

3 Cooperation as a classic problem in behavioural biology



MICHAEL E. PRICE

CHAPTER OUTLINE

WHY HAS COOPERATION BEEN SUCH A BIOLOGICAL PUZZLE? 74

INDIVIDUAL-LEVEL SOLUTIONS TO THE PUZZLE: SELFISH REPLICATORS, COOPERATIVE VEHICLES 76

COOPERATION VIA GENIC SELF-FAVOURITISM (KIN SELECTION AND GREENBEARD ALTRUISM) 77

COOPERATION VIA RETURN BENEFITS (RECIPROCAL ALTRUISM, INDIRECT RECIPROCITY AND COSTLY SIGNALLING) 82

SUMMARY OF INDIVIDUAL-LEVEL THEORIES OF COOPERATION 86

GROUP SELECTION 87

COMPLEX HUMAN COOPERATION: COLLECTIVE ACTION 91

CONCLUSION 98

ACKNOWLEDGEMENTS 100

REFERENCES 100

Cooperative behaviour is, and has always been, a centrally important aspect of human sociality: cooperation in economic exchange, and in teams, groups and organisations of all kinds (religious, political, foraging, military and so on) is a defining characteristic of our species. Cooperative behaviour has also been a centrally important problem in behavioural biology for decades, both as a challenge that has stimulated major scientific advances and as a puzzle that has generated great confusion. Because of the importance of cooperation both in human social life and as a scientific puzzle, the evolution of human cooperation has received decades of intense attention from scholars representing diverse fields, including biology, psychology and all of the social sciences. In this chapter, I will review the history of the most important attempts that have been made to explain cooperation, examine the current state of our ability to explain it and make some suggestions about which future roads we should take in order to continue to enhance our understanding of this important topic.

WHY HAS COOPERATION BEEN SUCH A BIOLOGICAL PUZZLE?

Cooperation has been a puzzle for biologists primarily because it often involves altruism. The traditional definition of altruism in biology has been an action that lowers the actor's fitness while enhancing the fitness of some other individual (Hamilton, 1964; Maynard Smith, 1964; Williams and Williams, 1957). As Tooby and Cosmides (1996) note, this definition is problematic in that it does not give sufficient consideration of altruism's adaptive functionality, that is of the extent to which the altruistic behaviour was designed by natural selection for the purpose of delivering benefits to another individual. For example, a moth that has flown into a spider's web will benefit the spider's fitness at the expense of its own, but this 'altruism' will have been due to successful predation by the spider, rather than to some functional mechanism in moths that evolved in order to deliver benefits to spiders. Therefore, a more appropriate definition of altruism would be behaviour that has been designed by selection to benefit someone else's fitness at the expense of one's own (Tooby and Cosmides, 1996).

Cooperation can be broadly defined as two or more biological entities acting together to accomplish some common goal. The biological world is rife with cooperative behaviour. For example, a mother and the offspring she is nursing both have an interest in making sure the offspring gets fed. When you buy a pack of gum at a corner store, both you and the shopkeeper have the goal of carrying out an exchange transaction. In multicellular species, different specialised cellular types act in concert to help each other, because they all serve the larger goal of promoting the survival and

reproduction of the whole organism. The members of a honeybee colony all strive to benefit the colony in which they live, and will sacrifice their own lives in defence of it. Wolves hunt in groups in order to kill a deer that they will consume together. Note that all of these examples involve altruism, because they all involve actors that incur some kind of fitness cost – for instance an expenditure of time and energy, an exhibition of restraint in the pursuit of one’s own selfish interests or suicidal self-sacrifice – in the course of acting in a way that enhances the fitness of beneficiaries.

Besides being central to cooperation, altruism is also one of the most historically problematic issues in evolutionary biology. Altruism has achieved this status because it seems to challenge the idea that the primary engine of adaptation by natural selection, as specified by Darwin (1859/1958), is reproductive competition among individuals. Shouldn’t all genes governing behaviour focus on enabling their individual carrier to out-compete rivals in the struggle to survive and reproduce, as opposed to altruistically benefiting these rivals? Shouldn’t any gene that benefited these rivals, at the expense of the carrier, be promptly eliminated by selection? These are the basic questions that have driven biologists’ intense efforts to understand altruistic cooperation over the past several decades. Darwin (1859/1958, p. 190) himself regarded such cooperation as a potential threat to his theory: ‘If it could be proved that any part of the structure of any one species had been formed for the exclusive good of another species, it would annihilate my theory, for such could not have been produced through natural selection.’ Over a century later, E. O. Wilson (1975, p. 3) still regarded altruism as ‘the central theoretical problem of sociobiology’.

Table 3.1 helps define altruism in the context of other kinds of social action. In this table, ‘social action’ refers to any behaviour that has a negative or positive impact on the fitness of another individual (the ‘recipient’). Two of these behaviours, selfishness and spite, are non-cooperative. As selfishness involves a fitness gain for the actor, it does not present any special challenge for Darwin’s theory; spite, on the other hand, presents more of a challenge (Hamilton, 1970). Spite, as a costly action that harms a recipient but that provides no benefit to the actor, should not usually be favoured by selection; accordingly, spite appears to be rare in nature (Foster, Wenseleers and Ratnieks, 2001). The other two behaviours in Table 3.1, altruism and mutualism, can be considered kinds of cooperation. Altruism, as an action that harms the actor while benefiting the recipient, presents the puzzle that is the main subject of this chapter. Mutualism, as a mutually beneficial action, is much less puzzling. However, in reality, mutualism and altruism may be more difficult to distinguish than Table 3.1 suggests. For example, ‘reciprocal altruism’ (Trivers, 1971) can involve a protracted process of mutually beneficial exchange. A reciprocally altruistic actor may deliver a benefit to a recipient, and then much later receive a benefit from that recipient in return. In the long term, this interaction will look like ‘mutualism’ according to Table 3.1, but in the short term – before the actor has received the return benefit – it will resemble ‘altruism’. The differences between reciprocal altruism and mutualism will be discussed in more detail later in the chapter, but this topic is raised here simply to make the point that Table 3.1 to some extent oversimplifies a more complex reality.

Table 3.1. Social action classified by effects on fitness of actor and recipient.

		Effect on Recipient	
		+	–
Effect on Actor	+	Mutualism	Selfishness
	–	Altruism	Spite

Each of the four social actions described in this table can have effects that are positive (+) or negative (–) for the respective fitnesses of actor and recipient.

Source: Based on Pizzari Foster (2008).

INDIVIDUAL-LEVEL SOLUTIONS TO THE PUZZLE: SELFISH REPLICATORS, COOPERATIVE VEHICLES

As will be discussed later in this chapter, there are still important unresolved issues related to the topic of how altruistic cooperation evolves. Still, we can safely say that a great deal of progress towards solving this problem has been made since Darwin's era. Most of this progress occurred after biologists began switching their focus from the reproducing individual to the replicating gene. No theorist was more influential than Hamilton (1963, 1964) in promoting the realisation that the gene should be regarded as fundamental, both as the unit of selection and as the ultimate unit of analysis in evolutionary and behavioural biology. From a post-Hamiltonian perspective, a general solution to the puzzle of cooperation suggests itself: genes promote cooperation as a means of promoting their own replication. But how, exactly, can genes enable their own replication by building cooperative individuals?

Genes replicate by building individuals who are strongly motivated to help the genes make copies of themselves, that is individuals who act to promote their own inclusive fitness (or, technically, who act in ways that were inclusive fitness-promoting in the ancestral past). In other words, genes are 'replicators', and individual bodies are 'vehicles' they build in order to enable themselves to replicate (Dawkins, 1976). As fundamental vehicles of selection, individuals are endowed by genes with many adaptations which enable them to behave in fitness-promoting ways. Therefore, when researchers attempt to understand how cooperation evolved, they typically examine the ways in which individuals are adapted for cooperation. This focus on individual-level adaptation has always been a prominent feature of modern evolutionary

biology; Darwin's theory of adaptation by natural selection, for example, was primarily a theory about individual adaptation.

However, research on cooperation also sometimes focuses on vehicles of selection other than the individual, especially the social group. In many species, individuals cooperate in groups, and groups can be considered important vehicles of selection because they are organisations that genes produce as a means of promoting their own replication. If individuals and groups are both important vehicles of selection, however, at which level of selection should we be searching for evidence of adaptation? This issue often comes up in the context of discussions about group selection and multilevel selection, and these will be discussed later in the chapter. First, however, we will review theories that focus on the individual as the primary vehicle of selection. These theories address the puzzle of cooperation in two general ways: by suggesting either that cooperative genes benefit replicas of themselves that are located in other individuals or else that cooperative individuals somehow reap benefits in return for their cooperation.

COOPERATION VIA GENIC SELF-FAVOURITISM (KIN SELECTION AND GREENBEARD ALTRUISM)

Kin selection

The essential insight that Hamilton (1963, 1964) developed is that a gene can succeed in replicative competition not only by promoting the reproduction of its carrier but also by promoting the reproduction of individuals who carry exact copies of itself. For example, if exact copies of Gene A are carried by Individual 1 and Individual 2, then the copy of Gene A located in Individual 1 can create exact replicas of itself by promoting the reproduction of either Individual 1 or Individual 2. The main problem that the gene must overcome in order to engage in such self-favouritism is that of how it can 'know' whether a replica of itself is likely to reside in another individual. One basic and relatively reliable route to such knowledge is to determine the likelihood that the individual is a genetic relative, because the closer two individuals are related genetically, the more likely they will be to share genes that are identical by descent. If Individual 1 and Individual 2 are full siblings, for instance, then the probability that they will share two alleles by virtue of common descent is $\frac{1}{2}$ (because they each must have inherited the allele from one parent or the other). This probability is known as the 'coefficient of relatedness', and it decreases rapidly as genetic distance grows. For example, this coefficient is $\frac{1}{4}$ for half-siblings, for aunts/uncles and nieces/nephews,

and for grandparents and grandchildren, and it drops off to 1/8 for first cousins and to 1/32 for second cousins.

Hamilton incorporated this coefficient into a simple inequality, now known as 'Hamilton's rule', which states that altruism will be likely to evolve when $rB > C$, where r is the coefficient of relatedness, B is the benefit to the recipient, and C is the cost to the altruist. For example, for an act of altruism between two full siblings ($r = 1/2$) to be favoured by selection, it must benefit the recipient more than twice as much as it costs the altruist. Hamilton's rule predicts that altruism will be more likely to evolve when the two actors are more closely related, and when the benefits to the recipient are great and the costs to the altruist are low. Several new terms entered the biologist's lexicon following the publication of Hamilton's theory. If a gene causes an act of altruism to benefit a copy of itself that is located in another body, then the altruist was now said to have acted to increase his or her 'inclusive fitness'. Thus, an act which was detrimental to fitness in the sense that Darwin used the term (i.e. 'classical fitness') could be enhancing to inclusive fitness. The process that governs the evolution of inclusive fitness-enhancing behaviours among genetic relatives became known as 'kin selection'.

