant breeding pairs coerce their offspring and other group members into forfeiting reproduction to serve as helpers (Mulder & Langmore 1993; Clutton-Brock & Parker 1995). These examples of cooperation maintained by functional punishment demonstrate how behaviour that is harmful to the punisher can be discouraged. Functional punishment benefits the actor and is therefore an evolutionarily selfish strategy exercised by individuals which are in a position to exploit others, such as when dispersal and reproductive options for subordinates are limited.

However, it is surprising to discover that there are many instances in which there is no functional punishment for non-cooperative behaviour, and relatively few examples in which there is. This may be owing to a lack of attention to functional punishment, but there is likely to be even more underreporting of observations of non-events. For example, in cooperative breeders, there is very little evidence that non-cooperative behaviours are punished. Dominant meerkat males will aggress against subordinate males for 'false feeding', namely failing to provide food for pups (Clutton-Brock *et al.* 2005), but there is little evidence for ejection of lazy individuals from groups (Clutton-Brock 2002). Helpers in colonies of naked mole rats (*Heterocephalus glaber*) will continue to help even if dominants are removed (Reeve 1992). Furthermore, 'false feeding', at least in the bell miner (*Manorina melanophrys*), may not be a deceptive behaviour and therefore not a non-cooperative behaviour in need of correction (McDonald *et al.* 2007). Within primates, accounts of functional punishment targeted at non-cooperative behaviours are anecdotal; there is, as yet, no systematic evidence for it. There is one reported observation in captivity of one male chimpanzee (*Pan troglodytes*) attacking another, supposedly for failing to provide support in a conflict (de Waal 1982), and another single observation in the wild of males attacking a younger male, apparently due to his insubordination (Nishida *et al.* 1995). However, in perhaps the only systematic study of reciprocity and aggression in chimpanzees there was no functional punishment of any sort for failure to reciprocate grooming or support (Koyama *et al.* 2006). There is one suggestive example of functional punishment of non-cooperative behaviour in rhesus macaques (*Macaca mulatta*) in which higher ranking individuals attacked lower ranking individuals when they failed to give food calls (Hauser 1992). The suggestion was that dominant individuals were functionally punishing the functionally selfish behaviour of withholding information. While an attractive hypothesis, it failed to rule out a more plausible explanation, namely that conflict over food arose when individuals finding it failed to establish possession by giving food calls, something that was demonstrated in white-faced capuchin monkeys (*Cebus capucinus*; Gros-Louis 2004).

To elucidate whether chimpanzees functionally punish non-cooperative behaviours, an experiment presented captive subjects with three different scenarios, all involving food loss (Jensen *et al.* 2007*a*). In the loss condition, the food was moved away from the subject by the experimenter to an empty, adjacent room; this was a baseline measure of general frustration to losing food. In the unfairness condition (on which more will be said later), the experimenter moved the food towards another chimpanzee who was in that room. Finally, there was a theft condition in which another chimpanzee stole the food away from the subject by pulling a rope—a decidedly non-cooperative behaviour. In all conditions, the subjects could never get the food back, but they could collapse the table, preventing anyone from having it. Chimpanzees reliably collapsed the table more often when it was stolen than in either of the other two conditions. The chimpanzees were vengeful (functionally punitive) in that they retaliated aggressively in the only way possible. That they did so most often in the theft condition suggests that they were sensitive to the harmful behaviour of conspecifics. Consistent with functional punishment (though also consistent with intimidation), dominant individuals were more likely to collapse the table than were subordinates (though subordinates were just as likely to steal food). However, theft increased over time while retaliation decreased, suggesting that in the absence of immediate pay-offs—dominants normally chase off subordinates when food is contested (e.g. Hare *et al.* 2000)—functional punishment failed to enforce cooperative behaviour (see also Jensen & Tomasello in press).

All of the above are examples of second-party (do-it-yourself) functional punishment; the punisher reaps the benefits of changes in the target's behaviour. This appears to be the dominant form of functional punishment in small-scale human societies (Wiessner 2005; Marlowe & Berbesque 2008; Hill *et al.* 2009). Much cooperative human behaviour can probably be explained as a form of correcting the behaviour of someone else for personal, though delayed benefits. However, there is more to human cooperation than 'might makes right'. Norms of cooperation allow people of any rank to use low-cost punishments such as scolding to reign in free-riders. For instance, if someone jumps to the head of a queue, he will be told off, and not just by the person at the head of the queue or the biggest person there. In one amusing anecdote demonstrating the potential costs of functional punishment, a bank robber brandishing a handgun was remonstrated by a customer at the head of the queue and told to wait his turn. Discouraged, the would-be thief left the bank and was later arrested (Bryson 1995). The difference between human queues and something like reproductive 'queuing' in fish is that dominance relationships—coercive cooperation—are not needed. Since functional punishment is costly, such as through retaliation against punishers (Denant-Boemont *et al.* 2007; Janssen & Bushman 2008), it makes little sense to punish if there are no direct benefits. Yet people do this routinely, which brings the discussion to a special form of functional punishment.

(b) *Functional altruistic and third-party punishment*

Second-party functional punishment is not likely to be sufficient to maintain large-scale cooperation simply because individuals in a position of dominance can exploit others, coercing them to work in their favour, and retaliation can make functional punishment too costly. Cooperative outcomes are fortuitous, but not inevitable. As discussed elsewhere in this volume (see Gächter *et al.* 2010), functional punishment is important in maintaining cooperation in humans, perhaps in a way not seen in other animals (Fehr & Fischbacher 2004*a*). One basic reason for this is that humans will punish others for social violations even when they personally stand nothing to gain. One suggestion is that humans have a tendency to behave prosocially and, additionally, are inclined to punish (e.g. Fehr & Gächter 2002). This is referred to as strong reciprocity (Gintis 2000). Because the costs are borne by the individual but the benefits accrue to the group, the functional punishment is called 'altruistic punishment'. Altruistic functional punishment is distinguished from second-party functional punishment in that the former produces group benefits (Fehr & Gächter 2000, 2002; Bowles & Gintis 2003; Boyd *et al.* 2003).

Evidence for functionally altruistic punishment comes from economic experiments such as the public goods game (Fehr & Gächter 2002). In the public goods game, several participants (players) who do not know each other are each given an endowment of money. They can put as much or as little of this endowment into a public pool as they choose. Money in the public pool is increased by some ratio by the experimenter and then divided equally among all the players. The public goods game is effectively an *n*-person Prisoner's Dilemma in which the best collective outcome is for everyone to cooperate, but the best individual strategy is to defect (contribute nothing) while the others contribute maximally. The presence of defectors causes a decline in public contributions over successive trials, even though each individual never plays against the same group of players more than once. However, allowing players to inflict a cost on others by giving up a smaller portion of their endowment has the effect of punishing defecting. As a result, cooperation in the form of giving money to the public pool stabilizes at a high level. The reason that altruistic functional punishment is functionally altruistic is that the punishers pay an additional cost to harm the target, even though they never again interact with the reformed defector and do not gain recognition or any other material benefit, and any benefits go to other anonymous individuals. Functional punishment in these games has therefore been called a second-order public good (Panchanathan & Boyd 2004). A minority of strong reciprocators in a group creates a cooperative 'culture', whereas a functionally punishment-free group loses its members to the more successful sanctioning institution (Gürerk *et al.* 2006). Moreover, people are more likely to functionally punish non-cooperators within their own group than out-group defectors since such functional punishment increases benefits (in terms of reforming free-riders) within the punisher's group (Shinada *et al.* 2004; though see Bernhard *et al.* 2006).

Similarly, third-party functional punishment (what social psychologists mean when they use the term 'punishment') involves a disinterested individual intervening and inflicting costs on violators. This occurs when a judge or a police officer metes out penalties for social violations. Third-party functional punishment has also been demonstrated in economic experiments (e.g. Fehr & Fischbacher 2004*b*). In a third-party punishment experiment, an observer witnesses a transgression such as defection in a Prisoner's Dilemma game played between two other participants. This anonymous observer can give up part of his endowment to inflict a cost on the violator even though he can gain nothing from his actions. Canonical economic models of rational self-interest predict that the observer should give up nothing, but some people will still impose a cost on violations of cooperative norms, a finding that has been replicated in various cultures (e.g. Henrich *et al.* 2005).

There is little, if any, solid evidence for functional altruistic or third-party punishment in non-human animals. The most suggestive evidence comes from studies of policing. Policing occurs when one animal intervenes on behalf of another in a conflict. Ruling out third-party interventions on behalf of kin, there are only a handful of examples in which the intervener appears to be neutral to the outcomes. For instance, chimpanzees (de Waal 1982; de Waal & Luttrell

1988) and monkeys such as bonnet macaques (*Macaca radiata*; Silk 1992) will intervene in conflicts. However, the evidence tends to be indirect, such as the observation that there is an increase in the number of conflicts in groups of monkeys (pigtailed macaques, *Macaca nemestrina*) after the removal of the dominant individuals (Flack *et al*. 2006); however, this may just reflect an increase in conflicts as the sub-dominants jockey for position in the resulting power vacuum. Or it may be the case that the 'punisher' achieves immediate or delayed direct benefits such as reducing the amount of noise in the group, or reduces harm among females in his harem (e.g. Schradin & Lamprecht 2000). Policing in social insects is a special case since the destruction of eggs for the benefit of the remainder of the hive benefits the punishers indirectly through kin benefits (Ratnieks & Wenseleers 2008), a point that will be expanded upon in §4. In one experiment, male cleaner fish aggressed against female partners for 'cheating' by taking the preferred food from a plastic plate, resulting in the immediate removal of the common food source (Raihani *et al*. 2010). As a result, the females were less likely to take the preferred food in subsequent trials. While Raihani *et al*.'s (2010) study was designed to make a point about third-party functional punishment, it was actually a test of second-party functional punishment since there was no third party, and since the punisher benefited directly by altering the behaviour of his partner to his benefit. While clients may benefit in natural settings, this study demonstrated that third-party benefits would be a by-product of a coercive strategy. There is, as yet, no published experimental work on third-party functional punishment in non-human animals, a gap that sorely needs to be filled.

(c) *Psychological punishment*

The previous discussion addressed the function of punishment, which may say something about the adaptive significance of punitive strategies in maintaining cooperation, while at the proximate, psychological level the issue is what motivates one individual to punish another. A behaviour that is motivated for its effect on another individual—not on the actor—is a social motivation (Jensen in press). A social motivation can be influenced by sensitivity to the welfare of others or by sensitivity to the outcomes affecting others (social concern). If an individual faces a conflict between personal outcomes and consequences for others, and it chooses the latter, it is said to have a social (or other-regarding) preference (see also Brosnan 2006; Silk 2009; Jaeggi *et al*. 2010). In all of these cases, the motivations, concerns and preferences can be prosocial (as in positive other-regarding preferences) or antisocial (as in negative other-regarding preferences). As an example, empathy—having the emotions appropriate to the circumstances of others (e.g. Hoffman 1982; Preston & de Waal 2002)—is a prosocial concern and can induce prosocial acts of functionally altruistic behaviour (Batson 1991; see also de Waal & Suchak 2010). It is important to note that prosocial and antisocial outcomes can arise as by-products of social indifference. For example, if one leaves scraps of food on a picnic table when no longer hungry, any benefits to birds, squirrels, mice and other animals in the park are unintended and incidental. Motivations can only be said to be social if they have as their primary goal outcomes affecting others. Indifference is not a social preference.

In the case of punishment, the actions of the punisher have to be motivated for their effect on others (Jensen & Tomasello in press). These preferences can be either prosocial (positively other-regarding; think of a parent telling a child that she is being disciplined for her own good) or antisocial (negatively other-regarding; a desire to see the target of punishment suffer is satisfaction enough). They can also be normative or moral (punishing to maintain cooperation as a social good). However, in all of the examples of second-party functional punishment given above, it is quite probably the case that the goals of the punishers were non-social. The goal is only that the target refrains immediately from its harmful act, or becomes coerced into performing a behaviour congruent with the punisher's goals. The punisher does not need to be motivated by the results of its actions on the welfare of the target or others in the group. Any consequences for the well-being on the punished individual will be by-products. This might even be the case for third-party and altruistic functional punishment. Group beneficial outcomes do not require group beneficial intentions. That is not to say that cooperative behaviour will not result, just that such an outcome need not be the motivating force. Consider again a rose—it does not intend that animals do not eat it, nor does it intend that the animal suffer or learn to refrain from eating it. The rose's thorns produce the result. It does not need to intend outcomes because natural selection has honed the traits that lead to the adaptive outcome. The same can be said for punishment in social insects; attacking a queen from another hive, destroying eggs laid by other workers and so on are relatively invariant responses to biochemical cues (e.g. Monnin *et al*. 2002). While interesting as adaptive behaviours, from a cognitive point of view they are probably not much more interesting than rose thorns.

The flexibility of the behaviours of vertebrates, particularly large-brained species with complex social lives, makes it tempting to explain punishment in cognitively richer terms. Such is the argument of the social brain hypothesis (Jolly 1966; Humphrey 1976; Byrne & Whiten 1988; Dunbar 1998). It is difficult, however, to determine the intentions and motivations of animals. For instance, when fish aggress against harm, such as a client chasing away a cleaner that gleaned more than it should have, it is not clear what cognitive mechanisms are involved. Even though bitten clients will chase cheaters (Tebbich *et al*. 2002), simple learning (operant conditioning) processes could suffice; alternatively, innate mechanisms might also be at work. What is important is that the behaviour be performed flexibly in a variety of contexts. At present, there is not enough information to infer what intentional and motivational systems are involved, and more importantly, whether the behaviours are other-regarding (see Brosnan *et al*. 2010*b*).

Even in humans, which are without question the most behaviourally flexible animals in the world, and which are also the most studied, there is considerable debate about what motivates punitive behaviours. One suggestion is that punitive behaviours, which function to maintain cooperation by deterring free-riders, have a unique psychological mechanism such as a specialized cheater detection module (Cosmides 1989). Furthermore, humans may have a punitive sentiment, an evolved motivational system that imbues the punisher with a desire that the target be harmed (Price *et al.* 2002). This punitive sentiment, what Trivers (1971) and others call 'moral outrage', may be predicated on a belief in a sense of justice (e.g. Charlesworth 1991), of correcting a wrong. Perhaps the simplest form of justice is retributive, inflicting a harm for a harm. It is literally carved in stone: the Code of Hammurabi, from around 1700 BC, dictates 'an eye for an eye'. People will often state that offenders should be punished as a deterrence—a prospective motivation (e.g. Hoffman & Spitzer 1985). However, it is not always clear that this is the case. In practice, people are often retributive—a retrospective motivation—seeking 'just deserts' for perpetrators (e.g. Carlsmith & Darly 2002; Carlsmith 2006). In studies of altruistic functional punishment, it is not clear that people have altruistic motives—others may benefit as a result of changes in the free-riders' behaviours, but these altruistic benefits could be unintended by-products. The act may be antisocial in that it has as its primary motivation that the non-cooperators suffer (Herrmann *et al.* 2008). Psychological punishment in humans, then, can be attuned to the effects it has on others, not just the effect it has for the actor. It has also been suggested that even though the tests are done anonymously with single encounters, people may still act as though they are being observed and gaining a reputation as someone to not be trifled with (e.g. Johnstone & Bshary 2004; Barclay 2006; Kurzban *et al.* 2007). These alternate explanations are difficult to rule out, even in controlled experimental situations. The suggestion here is not that people are always motivated by a sense of moral or normative concern, but that they *can be* motivated in this way.

A further consideration on the topic of psychological punishment is the role of emotions. Contrary to what would be expected from moral philosophy, emotions play an important role in moral judgements (e.g. Frank 1988; Greene & Haidt 2002). People report being angry when punishing others in economic games, and they show concomitant physiological and neurological responses (Pillutla & Murnighan 1996; Fehr & Gächter 2002; de Quervain *et al.* 2004; van't Wout *et al.* 2006). Punishing should feel good, since material benefits would not always be immediately forthcoming. Proximate mechanisms in the forms of immediate motivational rewards are important for mediating punishment and negative other-regard (e.g. de Quervain *et al.* 2004). Similar results were found by Singer *et al.* (2006) in which men experienced increased activation in the reward circuit of the brain when they saw people who had previously cheated against them in a Prisoner's Dilemma game (actually confederates) receive a physically painful stimulus. Men also showed decreased activation in parts of the medial prefrontal cortex associated with empathy when they saw a fair opponent, as opposed to an unfair opponent, in pain. Humans are not the only angry species; anger is basic emotion that probably has deep evolutionary roots (Darwin 1899; Burrows *et al.* 2006; Parr *et al.* 2007). In the punishment experiment described above, chimpanzees also showed signs of anger (displays and tantrums) when food was stolen from them, and anger was correlated with collapsing the food table (Jensen *et al.* 2007*a*). However, although other species have primary emotions, secondary, social emotions such as moral outrage, pride, shame and guilt may be uniquely human (e.g. Fessler & Haley 2003).