Hamilton was not the first to appreciate the link between genetic kinship and altruism (e.g. see Williams and Williams, 1957). However, his generalisation and formalisation of this idea led to him becoming regarded as the father of kin selection. As a theory that is very simple but which nonetheless explains a large portion of behaviour in a huge range of species, kin selection has been a major scientific triumph. It provides a fundamental explanation for the high levels of altruism that are routinely observed between close kin in a vast variety of species, including humans and other primates (e.g. Chagnon, 1979; Lieberman, Tooby and Cosmides, 2007; Silk, 2005). Further, kin selection helps resolve some long-standing puzzles about cooperative behaviour, in particular that of eusocial species – extraordinarily cooperative species characterised by features such as division of labour and cooperative care of young. The insect order Hymenoptera, which includes ants, bees and wasps, includes many eusocial species. Hymenoptera colonies include sterile worker and soldier castes, which perplexed Darwin (1859/1958), since the existence of sterile castes seemed to undermine his theory of adaptation by natural selection: how could sterile individuals pass on adaptations to offspring? As it turns out, the Hymenoptera order is characterised by a breeding system known as 'haplodiploidy', whereby females are more related to their sisters ($r = 3/4$) than to their own offspring ($r = 1/2$); thus, these females' genes can replicate more effectively via altruism towards sisters than via altruism towards offspring. Although haplodiploidy cannot be the whole explanation for eusociality (Wilson and Wilson, 2007), it does help resolve the sterility problem which so puzzled Darwin.

Although kin selection is one of the most important elements of modern evolutionary theory, important aspects of it are often misunderstood or underappreciated. First, although the theory is powerful, one should appreciate its limitations. It is especially important to note that coefficient r refers specifically to the likelihood of genes for altruism that are identical by descent, and cannot be regarded as an

abstract measure of overall ‘genetic similarity’. Kin selection theory does not predict, for instance, that individuals from the same ethnic group should be more altruistic to one another than should ethnically different individuals (for related debates, see Rushton, 1989 and associated commentaries, e.g. Tooby and Cosmides, 1989). In fact, kin selection should act only weakly on all but the very closest genetic relationships. For example, for kin altruism to evolve between first cousins (who by many standards would be considered close genetic relatives), the required B/C ratio (at least eight to one) will often be prohibitively high.

Another important yet sometimes unappreciated feature of kin selection is that related genes in different bodies cannot directly ‘sense’ each other’s presence. Instead, kin altruism occurs as the result of some environmental regularity that enables altruistic behaviour to correlate positively with genetic kinship. For species with nervous systems, this regularity often comes in the form of some kind of information that can be cognitively processed and that reliably indicates the probability that another individual is a close relative. Such a kin detection mechanism may enable one to recognise kin based on, for instance, familiarity (as in ground squirrels and guppies), or on phenotypic matching via odour (as in paper wasps and wood frogs) (Hain and Neff, 2007; Sherman, Reeve and Pfennig, 1997). In humans, kin recognition cues include perinatal association with one’s own mother and enduring co-residence in the same household (Lieberman, Tooby and Cosmides, 2007; Westermarck 1891/1921). As Lieberman *et al.* note, such mechanisms are important not just for directing kin altruism but also for enabling incest avoidance. Because these mechanisms require highly specific kinds of informational inputs in order to operate effectively, as opposed to being ‘general purpose’ kin detectors, they fail in the absence of such inputs. Thus, siblings who are not exposed to kinship cues (e.g. if they are raised in different households) should be less likely to perceive each other as siblings: they should feel less altruism and more sexual attraction towards one another, even if they consciously ‘know’ (i.e. they are verbally informed) that they are siblings. By the same token, non-siblings who are raised in the same household should regard each other with more altruism and less sexual attraction.

The broadening of r : Greenbeard altruism

The general principle underlying kin selection is one of genic self-favouritism: a gene can promote its own replication by benefiting replicas of itself that are located in other individuals. Note that there is nothing in this principle that stipulates that these other individuals must be close genetic kin in general. Any gene that enabled its carrier to recognise other carriers of the same gene, and to direct altruism towards them, could in principle evolve, regardless of whether the carriers were closely related at other loci. So while r in Hamilton’s rule was initially regarded mainly as the probability of individuals sharing genes via common descent, its definition eventually broadened into the probability of individuals sharing the same genes for altruism, irrespective of general kinship (Hamilton, 1975).

In the absence of kinship, the problem of recognising other carriers of the same specific gene generally becomes more formidable. Kinship provides a relatively convenient solution to this problem, because kin relationships tend to be characterised by environmental regularities (e.g. spatial proximity of siblings) to which mechanisms for delivering benefits to kin can become adapted. This correlation between genetic kinship and environmental structure is a major reason for why kin altruism is by far the most commonly observed type of genic self-favouritism. Nevertheless, genic self-favouritism in the absence of general kinship may occur. This type of cooperation is often called ‘greenbeard altruism’, because of a colourful thought experiment proposed by Dawkins (1976) in his widely read popularisation of Hamilton’s theories. Dawkins asks readers to imagine a gene which endows carriers with a green beard (any conspicuous phenotypic label of carrierhood would do; Dawkins just happened to choose this one) and caused them to direct altruism towards other greenbeards. Because carriers would be so recognisable, it would be relatively easy to direct altruism towards them, and altruism among greenbeards could therefore evolve.

However, although greenbeard altruism seems plausible in theory, in reality it is much less common than kin altruism. A likely reason this is true (in addition to the reason mentioned above, i.e. the fact that genetic kinship is relatively likely to be correlated with aspects of environmental structure) is that labels of altruistic disposition are generally easier to fake than are indicators of genetic kinship. (However, even kinship can be faked, for example cuckoo chicks pass themselves off as kin to their hosts.) In a system of greenbeard altruism, the biggest winners will be those who deceptively display the green beard without actually engaging in altruism towards other greenbeards. Mutations that cause their carriers to display the green beard but to engage in a reduced level of altruism towards other carriers would come to dominate the population; these selfish greenbeards would gain all the benefits of the system while paying reduced costs, and would eventually exploit the altruistic greenbeards to extinction (Figure 3.1 illustrates that just as altruistic greenbeards are advantaged over non-greenbeards, selfish greenbeards are advantaged over altruistic greenbeards). However, despite the fact that greenbeard systems will often be vulnerable to deception problems, such systems have been reported in species such as fire ants (Keller and Ross, 1998), slime mould (Queller *et al.*, 2003), side-blotched lizards (Sinervo *et al.*, 2006) and yeast (Smukalla *et al.*, 2008), as well as in human maternal–foetal interactions (Summers and Crespi, 2005).

The further broadening of r

Genic self-favouritism enables cooperation to evolve because it allows genes for cooperation to assist copies of themselves. However, these genes cannot interact directly, but only via the behaviour they encode. Thus, genic self-favouritism is possible only because cooperative genotypes encode cooperative phenotypes, and ultimately it is the individual’s cooperative behaviour itself that delivers fitness benefits to other cooperative individuals. Therefore, in order for cooperation to evolve, it is not necessary

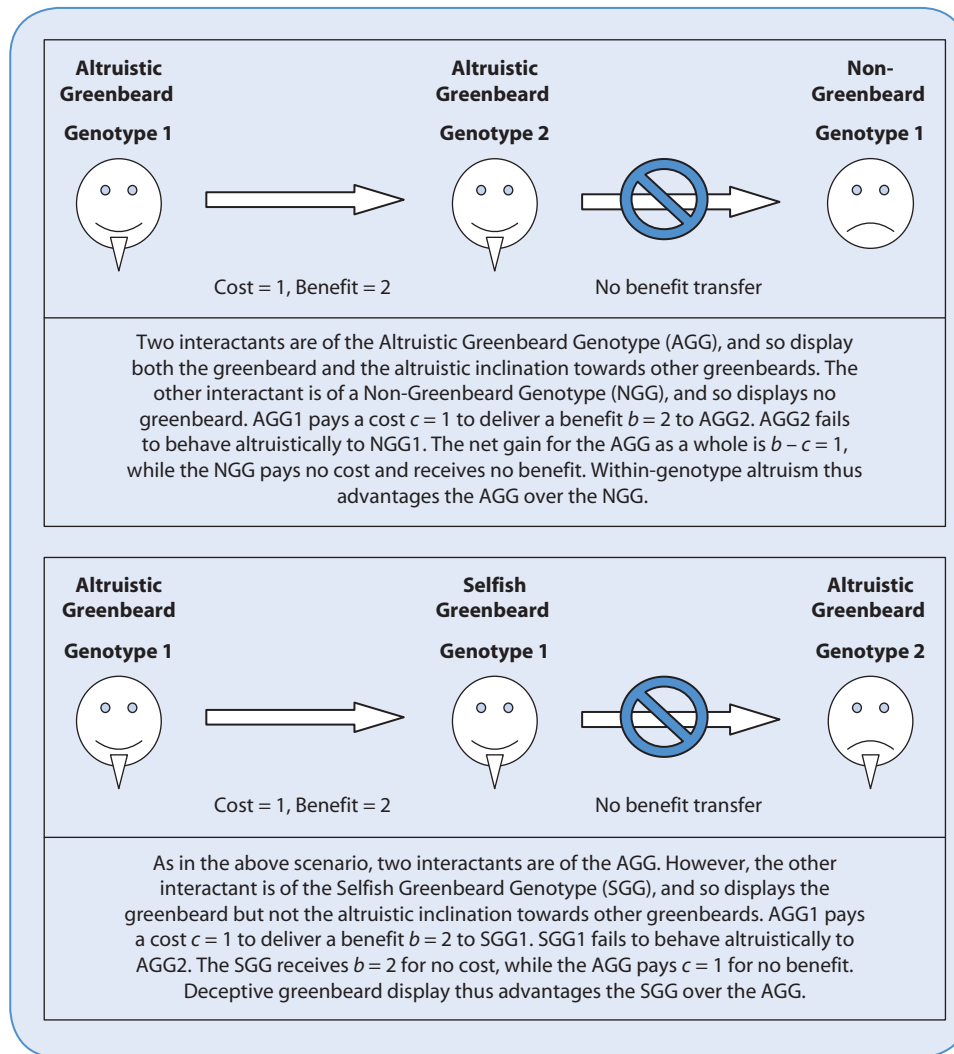


Figure 3.1. Altruistic greenbeards beat non-greenbeards, but selfish greenbeards beat altruistic greenbeards.

that interacting individuals be of the same cooperative genotype. Cooperative genotypes can proliferate by interacting preferentially with cooperative phenotypes, and r can be further broadened to represent the probability that an altruism beneficiary has a cooperative phenotype (Fletcher and Doebeli, 2006; Queller, 1985). In other words, a general rule for adaptive cooperation is 'cooperate with other co-operators', regardless of whether you share the exact same genes for cooperation with these others. If this general rule is followed, then cooperation can potentially evolve between genetic non-kin, indeed even between members of different species. In the next section, we will look first at the best-known example of how cooperation can evolve between cooperative phenotypes, irrespective of genotype: reciprocal altruism.

COOPERATION VIA RETURN BENEFITS (RECIPROCAL ALTRUISM, INDIRECT RECIPROCITY AND COSTLY SIGNALLING)

Reciprocal altruism

Shortly after Hamilton (1963, 1964) made the case for the evolution of cooperation via genic self-favouritism, a postgraduate in biology named Robert Trivers began thinking about other ways in which cooperation could evolve, and developed the theory of reciprocal altruism (Trivers, 1971). The basic principle of reciprocal altruism is that cooperation can evolve if interaction partners engage in a mutually beneficial exchange of costly altruistic acts. If you can produce a benefit for me more easily than I can for myself, and I can produce a benefit for you more easily than you can for yourself, then we can benefit mutually via exchange. As long as we can trust each other to reciprocate, the return benefit produced via exchange can more than compensate for the cost of benefit delivery, and so cooperation can be individually adaptive. The expectation of future reciprocity, however, must be accurate. If I pay a cost to benefit you, but you end up defecting rather than reciprocating, then my fitness will suffer. So if I have reason to think that you will not prove to be a reliable reciprocator in the future, and that you will instead attempt to gain a cheater's advantage, I should refuse to cooperate with you in the present. Trivers' theory famously found support in the computer simulations of Axelrod and Hamilton (1981), who demonstrated the adaptiveness of a simple 'tit-for-tat' strategy: cooperate at first with your partner, continue to cooperate as long as your partner continues to reciprocate and stop cooperating if your partner defects. Although the plausibility of reciprocal altruism became established in early formal models such as this one, new models continue to examine reciprocal altruism from novel angles (e.g. Imhof and Nowak, in press; Rand, Ohtsuki and Nowak, 2009).

Interactions involving true reciprocal altruism have a Prisoner's Dilemma type of pay-off structure. In a two-player Prisoner's Dilemma, each player gains more from mutual cooperation than from mutual defection, but can gain the biggest pay-off of all by defecting when the partner cooperates. Note that reciprocal altruism should not be confused with mutualism (sometimes referred to as 'by-product mutualism'; Brown, 1983), which occurs when an organism delivers a benefit to another organism in a cost-free manner, in the process of pursuing its own interests. Because this altruism is cost-free, defection ceases to be advantageous. For example, hummingbirds provide valuable pollinator services for flowering plants, but it costs them nothing to do so; this service occurs as a beneficial by-product of the hummingbird's efforts to consume plant nectar. Since the hummingbird's service is cost-free, it cannot gain a cheater's advantage by defecting. There are some costs involved for the plant, which must produce flowers in order to attract and feed hummingbirds, but these costs are just a fee that the plant must pay in order to ensure delivery of the hummingbird's positive externalities. Such mutually beneficial relationships are common in

nature, both within species and between them. For instance, by huddling together for warmth, puppies utilise each others' excess body heat; when cleaner shrimp eat parasites off of fish, they get fed and the fish get rid of pests; and when gut flora reside in the human digestive tract, they gain a place to live and humans acquire a host of beneficial services (Guarner and Malagelada, 2003).

Because both partners can gain from defection in reciprocal altruism, unlike in mutualism, reciprocally altruistic relationships tend to be less stable than mutualistic ones. Reciprocal altruism is also relatively cognitively demanding, since it requires that interactants remember past interactions with potential reciprocal partners and that they detect and avoid cheaters. For these reasons, reciprocal altruism appears to occur relatively rarely across species as compared to mutualism. Nevertheless, researchers have claimed to observe true reciprocal altruism between, for example, wrasse cleaner fish and their hosts (Bshary and Grutter, 2002; Bshary and Schäffer, 2002), predator-inspecting stickleback fish (Milinski, 1987, 1990), blood-exchanging vampire bats (Wilkinson, 1984, 1988) and grooming, alliance and sharing partners of various primate species (for a review of the mixed evidence on reciprocity in primates, see Silk, 2005).

However, identifying reciprocal altruism in non-human species is often a difficult and controversial undertaking, in part because the lines between mutualism and reciprocal altruism tend to blur easily. As a result, interactions identified as reciprocal altruism by some researchers have been considered mutualism by others. For example, mobbing of predators by flycatcher birds is considered reciprocal altruism by Krams *et al.* (2008) and mutualism by Russell and Wright (2008), and researchers disagree about whether predator inspection in fish constitutes reciprocal altruism or mutualism (Connor, 1996; Dugatkin, 1996). Another problem with identifying reciprocal altruism is that it is often surprisingly difficult to judge whether an act of altruism from A to B should be regarded as having been contingent on a previous act of altruism from B to A (Silk, 2005). So, despite the very compelling theoretical plausibility of reciprocal altruism, biologists have identified relatively few uncontroversial examples of its occurrence in nature. This dearth, however, in no way diminishes the truth of the observation that reciprocity is massively important in human social life. Human social interactions are universally saturated with the logic of reciprocity (Brown, 1991; Gouldner, 1960; Trivers, 1971), and humans are generally considered the clearest and best example of a truly reciprocally altruistic species. Indeed, the theory of reciprocal altruism has served as the foundation for some of the best-known research in evolutionary psychology, that on cheater detection in the Wason Selection Task (e.g. Cosmides and Tooby, 1992, 2005); this research suggests that the human mind contains a mechanism which is specifically devoted to detecting violators of reciprocal social contracts.

Reciprocal altruism is not always necessarily distinct from genic self-favouritism

Note that there is no necessary reason why reciprocal altruism must involve genic self-favouritism. All reciprocal altruism requires is reciprocating phenotypes, and

it could even involve partners of different species. Even if interactants are of the same species, reciprocal altruism need not involve genic self-favouritism. For instance, a variety of converging evidence suggests that, in chimpanzees, females provide sex to males in exchange for meat (Galdikas and Teleki, 1981; Stanford *et al.*, 1994; Teleki, 1973; Tutin, 1979); however, the genes enabling males to share meat are not necessarily the same as those enabling females to provide sex in exchange. On the other hand, although reciprocal altruism does not necessarily depend on genic self-favouritism, it may potentially involve it. If members of the same species engage in the exchange of the same kind of resource (e.g. if they groom one another), or in a relatively generalised type of exchange in which any type of resource can be traded (as with humans), then it is reasonable to expect that the same adaptations and hence genes are enabling this cooperativeness in each interactant (Price, 2006). If the genes encoding cooperative behaviour were the same in the interactants, then it would be reciprocal altruism *and* genic self-favouritism; if the genes were different, then it would just be reciprocal altruism. This example shows how reciprocal altruism and genic self-favouritism are often not as distinct as is typically assumed, and how they can easily shade into one another (Humphrey, 1997; Rothstein, 1980).

Indirect reciprocity

With reciprocal altruism, the assumption is that Individual A will learn about Individual B's history of cooperation and defection based on A's own interactions with B. With indirect reciprocity (Alexander, 1979, 1987), the assumption is that A could learn this history based on information about B's interactions with other partners. A could obtain this information, for example, by observing B's interactions with others, or via reputational information about B that was reported by others. If A makes a decision about whether to cooperate with B based on such second-hand experience – for example if A chooses to act cooperatively with B because A believes that B acted cooperatively with C in the past – then indirect reciprocity will have occurred. Indirect reciprocity is fundamentally similar to reciprocal altruism in several major ways. First, as with reciprocal altruism, verbal models of indirect reciprocity have been extensively supplemented and supported with more formal models (Leimar and Hammerstein, 2001; Nowak and Sigmund, 1998, 2005; Ohtsuki, Iwasa and Nowak, 2009; Panchanathan and Boyd, 2004). Second, while indirect reciprocity has been well-documented in humans (Rockenbach and Milinski, 2006; Semmann, Krambeck and Milinski, 2004, 2005; Simpson and Willer, 2008; Stanca, 2009; Wedekind and Milinski, 2000), its importance to other species is significantly less clear. And finally, indirect reciprocity may or may not involve genic self-favouritism. In order to evolve, indirect reciprocity must at a minimum enable interactants with cooperative phenotypes to have a better-than-random chance of interacting with one another. If these phenotypes are encoded by the same genotypes, then the indirect reciprocity will also constitute genic self-favouritism; if not, it will qualify as indirect reciprocity only.

Costly signalling

The costly signalling theoretical framework was developed in biology by theorists such as Zahavi (1975, 1977) and Grafen (1990) and more recently has been used to explain human cooperation (Bird, Smith and Bird, 2001; Gintis, Smith and Bowles, 2001; Gurven *et al.*, 2000; Hardy and van Vugt, 2006; Iredale, van Vugt and Dunbar, 2008; Nelissen, 2008; Smith and Bird, 2000). The basic idea of costly signalling theory is that a costly trait, such as engagement in altruistic behaviour, can bring reputational benefits to the signaller and thus make him or her seem more attractive to others as a social partner. The return benefits proposed by costly signalling theory, then, come from the increased social opportunities that one can acquire via engagement in conspicuous cooperation.

What types of desirable qualities might one be advertising through cooperative acts? An obvious answer would be cooperative disposition itself: what better way to signal your cooperativeness, and facilitate positive assortment with other cooperators than by broadcasting your cooperative deeds (Gurven *et al.*, 2000; Hardy and van Vugt, 2006; Nelissen, 2008)? However, it is also possible that a cooperative act could be used to advertise qualities that are only incidentally related to cooperative disposition itself. Smith and Bird (2000), for instance, suggest that among the Meriam of Australia, men make a great effort to hunt turtles, and share the meat widely, so that they can advertise qualities such as hunting ability, health and vigour. These qualities may make them more attractive as allies and mates, but this attractiveness is not due to them being perceived as particularly good cooperators.

So one version of costly signalling theory proposes that cooperators advertise cooperativeness itself (let's call this the 'auto-signal' theory), and another version suggests that they advertise qualities that are only incidentally related to cooperativeness (let's call this the 'other-signal' theory). While both versions have appeared in the literature on altruism and costly signalling in humans, the other-signal theory has been more prevalent in literature on non-human animal behaviour. The other-signal explanation, for example, has commonly been used to explain food sharing in birds. It has been suggested that males share with females in order to advertise their mate quality (Helfenstein *et al.*, 2003; Wiggins and Morris, 1986), and that males may share with other males in order to display their dominance (Kalishov, Zahavi and Zahavi, 2005), but it not usually suggested that sharing serves as a signal of general cooperative disposition.