Functional punishment, then, is a harm-causing behaviour that provides delayed benefits at some cost to the actor. Because of these costs, it can be used to manipulate the targets into performing behaviours that benefit the actor and, superficially at least, maintain cooperative outcomes such as cooperative breeding. Functional punishment can certainly deter free-riding. The inference is that harm-causing behaviour is adaptive, but it is difficult for any given case to distinguish functional punishment from other aggressive behaviours such as harassment and redirected aggression such as when gulls 'attack' grass after losing a conflict (Lorenz 1966). Experimental work is helpful in this regard. At present, there are very few experimental studies of functional punishment and none on altruistic and third-party functional punishment in non-human animals, a situation that will hopefully be remedied. It will also be important to probe the psychological aspects of functional punishment to determine what it is that motivates the punisher, particularly with regard to the effects on the target. A way forward will be to look at cases of harm-causing behaviour where the only reason for inflicting harm is to see the target suffer.

4. SPITE

(a) *Functional spite*

Functional punishment, because it is costly to the actor at the time it is performed—despite any direct fitness benefits that may result in the future—is sometimes labelled as delayed benefits spite (Trivers 1985; Clutton-Brock & Parker 1995). To evolutionary biologists, this can be discomfiting. Evolutionary spite involves lifetime fitness costs to both actor and target. Since evolutionary spite does not directly help others, and since reciprocity in kind would be harmful, evolutionary spite seems even less likely to evolve than evolutionary altruism. However, evolutionary spite can yield inclusive fitness benefits to the actor through indirect fitness if the individuals harmed are less related to the actor than the average individual in the population or if third parties sharing genes with the actor benefit as a result of the action (Hamilton 1970; Wilson 1975). Evolutionary spite, then, is a form of evolutionary altruism in which the actor suffers a fitness cost to indirectly provide benefits to individuals sharing genes with it by reducing competition from

individuals not sharing those genes (Gardner & West 2004*b*, 2006; West & Gardner 2010). Evolutionary spite is extraordinarily rare in nature. Only embryonic parasitoid wasps (*Copidosoma floridanum*), red fire ants (*Solenopsis invicta*), the bacterium *Wolbachia* and some colonial bacteria (e.g. *Photorhabdus luminescens*) satisfy the strict requirements (Keller *et al.* 1994; Foster *et al.* 2001; Gardner & West 2006).

Functional spite, on the other hand, may be more common. It is true that from an evolutionary perspective, if the actor benefits in any way as a result, functional spite, like functional punishment, is ultimately selfish. However, like functional punishment, functional spite is still a phenomenon that requires explanation. Overly exclusive definitions overlook interesting examples of social behaviour (Gadagkar 1993). For instance, western and herring gulls (*Larus occidentalis* and *Larus argentatus*) were observed to destroy the eggs of rivals if they had lost their own eggs (Pierotti 1980). While there was no net reduction in the actor's fitness (Waltz 1981), the behaviour is consistent with functional spite (Pierotti 1982; Gadagkar 1993) in that the plausible adaptive explanation of the act is to reduce the fitness of rivals. Relative fitness gains come from a decrease—or failure to increase—in a rival's fitness relative to the actor's. As another example, Brereton (1994) suggested that when stumptail macaques (*Macaca artoides*) interfere with copulating pairs, they risk aggression (naturally), but they could benefit in the future by reducing the likelihood of the reproduction of their rivals. Other examples include wasteful feeding by vervet monkeys (*Cercopithecus aethiops*; Horrocks & Hunte 1981), harassment of infants and juveniles in macaques (Trivers 1985), and post-copulatory mate guarding and sexual swelling in cercopithecines (Pagel 1994). However, there are very few published examples of functional spite in the animal behaviour literature, and all of these would need to be carefully scrutinized to rule out immediate gains or delayed direct benefits such as dominance or sexual coercion. Experimental work will be particularly valuable in teasing apart the alternatives.

Unsurprisingly, most experiments have been conducted on humans. The most widely used test that results in functionally spiteful outcomes is the ultimatum game. (This is a test of fairness preferences, a topic that will be discussed in the following section.) In this economic experiment, one player, the proposer, is given a sum of money by the experimenter, and he can share this amount with the second player, the responder. If the responder accepts the offer, both take home their share, and if he rejects it, both get nothing (Güth *et al.* 1982; Camerer 2003). If responders behave in a rational, self-interested way, they should accept any offer because something is better than nothing, and as a result, proposers should make minimal offers. However, this is not what people do; responders routinely reject unfair offers, and as a result, proposers tend to make fair offers. (In the dictator game, in which the second player has no power, first players tend to offer something, but far less than in the ultimatum game; Kahneman *et al.* 1986; Camerer 2003). The threat of harm induces the proposers to behave more cooperatively.

Experimental economic approaches are now being used to probe other-regarding preferences in other animals. One such study allowed chimpanzees to choose between prosocial outcomes and antisocial outcomes (Jensen *et al.* 2006). Chimpanzees could pull a tray with food closer while at the same time causing the other tray to move further away. In one of the experiments, the actor would receive no food for any of her choices, but she could prevent the partner from getting anything (a functionally spiteful outcome) by pulling the opposite table away. If she did nothing, the partner received the food automatically. There was no preference for functionally spiteful (or functionally altruistic) outcomes. Using another approach described earlier, chimpanzees could negatively impact the food outcomes of a partner by collapsing a table (Jensen *et al.* 2007*a*). This is similar in spirit to the money burning game (Zizzo & Oswald 2001). In addition to the theft condition already discussed, there was an unfair outcome condition in which the experimenter moved the food away from the subject and gave it to a conspecific. Chimpanzees were no more likely to collapse the table in this condition than in the loss condition in which no one benefited, nor were they angrier, suggesting that they were not spitefully motivated. In another study, chimpanzees were presented with a reduced form of the ultimatum game called the mini-ultimatum game (Jensen *et al.* 2007*b*). In the mini-ultimatum game, proposers are given a choice of two outcomes, one of which is always unfair and typically rejected, in four different games with differing degrees of unfairness between the options (Falk *et al.* 2003). Proposer payoffs are shown before the slash, and the amount for the responder is after the slash; for example, 8/2 indicates that 80 per cent of the reward goes to the proposer while 20 per cent goes to the responder. Adults in the Falk *et al.* (2003) study responded by rejecting the unfair (8/2) option most often when they could have been offered the fair (5/5) outcome by the proposer. There were fewer rejections when the proposer was faced with a generous option (2/8). Responders sometimes rejected 8/2, though less often, when the proposer had no choice (8/2 versus 8/2), presumably because they were sensitive to the outcome disparity. Some even rejected 8/2 when the alternative was 10/0 (nothing for them), possibly out of malice. Chimpanzees, however, showed no such sensitivity. Regardless of what options the proposer faced, responders never rejected any non-zero offer, though they would reject offers of zero. Chimpanzee behaviour was consistent with the standard economic model of rational self-interest. They were not willing to pay a cost to see another individual suffer a greater cost.

What distinguishes functional spite from functional punishment is that functional spite does not require any change in the target's subsequent behaviour. The end goal is the harm incurred by the target. There may be indirect benefits—otherwise the behaviour would not be functional—but these are less tangible than for functional punishment. Whereas functional punishment emphasizes the delayed benefits to the punisher, functional spite emphasizes the immediate costs to the target; negative consequences for the

target are the *raisons d'être* for spiteful acts. Functional punishment is a means to an end; functional spite is an end in itself. The benefits that accrue to the actor would therefore be indirect; the target's loss is the actor's gain. Here, losses and gains are not evaluated in absolute terms as with functional punishment, but in relative terms; the actor need not benefit directly, but the target has to suffer a greater relative cost. For instance, with cooperative breeding, functional punishment requires that the punisher succeed in coercing others to forfeit reproduction so that the punisher gains reproductive help, whereas in functional spite the purpose of the harmful act is to have the target reproduce less. This can indirectly benefit the actor by resulting in less competition for the actor's offspring or for the actor itself. As for spiteful acts in humans, since much of the evidence comes from studies addressing the motivations, these will be discussed in the next section.