The distinction between the auto-signal and other-signal versions is important, because it affects how costly signalling theory relates to the other theories of cooperation that we have discussed so far. The auto-signalling theory is in fact difficult to distinguish from theories of reciprocal altruism and indirect reciprocity (both of which, as we have seen, could potentially involve greenbeard effects), because the signaller is seen as cooperating in order to attract cooperative partners for reciprocal relationships. In contrast, the other-signalling theory is qualitatively different from any other theory we have discussed. For example, if sharing meat attracts mates because it signals hunting ability, health and vigour, then reciprocal altruism and indirect reciprocity should not be important (since the signaller's value as a reciprocal partner is not

what is being evaluated), and genic self-favouritism should also be irrelevant (because the genes that cause the signaller to share meat should be different from those which cause females to be attracted to him).

If the auto-signal theory is correct, then it may (ironically) be unnecessary, because it seems redundant with the two reciprocity theories. However, the other-signal theory is sometimes cast explicitly as an alternative to reciprocity theories (Bird, Smith and Bird, 2001; Gintis, Smith and Bowles, 2001; Smith and Bird, 2000), and could very well capture some dynamics of cooperation that are overlooked by reciprocity theories. Price (2003) tested between the other-signal and reciprocity theories in order to see which theory would better predict behaviour in a small-scale society; he found better support for reciprocity, but additional studies are needed.

SUMMARY OF INDIVIDUAL-LEVEL THEORIES OF COOPERATION

Table 3.2 summarises the six individual-level theories of cooperation that are discussed above, and notes where each one falls on the spectrum from genic self-favouritism to return benefits. On one end of the spectrum, we have kin and greenbeard altruism, two kinds of genic self-favouritism which can evolve without return benefits for cooperative individuals; the only required beneficiary is the shared gene for cooperation. (However, even though genic self-favouritism does not require return benefits, it will nonetheless often involve them, since there will often be a positive correlation

Table 3.2. Comparing individual-level evolutionary theories of cooperation.

	<i>Necessarily involves genic self-favouritism?</i>	<i>Potentially involves genic self-favouritism?</i>	<i>Necessarily involves return benefits for cooperator?</i>	<i>Potentially involves return benefits for cooperator?</i>
Kin altruism	Yes	—	No	Yes
Greenbeard altruism	Yes	—	No	Yes
Reciprocal altruism	No	Yes	Yes	—
Indirect reciprocity	No	Yes	Yes	—
Costly signalling ('auto-signal')	No	Yes	Yes	—
Costly signalling ('other-signal')	No	No	Yes	—

between receiving aid from an individual who shares one's cooperative genes and delivering aid to that individual. For instance, if A and B are siblings, then just as A will be inclined to aid B, B will be inclined to aid A.)

In the middle of the spectrum we have the reciprocity theories: reciprocal altruism and indirect reciprocity, and the auto-signal version of costly signalling theory (as we have seen, this last theory is largely redundant with the reciprocity theories). The reciprocity theories do require return benefits, but do not require genic self-favouritism and can even account for cooperation between members of different species. However, reciprocity can potentially involve genic self-favouritism, and would be likely to if carriers of the same gene(s) for reciprocity had some means of interacting preferentially with one another. On the far return benefits end of the spectrum we have the other-signal version of costly signalling theory, in which cooperative behaviour signals some attractive quality that is only incidentally related to cooperation itself. The other-signal theory must involve some sort of return benefit for the signaller (usually in the form of social partners such as mates or allies), but it does not involve genic self-favouritism.

GROUP SELECTION

Up to now our discussion has focused on individual-level theories of cooperation, that is theories which focus on the individual cooperator as a vehicle for replicating genes. However, as mentioned above, in many species social groups can also be considered as relevant vehicles of selection. When individuals assort into groups, and selection operates among groups (i.e. if the members of some groups have higher average fitnesses than the members of other groups), then the search for cooperative adaptations potentially becomes more complicated. For example, an adaptation at one level may conflict with the functioning of an adaptation at another level, if it is good for the group but bad for the individual or vice versa.

The issue of how evolution may favour traits that benefit a whole group is known, in general, as 'group selection'. (The term 'group selection' has held various more specific meanings to different researchers, and the issue has been complicated by semantic confusion, as discussed below.) The idea that group selection may be relevant to the evolution of altruism has a long history in behavioural biology. For most of the twentieth century it was common for biologists to engage in what is now considered 'naive group selectionism' (Wilson and Wilson, 2007). In this type of thinking, apparently altruistic individual behaviour is explained in terms of the benefits that it brings for the individual's group, without adequate consideration of how the behaviour could have avoided being selected against at the individual level. The classic example of naive group selectionism is Wynne-Edwards (1962), who proposed that animals refrain from reproducing too much at the individual level so that they can avoid overexploiting their resources at the population level.

The publication that was most instrumental in ending the era of naive group selectionism was George Williams' (1966) influential critique of evolutionary thought,

Adaptation and Natural Selection. Williams believed that selection at the individual level would, under normal circumstances, be much more intense than selection at the group level, and so he agreed with Darwin's (1859/1958) usual view that adaptations evolve to benefit not the success of groups or species but rather the fitness of individuals. In most species, most of the time, successful gene replication will depend much more on the fitness of the individual carrier, as opposed to the average fitness of the carrier's group. Therefore, when there is a conflict between individual-level and group-level fitness, the individual level will prevail. Williams thus emphasised that when an individual-level explanation for the evolution of altruism is sufficient, one should not invoke a group-selectionist explanation: 'When recognized, adaptation should be attributed to no higher a level of organization than is demanded by the evidence' (Williams, 1966, p. v). In the wake of Williams' critique, the term 'group selection' acquired a pejorative meaning to most biologists. A group selectionist was considered someone who believed that in a conflict between individual and group fitness, the group level would prevail. Most biologists became highly sceptical about this kind of group selectionism, and individual-level adaptation became the focus of most studies in behavioural biology and evolutionary psychology.

However, despite all the scepticism, group selection has been seriously considered by many influential theorists. Darwin himself, deviating from his normal emphasis on individual fitness, entertains group selectionist ideas in the *Descent of Man* (1871/1981). (However, he also considers problems with these ideas. For more discussion of Darwin's views on group selection, see Box 3.1.) And while Williams' critique did discourage naive group selectionist theories, it also encouraged more rigorous thinking about group selection. Shortly after Williams' critique, for example, G. R. Price (1970, 1972) published his eponymous equation showing how gene frequency changes in a population are the joint product of within-group and between-group selection. Interest in group selection is probably as strong now as it has been at any time over the past few decades, with theorists continuing to develop new models of how group selection could permit cooperation to evolve, and to argue that its semi-banishment from mainstream evolutionary theory was hasty and unwarranted (e.g. Gintis, 2000; Wilson and Sober, 1994; Wilson and Wilson, 2007). Like the Price equations, recent models have tended to emphasise multilevel selection, that is the joint effects of selection operating at both the individual and group levels, and sometimes at other levels (e.g. intragenomic, species) as well (Wilson and Wilson, 2007).

If group selectionism seems inconsistent with the fundamental focus in evolutionary biology of individual-level fitness and adaptation, then why has it made something of a comeback in recent years? There are several reasons, the first and least substantive of which is recent semantic confusion about what constitutes 'group selection'. Some arguments in favour of 'group selection' seem to actually be attempts to broaden its definition, in order to subsume processes that had previously been considered cases of individual selection. For instance, Wilson and Wilson (2007) suggest that the evolution of reciprocal altruism in a population of defectors could be considered a case of group selection, because the reciprocators and the defectors can each be considered a separate group, and the reciprocators have a higher average fitness than the defectors. Similarly, Wilson and Dugatkin (1997) suggest that assortative interactions among



BOX 3.1. DARWIN'S SPECULATIONS ABOUT GROUP SELECTION IN THE *DESCENT OF MAN*

When advocates of group selection or multi-level selection want to demonstrate that Darwin seemed open to these ideas, they often (e.g. Sober and Wilson, 1998; Wilson, 2002) quote the following passage from Chapter 5 of the *Descent of Man, and Selection in Relation to Sex*:

There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage and sympathy, were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection.

(Darwin, 1871/1981, p. 166)

It seems clear from this passage that Darwin is entertaining group selectionist ideas. However, broader consideration of Darwin's speculations in this chapter reveals that he was not simply engaging in 'naive group selectionism', and on the contrary, that he was well aware of the problems for group selection that would be caused by individual selection (Wilson, 2007; Wilson and Wilson, 2007). For example, consider Darwin's (1871/1981, p. 163) musings from earlier in the chapter:

But it may be asked, how within the limits of the same tribe did a large number of members first become endowed with these social and moral qualities, and how was the standard of excellence raised? It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those which were the most faithful to their comrades, would be reared in greater number than the

children of selfish and treacherous parents of the same tribe. He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature. The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others, would on an average perish in larger number than other men. Therefore it seems scarcely possible (bearing in mind that we are not here speaking of one tribe being victorious over another) that the number of men gifted with such virtues, or that the standard of their excellence, could be increased through natural selection, that is, by the survival of the fittest.

These passages reveal that in the case of human morality, Darwin perceived a conflict between the individual and group levels of selection: while adaptations for moral behaviour would have advantaged everyone in a group, they would not have advantaged the individuals who bore them. Darwin seems perplexed by this puzzle, and realises that a complete solution is beyond the scope of his work. However, he does suggest how the puzzle may begin to get solved, when he proposes some psychological mechanisms by which morality may satisfy individual interests. Interestingly, these mechanisms are reminiscent of two types of individually adaptive cooperative behaviours: reciprocal altruism and indirect reciprocity, which we have already discussed:

Although the circumstances which lead to an increase in the number of men thus [morally] endowed within the

same tribe are too complex to be clearly followed out, we can trace some of the probable steps. In the first place, as the reasoning powers and foresight of the members became improved, each man would soon learn from experience that if he aided his fellow-men, he would commonly receive aid in return . . . But there is another and much more powerful stimulus to the development of

the social virtues, namely, the praise and blame of our fellow men.

(Darwin, 1871/1981, p. 163–4)

In these allusions to reciprocal altruism ('receive aid in return') and indirect reciprocity (reputational effects associated with 'praise and blame'), Darwin is anticipating the kinds of individual-level solutions to the puzzle of cooperation that continue to be discussed and explored today.

altruists (e.g. greenbeard altruism) could be considered group selection, since the altruists would constitute a group in which members had relatively high average fitness. These re-labelling attempts are confusing to those who are used to thinking of group selection as an *alternative* to individual selection. As mentioned above, group selectionism has traditionally implied preference for a group-level explanation when there is a conflict between the individual and group levels. However, in the cases of reciprocal and greenbeard altruism, there is no conflict between the individual and group levels.