(b) *Psychological spite: negative social preferences*

Functional spite may be indirectly selfish in that the actor benefits through the harm suffered by the target. The motivation to harm others may not be selfish, however, and any tangible benefits to the actor may be unintended. As discussed previously, an act that is motivated for its social effect is a social motivation, and the motivation is revealed through preferences for these social effects over personal outcomes. Negative, or antisocial, preferences will be motivated by concerns for the negative well-being of others (Jensen in press). Causing harm for harm's sake is a spiteful motivation, and it can be underpinned by a comparison of oneself to others. Again, indifference is not a social preference. If an individual acting for its own selfish ends causes unintended harm to others, then this is not an antisocial preference. There is no ulterior motive in psychological spite: the suffering of others is not the means to an end, but is an end in itself.

A key facet of negative social concern is the fact that individuals evaluate themselves relative to others. Social comparisons typically are done for one's abilities and opinions relative to those of others (Festinger 1954). Positive evaluations, which can improve self-esteem, come from downward social comparison, that is comparing oneself to others worse off. Doing so makes one's own situation seem better in comparison. Negative evaluations from upward social comparison can be more complicated. On one hand, if the individual identifies himself with the comparison group, the evaluations can be positive. On the other hand, they can diminish one's self-esteem by seeing that others are better off. For instance, it may feel good to buy a new, state-of-the-art television, particularly if one's co-worker's model is not as nice, but the good feeling will go away if the neighbour buys a better one; yet, if the neighbour's television stops working, positive feelings will return. Comparing one's own gains to others causes some individuals to make personally harmful decisions so that they are not worse off relative to others (though they end up worse off in absolute terms; Saijo & Nakamura 1995). Feelings such as jealousy, envy, *schadenfreude*, gloating and other such misanthropic sentiments are fortunes-of-others emotions (Ortony *et al.* 1988), and these may be tuned to social comparison. All of these sentiments can be regarded as spiteful in that they are driven by a regard for the misfortunes—the negative welfare—of others.

Economists also note that people compare themselves to others with the emphasis on material outcomes such as wealth, namely that they are sensitive to fairness, particularly disadvantageous inequity. According to the simplest accounts of fairness sensitivity, people attend not only to their own losses and gains, but compare these to the losses and gains of others (Loewenstein *et al.* 1989; Fehr & Schmidt 1999; Bolton & Ockenfels 2000). An aversion to disadvantageous inequity—having less than others—motivates people to correct an unfair situation. While outcome-based theories are simpler than psychological attempts to model sensitivity to fairness, they do not fully account for making or rejecting unfair offers in economic experiments (e.g. Forsythe *et al.* 1994; Blount 1995). The suggestion, then, is that people are sensitive to unfair intent (Rabin 1993; Levine 1998; Dufwenberg & Kirchsteiger 2004). It is quite likely the case that both outcomes and intent influence sensitivity to fairness (Falk & Fischbacher 2006). While the exact nature of how people are influenced by unfairness is unresolved, the proposal is that other-regarding preferences are the underlying motivation behind altruistic punishment and strong reciprocity (Fehr & Fischbacher 2003, 2005). The specifics of what constitutes unfairness vary because cultures have different norms or rules of behaviour (Henrich *et al.* 2005). What is consistent is this: people have other-regarding preferences (Andreoni 1990; Fehr & Camerer 2007).

The ultimatum game, described above, is a useful tool for probing social preferences, particularly sensitivity to fairness. Rejections of unfair offers are irrational from a purely self-regarding perspective, but people respond emotionally, angrily rejecting unfair offers (Pillutla & Murnighan 1996; Sanfey *et al.* 2003); the fairness sensitivity is not cool and calculated. While they appear to be more sensitive to the intentions of the proposer, for instance by not rejecting unfavourable outcomes if the choices were not determined by the proposer (Blount 1995), they still reject unfavourable outcomes even when the proposer could not have done differently (Falk *et al.* 2003) and they will destroy the wealth of others in a money burning game in which the unfair outcomes have nothing to do with the intentions of the target (Zizzo & Oswald 2001). The intuitive interpretation of responder rejections in the ultimatum game is that people functionally punish others out of a sense of fairness, even though this makes them worse off in absolute terms than if they accept any offer. However, because people reject offers when they are generous (Herrmann *et al.* 2008), or when the proposer had no unfair intent (Falk *et al.* 2003)—and since all studies are one-shot games—it seems that fairness motives are not the only factor influencing rejections.

People appear to be vindictive, namely they are willing to pay a cost to inflict harm for the sake of having the proposer suffer a loss (Fehr & Fischbacher 2005; Fehr *et al.* 2008). This effect does not only occur when getting less than a fair share. The motivations behind these harmful acts can be called 'do-gooder derogation', dominance, revenge, malice, competition, payoff maximization and so on; they are all negatively other-regarding preferences. Ultimatum rejections are spiteful in that the immediate motivation is that the targets suffer (Fehr *et al.* 2008; Herrmann *et al.* 2008). In other words, the intuitive interpretation may not be correct. The harm inflicted, if it is not intended to change the target's behaviour, is not psychological punishment. If there is no ulterior motive, then the motive is psychological spite.

There is considerable debate about whether non-human animals compare outcomes with others and therefore show a sensitivity to disadvantageous inequity. In the studies described above in which subjects could control and respond to outcomes, there did not appear to be any comparison of gains and losses relative to others (Jensen *et al.* 2006, 2007*a*,*b*). A paradigm that is widely used has subjects react to differential outcomes without being able to control them as a demonstration of inequity aversion. In these tests, subjects receive a lower quality food reward while the partner receives a better quality reward, either contingent on effort—typically trading an object with the experimenter—or not (Brosnan & de Waal 2003). Brown capuchin monkeys (*Cebus apella*) were first shown to be averse to inequity (Brosnan & de Waal 2003), but results with capuchin monkeys, great apes, cotton-top tamarins (*Saguinus oedipus*) and common marmosets (*Callithrix jacchus*), as well as dogs (*Canis familiaris*; Range *et al.* 2009) and New Zealand rabbits (Heidary *et al.* 2008) have been mixed (for reviews, see Brosnan 2006; Silk 2009; Brosnan *et al.* 2010; Jensen in press; see also de Waal & Suchak 2010). However, rejecting unfair offers when doing so has no effect on others does not decrease inequity but actually increases it (Henrich 2004). This is certainly not a rational thing to do, and people playing the impunity game, in which rejecting has no effect on proposers, tend not to reject unfair offers (Bolton & Zwick 1995; Hachiga *et al.* 2009) though some may do so as a signal of emotional commitment (Yamagishi *et al.* 2009). At present, the results for social comparison in non-human animals are inconsistent. Inequity aversion, if it is exhibited in other animals, does not appear to be robust. It also does not seem to translate into functionally spiteful actions. While it is not possible to draw strong conclusions on social comparisons yet, it does seem that humans are much more spitefully motivated than are other animals. If this indeed is the case, the obvious question is, how can the most prosocial species on the planet also be the most antisocial?

(c) *Hyper-competition: the adaptive value of functional and psychological spite*

Much has been made of the fact that humans cooperate on a large scale with non-kin and engage in coordinated activities involving a division of labour (e.g. Fehr & Fischbacher 2003; Richerson & Boyd 2005; Tomasello *et al.* 2005; Hill *et al.* 2009). Prosocial motivations such as empathy are likely to be fundamental to prosocial acts directed towards strangers (Batson 1991). Negative sentiments such as psychological punishment and sensitivity to unfairness are also likely to play an important role because they can impel people to punish free-riders. However, functional punishment can maintain any behaviour, not just cooperation (Boyd & Richerson 1992). For instance, people will ostracise others who fail to conform to norms of dress, worship or any other arbitrary behaviour. Functional punishment may be an important component of large-scale cooperation because groups with functional punishers—particularly altruistic or third-party functional punishers—are more successful than those with only functional altruists (which become exploited by free-riders) or only non-cooperators (Boyd *et al.* 2003; Panchanathan & Boyd 2004; Gürerk *et al.* 2006; Hauert *et al.* 2007; De Silva *et al.* 2010). Functional altruistic punishment, combined with social learning mechanisms, notably imitation, constitute cultural group selection (e.g. Fehr & Fischbacher 2003; Richerson *et al.* 2003; Mesoudi *et al.* 2004; Richerson & Boyd 2005), which may explain why humans—which are the only species with cumulative culture (e.g. Tomasello *et al.* 2005; Herrmann *et al.* 2007)—are able to overcome the free-rider problem in large groups. On the other hand, there are arguments against cultural group selection and the experimental evidence used to support it (e.g. Burnham & Johnson 2005; Hagen & Hammerstein 2006; West *et al.* 2007, 2008). It is beyond the scope of this paper to evaluate the merits of cultural group selection, but the insight I want to draw on here is that functional punishment—particularly when the punisher does not benefit directly—may be necessary for non-kin cooperation in large groups. And if altruistic and third-party functional punishment are shown to be unique to humans—a matter that requires investigation—they will help explain uniquely human cooperation.

Large-scale non-kin cooperation of the kind exhibited by humans has been described as ultrasocial and hyper-cooperative (Richerson & Boyd 1998, 2005; Hill *et al.* 2009). But human social behaviour is hardly always positive. We exploit the environment—and each other—in ways that no other species do (Vitousek *et al.* 1997). Our cooperative behaviours are often directed towards group members while outgroup members are derogated, all of which can take as little as random assignment to a group in a camp or a t-shirt colour (Sherif *et al.* 1961; Turner *et al.* 1979). According to cultural group selection, competition between groups is the selective pressure that allows for the success of groups with cooperators (e.g. Sober & Wilson 1998; Richerson & Boyd 2005). While humans do form large groups, every group is made of sub-groups, which in turn are composed of sub-sub-groups. For instance, the UK can be thought of as a group, and will act as such in a war, but there will be numerous groups within that

such as Liverpudlians versus Mancunians, 'postal code gangs' within Manchester, gang members who wear low-riders and those who wear baggy trousers, baggy-trouser wearers who drink Newcastle ale and those who prefer Guinness stout and so on. Just as one can form a group from random individuals, take any two individuals and you have two groups. In a similar vein, Freud (1961) coined the phrase 'narcissism of small differences'. People will compare themselves to others, looking for distinguishing differences. As well, they will compare their losses and gains relative to others, and these social comparisons can lead to negative feelings. As a result, people will inflict costs on others, not only for violations of cooperative norms, not only for levelling differences in wealth, but to see that others do not fare better. Gains and losses are not reckoned in absolute terms, but relatively.

Such obsessive social comparison suggests that humans are hyper-competitive. As an example, consider a queue in a coffee shop. Normally, waiting in line is a cooperative activity in which norm violators (queue jumpers) might be punished, or at least given the evil eye. But what if the stakes are raised? For instance, imagine that a special deal is announced in which the first five customers will get as much free coffee as they want, even if this means depleting the shop's supply. You are eighth in a long line, and you know that self-regarding (selfish) individuals will take everything, leaving nothing for the rest. In such a competitive situation, you have several options. You can simply leave and go to another coffee shop (scramble competition). You can bully your way to the front of the line and hope you are stronger and more determined than the others (contest competition). Or you can release a stink bomb that you just happen to be carrying, scattering everyone and contaminating the coffee so that no one—not even you—will get any (spiteful competition). The first two are well known in behavioural ecology (Nicholson 1954; Maynard Smith 1982) and contribute to social problems such as the tragedy of the commons (Hardin 1968). Spiteful competition is not a term used in behavioural ecology, possibly because it does not exist outside of humans (though there may be a few exceptions such as egg destruction, food waste and reproductive interference, described earlier).

Antisocial motives would not seem to be intuitively adaptive. They would seem to be correlated by-products of prosocial motives; having positive social concerns is adaptive for cooperation, and the underlying mechanisms happen to spill over, resulting in negative social concerns. However, negative social concerns give people the ability to assess their outcomes in relative, rather than just absolute terms (a generic mobile phone is nice, but not as nice as your friend's latest iPhone). As a result, we flexibly adjust our cooperativeness and competitiveness to the size of the group; people will cooperate when competition is more global, compete when it is more local (West *et al.* 2006; see also Gardner & West 2004*a*,*b*). It is hard to imagine another species in which individuals flexibly adjust their competition and cooperation depending on the size of the group and the presence of other groups, compete for the spirit of competition, gauge success in relative terms, savour the failure of others and use these negative social concerns to seek the downfall of rivals as seen in parochialism, tribalism, war and so on (e.g. Darwin 1871; Hamilton 1975; West *et al.* 2006; Choi & Bowles 2007). In short, it is hard to imagine another species that is hyper-competitive. Taking pleasure in the misfortunes of others provides the immediate motivational reward for the delayed and relatively intangible benefits of relative gains to be reaped. Negative social concerns are essential elements of hyper-competitiveness, just as positive social concerns are likely to be essential to human hyper-cooperativeness and ultrasociality (Richerson & Boyd 1998, 2005; Hill *et al.* 2009).

Self-serving, second-party functional punishment that typifies vengeance and retaliatory aggression likely evolved first. Second-party functional punishment is not uncommon in the animal kingdom; the only thing that sets it apart from simple acts of aggression is the delay in benefits. This is not likely to be such a large step from immediately beneficial behaviours such as harassment, dominance and aggression, though these will entail some cognitive demands such as individual recognition. Altruistic and third-party functional punishment are more cognitively demanding. They will require concern for the welfare and suffering of others, and probably also an awareness of social norms, rules for how one ought to and ought not to behave. However, whether human altruistic and third-party functional punishment may be due, in part, to psychological spite rather than psychological punishment is unresolved. The selection pressure for altruistic and third-party functional punishment of non-cooperative behaviours might have required cultural group selection (Richerson & Boyd 2005), or kin selection writ large (West *et al.* 2007, 2008). Whatever the selective pressure, functional punishment of violations of cooperative norms may have only evolved once, and this is a question that begs an answer. Functional spite might lie between second-party and third-party functional punishment, having evolved after the former and before the latter (see Hauser *et al.* 2009 for an alternative scenario). Cognitively, functional spite would seem to require psychological spite, an ability to assess one's gains and losses in relative terms and to seek other's losses as primary goals; it would build upon basic emotions such as anger to produce socially evaluative emotions such as jealousy and schadenfreude. Whether such concerns are exhibited by other animals is a matter of active research and debate (e.g. Brosnan 2006; Silk 2009; Jensen in press). The question is an important one. If functional spite lies on the path between second-party and third-party functional punishment, tracking its evolution will illuminate human hyper-competitiveness, and in turn suggest something about our hyper-cooperativeness. The dark side of human nature may not only be a shadow of the light side, but may be integral to the foundation of large-scale cooperation.

I would like to thank Sarah Brosnan, Redouan Bshary, Stuart West and one anonymous reviewer for their helpful comments.

ENDNOTE

[1]For the sake of simplicity, I will overlook the fact that there are also different psychological levels of analysis (Seed *et al.* 2009).

REFERENCES

Andreoni, J. 1990 Impure altruism and donations to public goods: a theory of warm-glow giving? *Econ. J.* **700**, 464–477.

Andreoni, J., Harbaugh, W. & Vesterlund, L. 2003 The carrot or the stick: rewards, punishments, and cooperation. *Am. Econ. Rev.* **93**, 893–902. (doi:10.1257/000282803322157142)

Barclay, P. 2006 Reputational benefits for altruistic punishment. *Evol. Hum. Behav.* **27**, 325–344. (doi:10.1016/j.evolhumbehav.2006.01.003)

Batson, C. D. 1991 *The altruism question: toward a social-psychological answer.* Hillsdale, NJ: Lawrence Erlbaum.

Bernhard, H., Fischbacher, U. & Fehr, E. 2006 Parochial altruism in humans. *Nature* **442**, 912–915. (doi:10.1038/nature04981)

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

Bolton, G. E. & Ockenfels, A. 2000 ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **90**, 166–193.

Bolton, G. E. & Zwick, R. 1995 Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* **10**, 95–121. (doi:10.1006/game.1995.1026)

Bowles, S. & Gintis, H. 2003 Origins of human cooperation. In *Genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 429–443. Cambridge, MA: The MIT Press.