A second reason for renewed interest in group selection has been a proliferation of formal models showing how a kind of altruism known as 'strong reciprocity' could evolve by biological and/or cultural group selection (e.g. Boyd *et al.*, 2003; Gintis, 2000; Gintis *et al.*, 2003). A strong reciprocator is someone who acts cooperatively and punishes non-cooperators, even when doing so produces no direct return benefit for one's self. These formal models have been supplemented with a large body of cross-cultural data from experimental economic games that is presented as empirical evidence of strong reciprocity (e.g. Fehr, Fischbacher and Gächter, 2002; Henrich *et al.*, 2005). The interpretation of these results, however, has been questioned, mainly on the grounds of ecological validity: psychological mechanisms for cooperation which did produce return benefits in ancestral environments should not necessarily be expected to do so in the highly artificial environments of experimental economic games (Burnham and Johnson, 2005; Hagen and Hammerstein, 2006; Price, 2008; Trivers, 2004; West, Griffin and Gardner, 2007).

Finally, group selection continues to attract interest for the fundamental reason that it is an important biological phenomenon that does occur in nature. For example, in a strain of *Pseudomonas fluorescens* bacteria known as the 'wrinkly spreader', some individuals help produce a cellulosic polymer that enables the group to survive in the absence of oxygen. Polymer production is costly, and non-producers have higher relative within-group fitness; nevertheless, genes for producing the polymer are maintained via group selection (Rainey and Rainey, 2003). Group selection also helps explain why parasites often evolve to be less virulent to their hosts than they could be: parasites which reproduce relatively rapidly will increase in frequency within their host,

but will risk killing their host and thus reducing their own chances of spreading to another host. Their ‘cheating’ (reproducing too quickly) damages their own long-term fitness, as well as that of other parasites who reside within the same host (Frank, 1996). However, probably the most fascinating example of group selection is the individual organism itself. According to the most widely accepted framework for explaining the major transitions in the history of life (Margulis, 1970; Maynard Smith and Szathmáry, 1995), the evolution of increased biological complexity has progressed via group selection. The first genes were lone replicating chemicals, which eventually began to join forces and surround themselves with the earliest prokaryotic cells; these simple cells began to cooperate with one another to form more complex eukaryotic cells, and groups of cells were eventually selected to compose the first multicellular organisms. At every major transition, lone biological agents found they could solve adaptive problems by forming cooperative groups, and these groups were ultimately selected to become the highly integrated packages that today we perceive as individuals.

As the driving force in the major transitions in life, group selection cannot be ignored as a fundamental and important biological process. Still, it is important to remember that in most cases of group selection, such as those referenced above, conflicts between the individual and group levels are minimal. For example, the various cellular types in a multicellular organism will generally all have an interest in working together to promote organismal fitness, since each type’s own fate depends highly upon this fitness. If one cellular type acted to selfishly promote its own representation in the organism at the expense of another cellular type, and the result were organismal death, then this selfishness would be self-destructive. In other words, the best way that each type can promote its own long-term reproductive interests is to cooperate with other types, and this confluence of interest minimises the need to distinguish between the fitness interests of the ‘group’ (which in this case happens to be the organism) and the ‘individuals’ (the cell types). However, even this extremely high degree of alignment in fitness interests is not sufficient to completely preclude conflict. Various kinds of intragenomic conflict can cause within-organism cooperation to break down, sometimes with devastating consequences for organismal fitness (Burt and Trivers, 2006). For example, segregation distorter genes promote their own replication by increasing their representation in the organism’s gametes, and damaged DNA may cause one cellular type to reproduce at the expense of other types (e.g. cancer).

COMPLEX HUMAN COOPERATION: COLLECTIVE ACTION

We have now covered the major individual-level adaptationist theories of cooperative behaviour, in addition to the theory of group selection. While these theories apply to a huge range of species, and not just to humans, each one has been used at one time or another to explain some aspects of human cooperative behaviour. However, with

regard to one kind of human cooperative behaviour in particular, there is still a major lack of consensus about which kind of evolutionary explanation is correct.

Collective action: The remaining puzzle of human cooperation

Humans are a highly cooperative species, distinguished among other species by their remarkable ability to cooperate in huge groups of individuals who are not close genetic relatives (Boyd *et al.*, 2003). As mentioned at the outset of this chapter, this ability to cooperate in large groups is a fundamental feature of human sociality. However, it is also the aspect of human cooperation that is most difficult to explain. In order to understand the difficulty of explaining it, let's consider the nature of these groups in more detail. These groups typically take the form of collective actions in which any number of individuals engage in the joint production of a shared resource such as a public good (a public good is a resource that all group members will be able to access equally). For example, joint efforts to protect a whole group from attack from human or non-human predators, or to increase a whole group's access to food or water, would be considered public good-producing collective actions. In such collective actions, the extent of an individual group member's net benefit from the public good is determined by the extent to which he or she contributed to its production. Since all members have equal access to the good, those who sacrifice the most to produce it will end up reaping the lowest net benefit; therefore, each member should strive to contribute as little as possible (i.e. to free ride as much as possible on the efforts of other members), and collective actions should generally tend to unravel and fail (Olson, 1965). This classic quandary is variously known as the 'free rider problem', 'collective action problem', 'social dilemma' or 'tragedy of the commons' (Hardin, 1968). The free rider problem is not quite ubiquitous because it should be absent if the benefits of cooperation are high enough (see Box 3.2 for a discussion about the conditions under which free rider problems should arise in collective actions). But in most collective actions, sustained cooperation will depend on successful resolution of the free rider problem.

Research suggests that various types of solutions to the free rider problem do, thankfully, exist. Studies of collective actions often take the form of experimental economic public goods games. These studies suggest several ways in which group members strive to solve the collective action problem, that is how they attempt to contribute to the public good while minimising the extent to which they are exploited by free riders. First, they engage in conditional cooperation, which is a kind of reciprocal altruism in which they contribute more highly if they observe or expect high contributions from co-members (Fischbacher, Gächter and Fehr, 2001; Kurzban and Houser, 2005; Ledyard, 1995). Second, they engage in the punishment of free riding co-members, even when it is costly to the individual to engage in such punishment (Fehr and Gächter, 2002; Ostrom, Walker and Gardner, 1992; Price, 2005; Price, Cosmides and Tooby, 2002; Yamagishi, 1986). Third, they engage in partner choice: if they have some ability to choose their co-members, then higher contributors tend to choose to form groups with each other and to thus exclude free riders from the public



BOX 3.2. WHY THE FREE RIDER PROBLEM DISAPPEARS IN EXTREMELY PRODUCTIVE COLLECTIVE ACTIONS

To really understand why the free rider problem should in theory disappear when the benefits of cooperation become very high, it helps to understand the simple maths which describe the situations in which the free rider problem should and should not arise.

First, consider a collective action in which the value of the benefit produced is *not* extremely high. In this kind of interaction, free riding will be individually beneficial. Imagine that four villagers ($n = 4$) each have an equal amount of food growing in a shared garden. It has been raining for several days and three of the villagers decide to construct a wall to protect their garden from a flood; they cannot convince the fourth villager to help. Every one unit of contribution effort ($c = 1$) will produce two units of benefit ($b = 2$) in the form of protected food. Each of the three workers contributes $c = 1$, so they collectively produce $3 \times 2 = 6$ units of benefit. This benefit is shared equally by all four villagers, for a per capita benefit of $6 / 4 = 1.5$. Each worker receives a net gain of $1.5 - 1 = 0.5$, while the free rider receives 1.5 at no cost. If the free rider had contributed at the same level as the others ($c = 1$), then the group would have collectively produced $4 \times 2 = 8$ units of benefit, for a per capita benefit of $8 / 4 = 2$, and a per capita net gain of $2 - 1 = 1$. So had the free rider contributed like the others, he would have done worse (his profit would have been 1 instead of 1.5), and each other member would have done better (each one's profit would have been 1 instead of 0.5).

Next, consider a collective action in which the value of the benefit produced is extremely

high. In this kind of interaction, free riding will be individually costly. Imagine that all of the above details are the same, except that the four villagers have been enduring serious food shortages, and so their garden food has become more valuable to them. Now, every one unit of contribution effort ($c = 1$) will produce not 2 but 5 units of benefit ($b = 5$) in terms of protected food. Again, three of the villagers contribute to building a flood wall while the fourth free rides. Each of the three workers contributes $c = 1$, so they collectively produce $3 \times 5 = 15$ units of benefit, for a per capita benefit of $15 / 4 = 3.75$. Each worker receives a net gain of $3.75 - 1 = 2.75$, while the free rider receives 3.75 at no cost. If the free rider *had* contributed at the same level of the others ($c = 1$), then the group would have collectively produced $4 \times 5 = 20$ units of benefit, for a per capita benefit of $20 / 4 = 5$, and a per capita net gain of $5 - 1 = 4$. So had the free rider contributed like the others, he would have done better (his profit would have been 4 instead of 3.75), and so would each of the other members (each one's profit would have been 4 instead of 2.75). Instead of benefiting from his free riding, the free rider suffered for it, because in this collective action the benefits of cooperation were very high.

The individual incentive to free ride, then, falls off in highly productive collective actions. Specifically, when the cost of an individual contribution is set at 1, free riding will be individually costly when the benefit produced by each contribution is greater than the number of collective action beneficiaries (i.e. when $c = 1$ and $b > n$).

good (Barclay and Willer, 2007; Ehrhart and Keser, 1999; Page, Putterman and Unel, 2005; Sheldon, Sheldon and Osbaldiston, 2000).

How well do the previously reviewed theories of cooperation explain behaviour in collective actions? Kin selection seems insufficient, because collective actions are often composed of non-kin. Other-signal costly signalling theory has been proposed as an explanation of some forms of public good provisioning (Bird, Smith and Bird, 2001; Gintis, Smith and Bowles, 2001; Smith and Bird, 2000), but it does not seem to predict key behaviours that have been observed in collective actions such as conditional cooperation and punishment of free riders. Because kin selection and other-signal costly signalling both seem to offer largely incomplete explanations of collective action, they will not be considered further here. Auto-signal costly signalling will also not be considered further as a distinct theory, since, as noted above, it is largely redundant with reciprocal altruism and indirect reciprocity. That leaves the following theories available for consideration: reciprocal altruism, indirect reciprocity, greenbeard altruism and group selection.