Boyd, R. & Richerson, P. J. 1992 Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **13**, 171–195. (doi:10.1016/0162-3095(92)90032-Y)

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

Brereton, A. R. 1994 Return-benefit spite hypothesis: an explanation for sexual interference in stumptail macaques. *Primates* **35**, 123–136. (doi:10.1007/BF02382049)

Brosnan, S. F. 2006 Nonhuman species' reactions to inequity and their implications for fairness. *Soc. Justice Res.* **19**, 153–185. (doi:10.1007/s11211-006-0002-z)

Brosnan, S. F. & de Waal, F. B. M. 2002 A proximate perspective on reciprocal altruism. *Hum. Nat.* **13**, 129–152. (doi:10.1007/s12110-002-1017-2)

Brosnan, S. F. & de Waal, F. B. M. 2003 Monkeys reject unequal pay. *Nature* **425**, 297–299. (doi:10.1038/nature01963)

Brosnan, S. F., Talbot, C., Ahlgren, M., Lambeth, S. P. & Schapiro, S. J. 2010*a* Mechanisms underlying responses to inequitable outcomes in chimpanzees, *Pan troglodytes*. *Anim. Behav.* **79**, 1229–1237. (doi:10.1016/j.anbehav.2010.02.019)

Brosnan, S. F., Salwiczek, L. & Bshary, R. 2010*b* The interplay of cognition and cooperation. *Phil. Trans. R. Soc. B* **365**, 2699–2710. (doi:10.1098/rstb.2010.0154)

Bryson, B. 1995 *Notes from a small island.* Toronto, ON: McClelland and Stewart.

Bshary, R. & Bergmüller, R. 2008 Distinguishing four fundamental approaches to the evolution of helping. *J. Evol. Biol.* **21**, 405–420. (doi:10.1111/j.1420-9101.2007.01482.x)

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

Burnham, T. C. & Johnson, D. D. P. 2005 The biological and evolutionary logic of human cooperation. *Analyse Kritik* **27**, 113–135.

Burrows, A. M., Waller, B. M., Parr, L. A. & Bonar, C. J. 2006 Muscles of facial expression in the chimpanzee (*Pan troglodytes*): descriptive, comparative and phylogenetic contexts. *J. Anat.* **208**, 153–167. (doi:10.1111/j.1469-7580.2006.00523.x)

Byrne, R. W. & Whiten, A. 1988 *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans.* Oxford, UK: Clarendon Press.

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

Carlsmith, K. M. 2006 The roles of retribution and utility in determining punishment. *J. Exp. Soc. Psychol.* **42**, 437–451. (doi:10.1016/j.jesp.2005.06.007)

Carlsmith, K. M. & Darley, J. M. 2002 Why do we punish? Deterrence and just desert motives for punishment. *J. Pers. Soc. Psychol.* **83**, 284–299. (doi:10.1037/0022-3514.83.2.284)

Charlesworth, W. R. 1991 The development of the sense of justice: moral development, resources, and emotions. *Am. Behav. Sci.* **34**, 350–370. (doi:10.1177/0002764291034003006)

Choi, J.-K. & Bowles, S. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640. (doi:10.1126/science.1144237)

Clutton-Brock, T. H. 2002 Breeding together: kin selection and mutualism in cooperative vertebrates. *Science* **296**, 69–72. (doi:10.1126/science.296.5565.69)

Clutton-Brock, T. H. 2009 Cooperation between non-kin in animal societies. *Nature* **462**, 51–57. (doi:10.1038/nature08366)

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

Clutton-Brock, T. H., Russell, A. F., Sharpe, L. L. & Jordan, N. R. 2005 'False feeding' and aggression in meerkat societies. *Anim. Behav.* **69**, 1273–1284.

Cosmides, L. 1989 The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* **31**, 187–276. (doi:10.1016/0010-0277(89)90023-1)

Darwin, C. 1859 *On the origin of species by means of natural selection.* London, UK: John Murray.

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

Darwin, C. 1899 *The expression of the emotions in man and animals.* New York, NY: Appleton.

Denant-Boemont, L., Masclet, D. & Noussair, C. 2007 Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Econ. Theory* **33**, 145–167. (doi:10.1007/s00199-007-0212-0)

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

De Silva, H., Hauert, C., Traulsen, A. & Sigmund, K. 2010 Freedom, enforcement, and the social dilemma of strong altruism. *J. Evol. Econ.* **20**, 203–217. (doi:10.1007/s00191-009-0162-8)

de Waal, F. B. M. 1982 *Chimpanzee politics: power and sex among the apes*. New York, NY: Johns Hopkins University Press.

de Waal, F. B. M. & Luttrell, L. M. 1988 Mechanisms of social reciprocity in three primate species: symmetrical relationship characteristics or cognition? *Ethol. Sociobiol.* **9**, 101–118.

de Waal, F. B. M. & Suchak, M. 2010 Prosocial primates: selfish and unselfish motivations. *Phil. Trans. R. Soc. B* **365**, 2711–2722. (doi:10.1098/rstb.2010.0119)

Dufwenberg, M. & Kirchsteiger, G. 2004 A theory of sequential reciprocity. *Games Econ. Behav.* **47**, 268–298. (doi:10.1016/j.geb.2003.06.003)

Dunbar, R. I. M. 1998 The social brain hypothesis. *Evol. Anthropol.* **6**, 178–190. (doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8)

Falk, A. & Fischbacher, U. 2006 A theory of reciprocity. *Games Econ. Behav.* **54**, 293–315. (doi:10.1016/j.geb.2005.03.001)

Falk, A., Fehr, E. & Fischbacher, U. 2003 On the nature of fair behavior. *Econ. Inq.* **41**, 20–26. (doi:10.1093/ei/41.1.20)

Fehr, E. & Camerer, C. F. 2007 Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11**, 419–427. (doi:10.1016/j.tics.2007.09.002)

Fehr, E. & Fischbacher, U. 2003 The nature of human altruism. *Nature* **425**, 785–791. (doi:10.1038/nature02043)

Fehr, E. & Fischbacher, U. 2004*a* Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 187–190.

Fehr, E. & Fischbacher, U. 2004*b* Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87. (doi:10.1016/S1090-5138(04)00005-4)

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

Fehr, E. & Gächter, S. 2000 Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994.

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

Fehr, E. & Schmidt, K. M. 1999 A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868. (doi:10.1162/003355399556151)

Fehr, E., Hoff, K. & Kshetramade, M. 2008 Spite and development. *Am. Econ. Rev.* **98**, 494–499. (doi:10.1257/aer.98.2.494)

Fessler, D. M. T. & Haley, K. J. 2003 The strategy of affect: emotions in human cooperation. In *Genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 7–36. Cambridge, MA: MIT Press.

Festinger, L. 1954 A theory of social comparison processes. *Hum. Relat.* **7**, 117–140. (doi:10.1177/001872675400700202)

Flack, J. C., Girvan, M., de Waal, F. B. M. & Krakauer, D. C. 2006 Policing stabilizes construction of social niches in primates. *Nature* **439**, 426–429. (doi:10.1038/nature04326)

Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. 1994 Fairness in simple bargaining experiments. *Games Econ. Behav.* **6**, 347–369. (doi:10.1006/game.1994.1021)

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

Frank, R. H. 1988 *Passions within reason: the strategic role of the emotions*. New York, NY: Norton.

Freud, S. 1961 *Civilization and its discontents* (ed. J. Strachey, transl.). New York, NY: W. W. Norton.

Gächter, S., Herrmann, B. & Thöni, C. 2010 Culture and cooperation. *Phil. Trans. R. Soc. B* **365**, 2651–2661. (doi:10.1098/rstb.2010.0135)

Gadagkar, R. 1993 Can animals be spiteful? *Trends Ecol. Evol.* **8**, 232–234. (doi:10.1016/0169-5347(93)90196-V)

Gardner, A. & West, S. A. 2004*a* Cooperation and punishment, especially in humans. *Am. Nat.* **164**, 753–764.

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

Gardner, A. & West, S. A. 2006 Spite. *Curr. Biol.* **16**, R662–R664. (doi:10.1016/j.cub.2006.08.015)

Gintis, H. 2000 Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179. (doi:10.1006/jtbi.2000.2111)

Greene, J. & Haidt, J. 2002 How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523. (doi:10.1016/S1364-6613(02)02011-9)

Gros-Louis, J. 2004 The function of food-associated calls in white-faced capuchin monkeys, *Cebus capucinus*, from the perspective of the signaller. *Anim. Behav.* **67**, 431–440. (doi:10.1016/j.anbehav.2003.04.009)

Gürerk, O., Irlenbusch, B. & Rockenbach, B. 2006 The competitive advantage of sanctioning institutions. *Science* **312**, 108–111. (doi:10.1126/science.1123633)

Güth, W., Schmittberger, R. & Schwarze, B. 1982 An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388. (doi:10.1016/0167-2681(82)90011-7)

Hachiga, Y., Silberberg, A., Parker, S. & Sakagami, T. 2009 Humans (*Homo sapiens*) fail to show an inequity effect in an 'up-linkage' analog of the monkey inequity test. *Anim. Cogn.* **12**, 359–367. (doi:10.1007/s10071-008-0195-7)

Hagen, E. H. & Hammerstein, P. 2006 Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theor. Popul. Biol.* **69**, 339–348. (doi:10.1016/j.tpb.2005.09.005)

Hamilton, W. D. 1964 The genetical evolution of social behaviour. I & II. *J. Theor. 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. 1975 Innate social aptitudes of man: an approach from evolutionary genetics. In *Biosocial anthropology* (ed. R. Fox), pp. 133–153. London, UK: Malaby Press.

Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243–1248. (doi:10.1126/science.162.3859.1243)

Hare, B., Call, J., Agnetta, B. & Tomasello, M. 2000 Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* **59**, 771–785. (doi:10.1006/anbe.1999.1377)

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A. & Sigmund, K. 2007 Via freedom to coercion: the emergence of costly punishment. *Science* **316**, 1905–1907. (doi:10.1126/science.1141588)

Hauser, M. D. 1992 Costs of deception: cheaters are punished in rhesus monkeys (*Macaca mulatta*). *Proc. Natl Acad. Sci. USA* **89**, 12 137–12 139. (doi:10.1073/pnas.89.24.12137)

Hauser, M., McAuliffe, K. & Blake, P. R. 2009 Evolving the ingredients for reciprocity and spite. *Phil. Trans. R. Soc. B* **364**, 3255–3266. (doi:10.1098/rstb.2009.0116)

Heidary, F., Mahdavi, M. R. V., Momeni, F., Minaii, B., Rogani, M., Fallah, N., Heidary, R. & Gharebaghi, R. 2008 Food inequality negatively impacts cardiac health in rabbits. *PLoS ONE* **3**, 1–3.

Henrich, J. 2004 Inequity aversion in capuchins? *Nature* **428**, 139. (doi:10.1038/428139a)

Henrich, J. *et al.* 2005 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–815.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B. & Tomasello, M. 2007 Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**, 1360–1366. (doi:10.1126/science.1146282)

Herrmann, B., Thöni, C. & Gächter, S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)

Hill, K., Barton, M. & Hurtado, A. M. 2009 The emergence of human uniqueness: characters underlying behavioral modernity. *Evol. Anthropol.* **18**, 187–200. (doi:10.1002/evan.20224)

Hoffman, M. L. 1982 Development of prosocial motivation: empathy and guilt. In *The development of prosocial behavior* (ed. N. Eisenberg), pp. 281–338. New York, NY: Academic Press.

Hoffman, E. & Spitzer, M. L. 1985 Entitlements, rights, and fairness: an experimental examination of subjects' concepts of distributive justice. *J. Legal Stud.* **14**, 259–297. (doi:10.1086/467773)

Horrocks, J. & Hunte, W. 1981 'Spite': a constraint on optimal foraging in the vervet monkey *Cercopithecus aethiops sabaues* in Barbados. *Am. Zool.* **21**, 939.

Humphrey, N. K. 1976 The social function of the intellect. In *Growing points in ethology* (eds P. P. G. Bateson & R. A. Hinde). Cambridge, UK: Cambridge University Press.

Jaeggi, A. V., Burkart, J. M. & Van Schaik, C. P. 2010 On the psychology of cooperation in humans and other primates: combining the natural history and experimental evidence of prosociality. *Phil. Trans. R. Soc. B* **365**, 2723–2735. (doi:10.1098/rstb.2010.0118)

Janssen, M. & Bushman, C. 2008 Evolution of cooperation and altruistic punishment when retaliation is possible. *J. Theor. Biol.* **254**, 541–545. (doi:10.1016/j.jtbi.2008.06.017)

Jensen, K. In press. Primate social concern. In *The evolution of primate societies* (eds J. Mitani, J. Call, P. Kappeler, R. Palombit & J. Silk).

Jensen, K. & Tomasello, M. In press. Punishment. In *Encyclopedia of animal behavior* (ed. S. J. Wood). Oxford, UK: Elsevier.

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. 2007*a* Chimpanzees are vengeful but not spiteful. *Proc. Natl Acad. Sci. USA Am.* **104**, 13 046–13 050. (doi:10.1073/pnas.0705555104)

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

Johnstone, R. A. & Bshary, R. 2004 Evolution of spite through indirect reciprocity. *Proc. R. Soc. Lond. B* **271**, 1917–1922. (doi:10.1098/rspb.2003.2581)

Jolly, A. 1966 Lemur social behavior and primate intelligence. *Science* **153**, 501–506. (doi:10.1126/science.153.3735.501)

Kahneman, D., Knetsch, J. L. & Thaler, R. 1986 Fairness as a constraint on profit seeking: entitlements in the market. *Am. Econ. Rev.* **76**, 728–741.

Keller, L., Milinski, M., Frishknecht, M., Perrin, N., Richner, H. & Tripet, F. 1994 Spiteful animals still to be discovered. *Trends Ecol. Evol.* **9**, 103. (doi:10.1016/0169-5347(94)90205-4)

Koyama, N. F., Caws, C. & Aureli, F. 2006 Interchange of grooming and agonistic support in chimpanzees. *Int. J. Primatol.* **27**, 1293–1309. (doi:10.1007/s10764-006-9074-8)

Kurzban, R., DeScioli, P. & O'Brien, E. 2007 Audience effects on moralistic punishment. *Evol. Hum. Behav.* **25**, 63–87.

Lehmann, L. & Rousset, F. 2010 How life history and demography promote or inhibit the evolution of helping behaviours. *Phil. Trans. R. Soc. B* **365**, 2599–2617. (doi:10.1098/rstb.2010.0138)

Leimar, O. & Hammerstein, P. 2010 Cooperation for direct fitness benefits. *Phil. Trans. R. Soc. B* **365**, 2619–2626. (doi:10.1098/rstb.2010.0116)

Levine, D. K. 1998 Modeling altruism and spitefulness in experiments. *Rev. Econ. Dyn.* **1**, 593–622. (doi:10.1006/redy.1998.0023)

Loewenstein, G. F., Thompson, L. & Bazerman, M. H. 1989 Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* **57**, 426–441. (doi:10.1037/0022-3514.57.3.426)

Lorenz, K. 1966 *On aggression.* London, UK: Methuen.

Marlowe, F. W. & Berbesque, J. C. 2008 More 'altruistic' punishment in larger societies. *Proc. R. Soc. B* **275**, 587–590. (doi:10.1098/rspb.2007.1517)

Maynard Smith, J. 1982 *Evolution and the theory of games.* Cambridge, UK: Cambridge University Press.

Mayr, E. 1961 Cause and effect in biology. *Science* **134**, 1501–1506. (doi:10.1126/science.134.3489.1501)

McDonald, P. G., Kazem, A. J. N. & Wright, J. 2007 A critical analysis of 'false-feeding' behavior in a cooperatively breeding bird: disturbance effects, satiated nestlings or deception? *Behav. Ecol. Sociobiol.* **61**, 1623–1635. (doi:10.1007/s00265-007-0394-2)

Melis, A. P. & Semmann, D. 2010 How is human cooperation different? *Phil. Trans. R. Soc. B* **365**, 2663–2674. (doi:10.1098/rstb.2010.0157)

Mencken, H. L. 1956 *Minority report.* New York, NY: Knopf. (Reprinted by John Hopkins University Press 1997.)

Mesoudi, A., Whiten, A. & Laland, K. N. 2004 Is human cultural evolution Darwinian? Evidence reviewed from the perspective of the Origin of Species. *Evolution* **58**, 1–11.

Monnin, T., Ratnieks, F. L. W., Jones, G. R. & Beard, R. 2002 Pretender punishment induced by chemical signalling in a queenless ant. *Nature* **419**, 61–65. (doi:10.1038/nature00932)

Mulder, R. A. & Langmore, N. E. 1993 Dominant males punish helpers for temporary defection in superb fairy-wrens. *Anim. Behav.* **45**, 830–833. (doi:10.1006/anbe.1993.1100)

Nicholson, A. J. 1954 An outline of the dynamics of animal populations. *Aust. J. Zool.* **2**, 9–65. (doi:10.1071/ZO9540009)

Nishida, T., Hosaka, K., Nakamura, M. & Hamai, M. 1995 A within-group gang attack on a young adult male chimpanzee: ostracism of an ill-mannered member? *Primates* **36**, 207–211. (doi:10.1007/BF02381346)

Ortony, A., Clore, G. L. & Collins, A. 1988 *The cognitive structure of emotions.* Cambridge, UK: Cambridge University Press.