The reciprocity theories as explanations for collective action

Traditionally, reciprocal altruism has been represented as simple dyadic exchange, a model which is obviously inadequate for explaining collective actions involving many members. However, efforts have been made to adapt reciprocal altruism to n -person collective actions (Tooby, Cosmides and Price, 2006). The first formal model of n -person reciprocity (Boyd and Richerson, 1988) modelled the reciprocity strategy as a discrete one which could contribute either fully or not at all (i.e. contribute either zero or one), depending on how many of its co-members contributed. The conclusion of this model was that this kind of reciprocity was not sufficient to overcome the free rider problem and could therefore not account for the evolution of 'sizeable' collective actions; discrete reciprocity performs reasonably well in very small groups (e.g. three to five members) but starts doing badly if groups get much larger than this. However, this conclusion seems incompatible with the observation that reciprocal altruism (conditional cooperation) is commonly observed in collective actions. A later model represented reciprocal altruism in a more realistic form, not as a discrete all-or-nothing strategy but rather as a continuous strategy which matches the average co-member contribution and can contribute any amount between zero and one. This strategy, which mimics the way in which conditional cooperators in public goods games have been observed to behave, in some respects performs significantly better in collective action contexts than the discrete strategy does (Johnson, Price and Takezawa, 2008). However, even continuous reciprocity has difficulty evolving (i.e. invading a population of unconditional defectors) in sizeable collective actions (e.g. 100 members), except for in those which are extraordinarily productive.

Further studies and more formal models are needed to investigate whether and how reciprocal altruism could evolve in collective actions. While the models described above suggest that reciprocity has difficulty evolving in large groups, it is possible that human adaptations for reciprocity evolved in small groups, and that these adaptations are nevertheless routinely deployed in the larger groups that frequently characterise

modern societies. Indeed, in human ancestral environments, small groups would have been the norm (e.g. Kelly, 1995; this issue, and the vital importance in evolutionary psychology of considering the nature of human ancestral environments in general, will be discussed in more detail in the conclusion to this chapter). Further, it may be the case that adaptations for reciprocity evolved simultaneously and synergistically with other cooperative behaviours, such as indirect reciprocity, positive assortment and punishment of free riders. For example, reciprocity and positive assortment interact synergistically in collective actions (Boyd and Richerson, 1988; Johnson, Price and Takezawa, 2008). If synergistic interactions between reciprocity and other behaviours occur, then models which examine the evolution of reciprocity in isolation, unaffected by these other behaviours, may underestimate its adaptiveness. Whatever the true story about the evolution of reciprocity in groups turns out to be, the ubiquity of reciprocity in collective actions lends credibility to the view that reciprocal altruism has indeed played a role in enabling the evolution of collective action.

It also seems plausible that indirect reciprocity has played a role in the evolution of collective action. In contrast to the case of reciprocal altruism, relatively little controversy has been associated with formal models showing how indirect reciprocity could evolve in collective action contexts (Nowak and Sigmund, 2005; Panchanathan and Boyd, 2004). Evidence from experimental economic games supports these formal theories: when contributors to collective actions engage in partner choice as a way of engaging other contributors and avoiding free riders, they tend to base their decisions on the information they have about potential partners' histories of cooperative interactions with other people (Barclay and Willer, 2007; Ehrhart and Keser, 1999; Page, Putterman and Unel, 2005; Sheldon, Sheldon and Osbaldiston, 2000).

Greenbeard altruism as an explanation for collective action

Greenbeard altruism's role in enabling the evolution of collective action remains a largely unexplored issue. If human cooperation has evolved significantly by greenbeard dynamics, then the conspicuous label of cooperative disposition required by these dynamics has likely been relatively non-arbitrary. Greenbeard systems become more unstable when they rely on a label of cooperative disposition that has a more arbitrary relationship with actual cooperative disposition. An arbitrary label like a green beard can easily be displayed by someone who does not truly possess a cooperative disposition, but it is more difficult to fake a label that can only be produced via actual engagement in cooperative behaviour. If greenbeard dynamics have been important in the evolution of human cooperation, therefore, then actual engagement in cooperative behaviour may have been the most common type of phenotypic label (Price, 2006; Wilson and Dugatkin, 1997). This system would evolve via mutations that caused people to engage in cooperative behaviour and to direct this behaviour towards people who displayed a similar tendency to engage in cooperative behaviour; as long as these behaviours were encoded by the same gene(s) in all of these interactants, it would be true genic self-favouritism. Thus, when high contributors to collective actions cooperate more with other high contributors, via either reciprocal

altruism (conditional cooperation) or indirect reciprocity (partner choice), an underlying process of greenbeard dynamics could potentially be involved.

Group selection as an explanation for collective action

Finally, could the evolution of collective action have been facilitated via a process of group selection? It is possible, even if the individual-level reciprocity processes (with or without greenbeard dynamics) have also been important. In a process of multi-level selection (Wilson and Wilson, 2007), collective action could have evolved via both group- and individual-level selection. The group-selection component of this process could retain genes for behaviour that was fitness-damaging to individuals but fitness-enhancing for groups, provided that selection among groups was sufficiently intense and that migration among groups was sufficiently limited. However, it has not proven easy to generate evidence for group-selected human cooperative behaviour. One problem is that researchers of this topic often look for negative as opposed to positive evidence: in controlled laboratory conditions, experimenters attempt to eliminate the possibility that public goods games players could be striving to gain individual-level benefits by cooperating (e.g. the experimenters impose strict anonymity on all players, and make the games one-shot, in order to preclude reputation effects and reciprocity), and if players still continue to behave cooperatively, this residual cooperativeness is considered evidence of group selection (Fehr, Fischbacher and Gächter, 2002; Henrich *et al.*, 2005). Critics of this approach point out that behaviour which appears selfless in environmentally novel laboratory conditions may not have been so in ancestral environments (Burnham and Johnson, 2005; Hagen and Hammerstein, 2006; Price, 2008; Trivers, 2004; West, Griffin and Gardner, 2007). If positive rather than negative evidence for group selection could be produced, and if this evidence were predicted by group-selection theory alone, then the relevance of group selection would be easier to demonstrate.

Despite the lack of unequivocal evidence, some evidence is interpreted by some researchers as indicating that group selection has been important in the evolution of collective action. For example, as noted above, many contributors in public goods games will engage in the costly punishment of free riding co-members. This punishment may seem to provide a selfish incentive for would-be free riders to cooperate, since cooperation allows them to avoid punishment. However, punishment does not provide a good general individual-level solution to the problem of collective action, since punishment itself seems to constitute a second-order public good: it is costly to provide, and it benefits the group by eliciting contributions from would-be free riders. Punishment simply creates a new, second-order problem of collective action (Boyd and Richerson, 1992; Yamagishi, 1986). Some modellers have suggested that a process of biological and/or cultural group selection is necessary to resolve this second-order problem (Boyd *et al.*, 2003; Gintis, 2000; Gintis *et al.*, 2003). On the other hand, some researchers maintain that the second-order free rider problem would not have prevented free rider punishment from evolving by standard individual-level selection (Gardner and Grafen, 2009; Price, 2003, 2005; Price, Cosmides and Tooby, 2002). For more in-depth discussion on the topic of free riders and the second-order problem, see Box 3.3.



BOX 3.3. FREE RIDER PUNISHMENT AND THE SECOND-ORDER PROBLEM

As noted in the main text, punishment can help solve the free rider problem in collective actions: when the costs of being punished outweigh the advantages of free riding, cooperation becomes more profitable from free riding. Evidence from both experimental and real-life collective actions suggest that groups cooperate more productively when free riders can be punished, and that group members will willingly accept the costs associated with punishing free riders (e.g. Fehr and Gächter, 2002; Ostrom, 2000; Price, 2005; Yamagishi, 1986). These costs may include not just the expenses of inflicting punishment (e.g. energy expenditure, monetary costs) but also those that may result from retaliation: after being punished, many free riders will attempt to retaliate against their punishers (Cinyabuguma, Page and Putterman, 2005; Nikiforakis, 2008).

However, while punishment at first glance seems to provide a solution to the free rider problem, a closer look suggests that it may simply replace this problem with a new problem of second-order free riding. When a collective action participant punishes a free rider, he or she produces resources for other participants because he or she coerces a contribution to the public good from the would-be free rider. But while all participants benefit from this contribution, only the punisher pays for it. Therefore, punishers should be disadvantaged relative to 'second-order free riders', that is participants who accept the benefits produced by free rider punishment but who do not help pay for these benefits by punishing free riders themselves (Boyd and Richerson, 1992; Yamagishi, 1986). For this reason, punishment of free riders is sometimes referred to as 'altruistic punishment' (Boyd *et al.*, 2003; Johnson *et al.*, 2009).

How can this second-order problem be solved? As noted, some theorists believe that the solution requires group selection (e.g. Boyd *et al.*,

2003; Gintis, 2000; Gintis *et al.*, 2003). However, other theorists assume that there must be some individual-level explanation for the evolution of free rider punishment and the punitive sentiment which motivates it (e.g. Gardner and Grafen, 2009; Hauert *et al.*, 2007; Price, Cosmides and Tooby, 2002). Given the problems associated with resorting to group selection when individual selection may in fact be sufficient, one should carefully consider all of the ways in which punishing free riders may have benefited ancestral punishers.

One individual-level explanation for how free rider punishment evolved is second-order punishment; this view assumes that punishments are directed not just at free riders but also at those who fail to punish free riders. (This may seem to lead to an infinitely recursive problem of third- and higher-order free riding – participants who fail to punish participants who fail to punish free riders, and so on – but the recursion problem seems to become trivial beyond the second order; for a discussion, see Kiyonari and Barclay, 2008.) However, the same sorts of experimental methods which have produced strong evidence for free rider punishment have not provided much evidence for punishment of non-punishers (Kiyonari and Barclay, 2008; Kiyonari, van Veelen and Yamagishi, 2008).

Fortunately, there are several more plausible individual-level explanations for free rider punishment. For instance, it may be that free riders are punished because they are shunned in private transactions outside of the context of collective actions, a process of indirect reciprocity that avoids the second-order problem (Panchanathan and Boyd, 2004). Alternatively, in collective actions where certain kinds of conditions apply (e.g. the collective good must be excludable), free rider punishment can evolve without the second-order problem if members

of the population can opt out of collective action participation and instead pursue productive solitary pursuits (Hauert *et al.*, 2007).

Finally, and perhaps most promisingly, punishers may receive some kind of private reputational benefit as the result of others observing their behaviour. Support for this view is provided by experiments suggesting that participants who are being observed by an audience punish cheaters more than do those whose actions are anonymous (Kurzban, DeScioli and O'Brien, 2007; Piazza and Bering, 2008). What might be the nature of these reputational benefits? They could make the punisher seem like a relatively good choice as a cooperative partner; studies suggest

that punishers are viewed as more trustworthy and are preferred as partners (Nelissen, 2008). Or they could make the punisher seem like a relatively bad choice as someone to cheat (Price, 2003): a willingness to punish free riders may send a message to the general public along the lines of 'mess with me and this is what you get'. This reputation could provide private benefits for the punisher in social contexts besides collective actions (e.g. dyadic exchanges), which could more than compensate for the punisher's disadvantage relative to second-order free riders in collective actions. For a discussion of some other possible individual-level solutions to the second-order free rider problem, see Price (2003).