Pagel, M. 1994 The evolution of conspicuous oestrous advertisement in Old World monkeys. *Anim. Behav.* **47**, 1333–1341. (doi:10.1006/anbe.1994.1181)

Panchanathan, K. & Boyd, R. 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499–502. (doi:10.1038/nature02978)

Parr, L. A., Waller, B. M., Vick, S. J. & Bard, K. A. 2007 Classifying chimpanzee facial expressions using muscle action. *Emotion* **7**, 172–181. (doi:10.1037/1528-3542.7.1.172)

Pierotti, R. 1980 Spite and altruism in gulls. *Am. Nat.* **115**, 290–300.

Pierotti, R. 1982 Spite, altruism, and semantics: a reply to Waltz. *Am. Nat.* **119**, 116–120. (doi:10.1086/283895)

Pillutla, M. M. & Murnighan, J. K. 1996 Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organ. Behav. Hum. Decis. Process.* **68**, 208–224. (doi:10.1006/obhd.1996.0100)

Preston, S. D. & de Waal, F. B. M. 2002 Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* **25**, 1–20.

Price, M. E., Cosmides, L. & Tooby, J. 2002 Punitive sentiment as an anti-free rider psychological device. *Evol. Human Behav.* **23**, 203–231. (doi:10.1016/S1090-5138(01)00093-9)

Rabin, M. 1993 Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **83**, 1281–1302.

Raihani, N. J., Grutter, A. S. & Bshary, R. 2010 Punishers benefit from third-party punishment in fish. *Science* **327**, 171. (doi:10.1126/science.1183068)

Range, F., Horn, L., Viranyi, Z. & Huber, L. 2009 The absence of reward induces inequity aversion in dogs. *Proc. Natl Acad. Sci. USA* **106**, 340–345. (doi:10.1073/pnas.0810957105)

Ratnieks, F. L. W. & Wenseleers, T. 2008 Altruism in insect societies and beyond: voluntary or enforced. *Trends Ecol. Evol.* **23**, 45–52. (doi:10.1016/j.tree.2007.09.013)

Reeve, H. K. 1992 Queen activation of lazy workers in colonies of the eusocial naked mole-rat. *Nature* **358**, 147–149. (doi:10.1038/358147a0)

Richerson, P. J. & Boyd, R. 1998 The evolution of human ultra-sociality. In *Ideology, warfare, and indoctrinability; evolutionary perspectives* (eds I. Eibl-Eibisfeldt & F. Salter), pp. 71–95. London, UK: Berghahn Books.

Richerson, P. J. & Boyd, R. 2005 *Not by genes alone: how culture transformed human evolution.* Chicago, IL: University of Chicago Press.

Richerson, P. J., Boyd, R. & Henrich, J. 2003 Cultural evolution of human cooperation. In *The genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 357–388. Cambridge, MA: MIT Press.

Saijo, T. & Nakamura, H. 1995 The 'spite' dilemma in voluntary contribution mechanism experiments. *J. Confl. Resolut.* **39**, 535–560. (doi:10.1177/0022002795039003007)

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. 2003 The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755–1758. (doi:10.1126/science.1082976)

Schradin, C. & Lamprecht, J. 2000 Female-biased immigration and male peace-keeping in groups of the shell-dwelling cichlid fish *Neolamprologus multifasciatus*. *Behav. Ecol. Sociobiol.* **48**, 236–242. (doi:10.1007/s002650000228)

Seed, A. M., Emery, N. & Clayton, N. 2009 Intelligence in corvids and apes: a case of convergent evolution? *Ethology* **115**, 401–420. (doi:10.1111/j.1439-0310.2009.01644.x)

Seymour, B., Singer, T. & Dolan, R. 2007 The neurobiology of punishment. *Nat. Rev. Neurosci.* **8**, 300–311. (doi:10.1038/nrn2119)

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R. & Sherif, C. W. 1961 *Intergroup conflict and cooperation: the robbers cave experiment.* Norman, OK: University of Oklahoma.

Shinada, M., Yamagishi, T. & Ohmura, Y. 2004 False friends are worse than bitter enemies: 'altruistic' punishment of in-group members. *Evol. Hum. Behav.* **25**, 379–393. (doi:10.1016/j.evolhumbehav.2004.08.001)

Silk, J. B. 1992 The patterning of intervention among male bonnet macaques: reciprocity, revenge, and loyalty. *Curr. Anthropol.* **33**, 318–324. (doi:10.1086/204073)

Silk, J. B. 2002 Practice random acts of aggression and senseless acts of intimidation: the logic of status contests in social groups. *Evol. Anthropol.* **11**, 221–225. (doi:10.1002/evan.10038)

Silk, J. B. 2009 Social preferences in primates. In *Neuroeconomics: decision making and the brain* (eds P. W. Glimcher, C. F. Camerer, E. Fehr & R. A. Poldrack), pp. 269–284. London, UK: Academic Press.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. 2006 Empathic neural responses are modulated by the perceived fairness of others. *Nature* **439**, 466–469. (doi:10.1038/nature04271)

Smith, A. 1759/2005 *The theory of moral sentiments* (ed. S. M. Soares). See http://metalibri.wikidot.com/title:theory-of-moral-sentiments:smith-a.

Smith, A. 1776/2007 *An inquiry into the nature and causes of the wealth of nations* (ed. S. M. Soares). See http://metalibri.wikidot.com/title:an-inquiry-into-the-nature-and-causes-of-the-wealth-of-nations:smith-a.

Sober, E. & Wilson, D. S. 1998 *Unto others: the evolution and psychology of unselfish behavior.* Cambridge, MA: Harvard University Press.

Tebbich, S., Bshary, R. & Grutter, A. S. 2002 Cleaner fish *Labroides dimidiatus* recognise familiar clients. *Anim. Cogn.* **5**, 139–145.

Tinbergen, N. 1963 On aims and methods of ethology. *Z. Tierpsychol.* **20**, 410–433.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. 2005 Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**, 675–735.

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

Trivers, R. 1985 *Social evolution.* Menlo Park, CA: Benjamin-Cummings.

Turner, J. C., Brown, R. J. & Tajfel, H. 1979 Social comparison and group interest in ingroup favouritism. *Eur. J. Soc. Psychol.* **9**, 187–204. (doi:10.1002/ejsp.2420090207)

van't Wout, M., Kahn, R., Sanfey, A. & Aleman, A. 2006 Affective state and decision-making in the Ultimatum Game. *Exp. Brain Res.* **169**, 564–568.

Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. 1997 Human domination of earth's ecosystems. *Science* **277**, 494–499. (doi:10.1126/science.277.5325.494)

Waltz, E. C. 1981 Reciprocal altruism and spite in gulls: a comment. *Am. Nat.* **118**, 588–592.

West, S. A. & Gardner, A. 2010 Altruism, spite, and greenbeards. *Science* **327**, 1341–1344. (doi:10.1126/science.1178332)

West, S. A., Gardner, A., Shuker, D. M., Reynolds, T., Burton-Chellow, M., Sykes, E. M., Guinnee, M. A. & Griffin, A. S. 2006 Cooperation and the scale of competition in humans. *Curr. Biol.* **16**, 1103–1106. (doi:10.1016/j.cub.2006.03.069)

West, S. A., Griffin, A. S. & 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)

West, S. A., Griffin, A. S. & Gardner, A. 2008 Social semantics: how useful has group selection been? *J. Evol. Biol.* **21**, 374–385.

Wiessner, P. 2005 Norm enforcement among the Ju'/hoansi bushmen, a case for strong reciprocity? *Hum. Nat.* **16**, 115–145. (doi:10.1007/s12110-005-1000-9)

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

Wong, M. Y. L., Buston, P. M., Munday, P. L. & Jones, G. P. 2007 The threat of punishment enforces peaceful cooperation and stabilizes queues in a coral-reef fish. *Proc. R. Soc. B* **266**, 1865–1870.

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S. & Cook, K. S. 2009 The private rejection of unfair offers and emotional commitment. *Proc. Natl Acad. Sci. USA* **106**, 11 520–11 523. (doi:10.1073/pnas.0900636106)

Yerkes, R. M. 1907/2005 *The dancing mouse: a study in animal behavior*. Project Gutenburg. See http://www.gutenberg.org/etext/8729. (doi:10.1037/10935-000)

Zizzo, D. J. & Oswald, A. J. 2001 Are people willing to pay to reduce others' incomes? *Ann. Econ. Stat.* **63–64**, 39–65.