Perhaps the best evidence in favour of group selection in humans is that people occasionally exhibit extremely altruistic behaviours in collective actions – for instance the deliberate sacrifice of one's own life during coalitional conflict – that seem particularly hard to explain in terms of individual-level theories (Sosis and Alcorta, 2008). These behaviours can be challenging to study, especially because it is often difficult or unethical to elicit them in the laboratory. Nevertheless, if efforts are made to better understand extremely self-sacrificial human behaviour, group selection may turn out to be a necessary part of the explanation. On the other hand, people who agree to kill themselves for their coalition – for example suicide bombers – often receive extraordinarily high status for doing so, before they are actually dead (Axell and Kase, 2002; Moghadam, 2003). It could be that in ancestral environments, attaining such high status was so highly positively correlated with fitness benefits – especially for males (Daly and Wilson, 1988) – that selection designed people to regard the attainment of such status as a deeply desirable end in itself. In modern societies characterised by environmental novelties (e.g. high explosives), status-striving (e.g. by suicide bombers) may lead relatively frequently to maladaptive, premature death; however, in ancestral environments, such striving may have on average been adaptive.

CONCLUSION

Cooperation and the evolution of altruism and have been centrally important topics in behavioural biology since the days of Darwin, and will continue to be for the foreseeable future. A great deal of progress has been made in understanding these topics, especially since the shift in the mid-1960s in perspective which saw the gene

take centre stage as the fundamental unit of selection. However, despite this progress, much important work remains to be done, especially for evolutionary psychologists interested in the evolution of collective action. Because collective action is so centrally important in human sociality, and because it still presents some of the greatest challenges facing behavioural biology, I will focus in this conclusion on emphasising two ways in which evolutionary investigations of collective action could most profitably proceed.

First, for reasons emphasised throughout this chapter, before we resort to higher-level explanations for the evolution of collective action, we should first focus on thoroughly applying existing individual-level theories, such as reciprocity theories and positive assortment theories like greenbeard altruism (Price, 2006). Moreover, as noted above, we should analyse these individual-level cooperative behaviours not just in isolation from one another but as strategies which can exist simultaneously and potentially interact synergistically. While this kind of 'bottom-up' approach would start by looking for individual-level solutions, it would not preclude an investigation into the effects of group selection. However if we start with an exhaustive investigation at the individual level, then by the time we work our way up to the group level we will be in a good position to understand how group-beneficial behaviours could have escaped being selected against at the individual level. Cutting-edge modelling work on the individual-level evolution of collective action is currently being carried out by theorists such as Martin Nowak, Christoph Hauert and colleagues (e.g. Hauert *et al.*, 2007, 2008; Wakano, Nowak and Hauert, 2009). In addition, particularly interesting new experimental tests of individual-level theories of collective action are being carried out by researchers like Louis Putterman and colleagues (e.g. Bochet and Putterman, 2009; Ertan, Page and Putterman, 2009), Pat Barclay and colleagues (e.g. Barclay and Willer, 2007; Kiyonari and Barclay, 2008) and Mark van Vugt and colleagues (e.g. O'Gorman, Henrich and van Vugt, 2008; van Vugt and Spisak, 2008).

Second, in attempting to identify human adaptations for collective action participation, we should always remember that these adaptations were designed by environments that were in important ways different from those of modern societies (Tooby, Cosmides and Price, 2006). These adaptations have specific environments of evolutionary adaptedness (EEA), and while they must have functioned adaptively in these ancestral environments, we should not always expect them to function adaptively in modern environments (Symons, 1990; Tooby and Cosmides, 1992). Nevertheless, as noted, some evolutionary analyses do make the mistake of assuming that if an adaptation promoted individual fitness in past environments then it should also do so in present environments – even when these present environments are radically different from those of the past, for example when they are the experimental laboratories of modern universities.

It would also be good to keep ancestral conditions in mind when we consider the adaptiveness of cooperative behaviour in large groups. For example, the influential model of the evolution of reciprocity in collective actions discussed above (Boyd and Richerson, 1988), which suggested that reciprocity would be unlikely to evolve in large collective actions, has led some to believe that when humans participate in large collective actions in modern environments, their behaviour cannot be explained in terms

of adaptations for reciprocity (e.g. Fehr, 2004; Henrich, 2004). The problem with this conclusion is that an adaptation that was designed by small-group conditions in ancestral environments could nevertheless still function in large groups. For example, a decision rule for contributing in a collective action like ‘match the mean co-member contribution’ may scale up to a group of any size, even if it functioned adaptively in small groups and maladaptively in large groups. If humans do possess adaptations for reciprocity in collective actions, then it is likely that these did evolve in small groups. An average hunter-gatherer band contains only about seven or eight male and female full-time foragers (Kelly, 1995), which suggests that a typical same-sex collective action (e.g. a male hunting party) in ancestral environments may have consisted of only about four members. Human adaptations for reciprocity may have evolved in such small groups, and could govern human behaviour in the large groups of modern societies, regardless of whether they lead to adaptive outcomes in large groups.

In conclusion, our understanding of collective action is certainly challenged by levels-of-selection considerations, by the need to thoroughly apply individual-level theories of cooperation and by the requirement that we must always remember that adaptations for cooperation were designed by ancestral environments that were significantly different from those of today. Nevertheless, these challenges are currently being faced by many researchers, and there is no doubt that our understanding of the evolution of cooperation will improve as a result. If this understanding continues to progress as steadily as it has for the past several decades, then in the coming decades the puzzle of cooperation may no longer be regarded as a puzzle at all.

ACKNOWLEDGEMENTS

Thanks to Oliver Curry for drawing my attention to the depth of Darwin’s thinking about levels of selection, and to Jade Price for suggesting revisions to the manuscript draft.

REFERENCES

- Alexander, R. D. (1979). *Darwinism and human affairs*. Seattle: University of Washington Press.
- Alexander, R. D. (1987). *The biology of moral systems*. Hawthorne, NY: Aldine de Gruyter.
- Axell, A. and Kase, H. (2002). *Kamikaze: Japan’s suicide gods*. New York: Longman.
- Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396. New York.
- Barclay, P. and Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B*, 274, 749–53.
- Bird, R. L. B., Smith, E. A. and Bird, D. W. (2001). The hunting handicap: Costly signaling in human foraging strategies. *Behavioral Ecology and Sociobiology*, 50, 9–19.

- Bochet, O. and Putterman, L. (2009). Not just babble: Opening the black box of communication in a voluntary contribution experiment. *European Economic Review*, 53, 309–26.
- Boyd, R. and Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132, 337–56.
- Boyd, R. and Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–95.
- Boyd, R., Gintis, H., Bowles, S. and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA*, 100, 3531–5.
- Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.
- Brown, J. L. (1983). Cooperation: A biologist's dilemma. *Advances in the Study of Behavior*, 13, 1–37.
- Bshary, R. and Grutter, A. S. (2002). Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, 63, 547–55.
- Bshary, R. and Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, 63, 557–64.
- Burnham, T. and Johnson, D. D. P. (2005). The evolutionary and biological logic of human cooperation. *Analyse and Kritik*, 27, 113–135.
- Burt, A. and Trivers, R. (2006). *Genes in conflict: The biology of selfish genetic elements*. Cambridge, MA: Belknap Press.
- Chagnon, N. A. (1979). Mate competition, favoring close kin, and village fissioning among the Yanomamö Indians. In N. A. Chagnon and W. Irons (eds), *Evolutionary biology and human social behavior* (pp. 86–132). North Scituate, MA: Duxbury Press.
- Cinyabuguma, M., Page, T. and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89, 1421–35.
- Connor, R. C. (1996). Partner preferences in by-product mutualisms and the case of predator inspection in fish. *Animal Behavior*, 51, 451–4.
- Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides and J. Tooby (eds), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.
- Cosmides, L. and Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: John Wiley & Sons, Ltd.
- Daly, M. and Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine de Gruyter.
- Darwin, C. (1859/1958). *The origin of species*. New York: Mentor.
- Darwin, C. (1871/1981). *The descent of man, and selection in relation to sex*. Princeton, NJ: Princeton University Press.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Dugatkin, L. A. (1996). Tit for tat: By-product mutualism and predator inspection: A reply to Connor. *Animal Behavior*, 51, 455–7.
- Ehrhart, K. and Keser, C. (1999). *Mobility and cooperation: On the run*. Working Paper 99s-24. Montreal: CIRANO.
- Ertan, A., Page, T. and Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, 53, 495–511.
- Fehr, E. (2004). Don't lose your reputation. *Nature*, 432, 449–50.
- Fehr, E., Fischbacher, U. and Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, 13, 1–25.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–40.
- Fischbacher, U., Gächter, S. and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public good experiment. *Economic Letters*, 31, 397–404.

- Fletcher, J. A. and Doebeli, M. (2006). How altruism evolves: Assortment and synergy. *Journal of Evolutionary Biology*, 19, 1389–93.
- Foster, K. R., Wenseleers, T. and Ratnieks, F. L. W. (2001). Spite: Hamilton's unproven theory. *Annales Zoologici Fennica*, 38, 229–38.
- Frank, S. A. (1996). Models of parasite virulence. *Quarterly Review of Biology*, 71, 37–78.
- Galdikas, B. M. F. and Teleki, G. (1981). Variations in subsistence activities of male and female pongids: New perspectives on the origins of human labor divisions. *Current Anthropology*, 22, 241–56.
- Gardner, A. and Grafen, A. (2009). Capturing the superorganism: A formal theory of group adaptation. *Journal of Evolutionary Biology*, 22, 659–71.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–79.
- Gintis, H., Smith, E. A. and Bowles, S. (2001). Cooperation and costly signaling. *Journal of Theoretical Biology*, 213, 103–19.
- Gintis, H., Bowles, S., Boyd, R. and Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153–72.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161–78.
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144, 517–46.
- Guarner, F. and Malagelada, J. R. (2003). Gut flora in health and disease. *Lancet*, 361, 512–519.
- Gurven, M. D., Allen-Arave, W., Hill, K. and Hurtado, A. M. (2000). It's a wonderful life: Signaling generosity among the Aché of Paraguay. *Evolution and Human Behavior*, 21, 263–82.
- Hagen, E. H. and Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, 69, 339–48.
- Hain, T. J. A. and Neff, B. D. (2007). Multiple paternity and kin recognition mechanisms in a guppy population. *Molecular Ecology*, 16, 3938–46.
- Hamilton, W. D. (1963). The evolution of altruistic behavior. *American Naturalist*, 97, 354–6.
- Hamilton, W. D. (1964). The genetical evolution of social behavior, I–II. *Journal of Theoretical Biology*, 7, 1–52.
- Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, 228, 1218–1220.
- Hamilton, W. D. (1975). Innate social aptitudes in man: An approach from evolutionary genetics. In R. Fox (ed.), *Biosocial anthropology* (pp. 133–55). London: Malaby Press.
- Hardin, G. J. (1968). The tragedy of the commons. *Science*, 162, 1243–8.
- Hardy, C. and van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32, 1402–13.
- Hauert, C., Traulsen, A., Brandt, H. et al. (2007). Via freedom to coercion: The emergence of costly punishment. *Science*, 316, 1905–7.
- Hauert, C., Traulsen, A., Brandt, H. et al. (2008). Public goods with punishment and abstaining in finite and infinite populations. *Biological Theory*, 3, 114–122.
- Helfenstern, F., Wagner, R. H., Danchin, E. and Rossi, J. M. (2003). Functions of courtship feeding in black-legged kittiwakes: Natural and sexual selection. *Animal Behaviour*, 65, 1027–33.
- Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*, 53, 3–35.
- Henrich, J., Boyd, R., Bowles, S. et al. (2005). 'Economic Man' in cross-cultural perspective: Ethnography and experiments from 15 small-scale societies. *Behavioral and Brain Sciences*, 28, 795–816, 838–55.

- Humphrey, N. (1997). Varieties of altruism and the common ground between them. *Social Research*, 64, 199–209.
- Imhof, L. A. and Nowak, M. A. (in press). Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B*.
- Iredale, W., van Vugt, M. and Dunbar, R. (2008). Showing off in humans: Male generosity as a mating signal. *Evolutionary Psychology*, 6, 386–92.
- Johnson, D. D. P., Price, M. E. and Takezawa, M. (2008). Renaissance of the individual: Reciprocity, positive assortment, and the puzzle of human cooperation. In C. Crawford and D. Krebs (eds.), *Foundations of evolutionary psychology* (pp. 331–52). New York: Lawrence Erlbaum Associates.
- Johnson, T., Dawes, C. T., Fowler, J. H. *et al.* (2009). The role of egalitarian motives in altruistic punishment. *Economics Letters*, 102, 192–4.
- Kalishov, A., Zahavi, A. and Zahavi, A. (2005). Allofeeding in Arabian babblers (*Turdoides squamiceps*). *Journal of Ornithology*, 146, 141–50.
- Keller, L. and Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, 394, 573–5.
- Kelly, R. (1995). *The foraging spectrum: Diversity in hunter-gatherer lifeways*. Washington: Smithsonian Institution Press.
- Kiyonari, T. and Barclay, P. (2008). Free-riding may be thwarted by second-order rewards rather than punishment. *Journal of Personality and Social Psychology*, 95, 826–42.
- Kiyonari, T., van Veelen, M. and Yamagishi, T. (2008). *Can second-order punishment solve the puzzle of human cooperation in one-shot games?* Manuscript submitted for publication.
- Krams, I., Krama, T., Igaune, K. and Mänd, R. (2008). Experimental evidence of reciprocal altruism in the pied flycatcher. *Behavioral Ecology and Sociobiology*, 62, 599–605.
- Kurzban, R., DeScioli, P. and O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75–84.
- Kurzban, R. and Houser, D. (2005). An experimental investigation of cooperative types in human groups: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102, 1803–7.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel and A. E. Roth (eds), *The handbook of experimental economics* (pp. 111–194). Princeton, MA: Princeton University Press.
- Leimar, O. and Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B*, 268, 2495–2501.
- Lieberman, D., Tooby, J. and Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445, 727–31.
- Margulis, L. (1970). *Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian Earth*. New Haven, CT: Yale University Press.
- Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, 201, 1145–7.
- Maynard Smith, J. and Szathmáry, E. (1995). *The major transitions in evolution*. New York: W. H. Freeman/Spektrum.
- Milinski, M. (1987). Tit for tat and the evolution of cooperation in sticklebacks. *Nature*, 325, 433–5.
- Milinski, M. (1990). No alternative to tit-for-tat cooperation in sticklebacks. *Animal Behavior*, 39, 989–91.
- Moghadam, A. (2003). Palestinian suicide terrorism in the Second Intifada: Motivations and organizational aspects. *Studies in Conflict and Terrorism*, 26, 65–92.

- Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, 29, 242–8.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92, 91–112.
- Nowak, M. A. and Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194, 561–74.
- Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291–8.
- O’Gorman, R. O., Henrich, J. and van Vugt, M. (2008). Constraining free-riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B*, 276, 323–9.
- Ohtsuki, H., Iwasa, Y. and Nowak, M. A. (2009). Indirect reciprocity provides only a narrow margin for efficiency for costly punishment. *Nature*, 457, 79–82.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14, 137–58.
- Ostrom, E., Walker, J. and Gardner, R. (1992). Covenants with and without a sword: Self governance is possible. *American Political Science Review*, 86, 404–17.
- Page, T., Putterman, L. and Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal*, 115, 1032–53.
- Panchanathan, K. and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432, 499–502.
- Piazza, J. and Bering, J. M. (2008). The effects of perceived anonymity on altruistic punishment. *Evolutionary Psychology*, 6, 487–501.
- Pizzari, T and Foster, K. R. (2008). Sperm sociality: Cooperation, altruism, and spite. *PLoS Biology*, 6, e130.
- Price, G. R. (1970). Selection and covariance. *Nature*, 227, 520–521.
- Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics*, 35, 485–90.
- Price, M. E. (2003). Pro-community altruism and social status in a Shuar village. *Human Nature*, 14, 191–208.
- Price, M. E. (2005). Punitive sentiment among the Shuar and in industrialized societies: Cross-cultural similarities. *Evolution and Human Behavior*, 26, 279–87.
- Price, M. E. (2006). Monitoring, reputation and ‘greenbeard’ reciprocity in a Shuar work team. *Journal of Organizational Behavior*, 27, 201–19.
- Price, M. E. (2008). The resurrection of group selection as a theory of human cooperation. *Social Justice Research*, 21, 228–40.
- Price, M. E., Cosmides, L. and Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior*, 23, 203–31.
- Queller, D. C. (1985). Kinship, reciprocity, and synergism in the evolution of social behavior. *Nature*, 318, 366–7.
- Queller, D. C., Ponte E., Bozzaro, S. and Strassmann, J. E. (2003). Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science*, 299, 105–6.
- Rainey, P. B. and Rainey, K. (2003). Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425, 72–4.
- Rand, D. G., Ohtsuki, H. and Nowak, M. A. (2009). Direct reciprocity with costly punishment: Generous tit-for-tat prevails. *Journal of Theoretical Biology*, 256, 45–57.

- Rockenbach, B. and Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature*, 444, 718–723.
- Rothstein, S. I. (1980). Reciprocal altruism and kin selection are not clearly separable phenomena. *Journal of Theoretical Biology*, 87, 255–61.
- Rushton, J. P. (1989). Genetic similarity, mate choice, and group selection. *Behavioral and Brain Sciences*, 12, 503–18.
- Russell A. F. and Wright, J. (2008). Avian mobbing: Byproduct mutualism not reciprocal altruism. *Trends in Ecology and Evolution*, 24, 3–5.
- Semmann, D., Krambeck, H. J. and Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, 56, 248–52.
- Semmann, D., Krambeck, H. J. and Milinski, M. (2005). Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology*, 57, 611–616.
- Sheldon, K. M., Sheldon, M. S. and Osbaldiston, R. (2000). Prosocial values and group-assortation within an N-person prisoner's dilemma. *Human Nature*, 11, 387–404.
- Sherman, P. W., Reeve, H. K. and Pfennig, D. W. (1997). Recognition systems. In J. R. Krebs and N. B. Davies (eds), *Behavioural ecology: An evolutionary approach* (pp. 69–96). Oxford: Oxford University Press.
- Silk, J. B. (2005). The evolution of cooperation in primate groups. In H. Gintis, S. Bowles, R. Boyd and E. Fehr (eds), *Moral sentiments and material interests* (pp. 43–73). Cambridge, MA: MIT Press.
- Simpson, B. and Willer, R. (2008). Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly*, 71, 37–52.
- Sinervo B., Chaine A., Clobert J. *et al.* (2006). Self-recognition, color signals, and cycles of green-beard mutualism and altruism. *Proceedings of the National Academy of Science USA*, 103, 7372–7.
- Smith, E. A. and Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior*, 21, 245–61.
- Smukalla S., Caldara M., Pochet N. *et al.* (2008). FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell*, 135, 726–37.
- Sober, E. and Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Sosis, R. and Alcorta, C. S. (2008). Militants and martyrs: Evolutionary perspective on religion and terrorism. In R. D. Sagarin and T. Taylor (eds), *Natural security* (pp. 105–25). Berkeley, CA: University of California Press.
- Stanca, L. (2009). Measuring indirect reciprocity: Whose back do we scratch? *Journal of Economic Psychology*, 30, 190–202.
- Stanford, C. B., Wallis, J., Mpongo, E. and Goodall, J. (1994). Hunting decisions in wild chimpanzees. *Behaviour*, 131, 1–18.
- Summers, K. and Crespi, B. (2005). Cadherins in maternal-foetal interactions: Red queen with a green beard? *Proceedings of the Royal Society B*, 272, 643–9.
- Symons, D. (1990). Adaptiveness and adaptation. *Ethology and Sociobiology*, 11, 427–44.
- Teleki, G. (1973). *The predatory behavior of wild chimpanzees*. Lewisburg, PA: Bucknell University Press.
- Tooby, J. and Cosmides, L. (1989). Kin selection, genic selection, and information-dependent strategies. *Behavioral and Brain Sciences*, 12, 542–4.
- Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides and J. Tooby (eds), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

- Tooby, J. and Cosmides, L. (1996). Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith and R. I. M. Dunbar (eds), *Evolution of social behaviour patterns in primates and man. Proceedings of the British Academy*, 88, 119–143.
- Tooby J., Cosmides L. and Price M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics*, 27, 103–29.
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Trivers, R. (2004). Mutual benefits at all levels of life. *Science*, 304, 964–5.
- Tutin, C. E. G. (1979). Mating patterns and reproductive strategies in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). *Behavioral Ecology and Sociobiology*, 6, 29–38.
- van Vugt, M. and Spisak, B. R. (2008). Sex differences in leadership emergence during competitions within and between groups. *Psychological Science*, 19, 854–8.
- Wakano, J. Y., Nowak, M. A. and Hauert, C. (2009). Spatial dynamics of ecological public goods. *Proceedings of the National Academy of Sciences, USA*, 106, 7810–914.
- Wedekind, C. and Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, 288, 850–852.
- West, S. A., Griffin, A. S. and Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity, and group selection. *Journal of Evolutionary Biology*, 20, 415–432.
- Westermarck, E. A. (1891/1921). *The history of human marriage*, (fifth edition). London: Macmillan.
- Wiggins, D. A. and Morris, R. D. (1986). Criteria for female choice of mates: Courtship feeding and parental care in the common tern. *American Naturalist*, 128, 126–9.
- Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, 308, 181–4.
- Wilkinson, G. S. (1988). Reciprocal altruism in bats and other mammals. *Ethology and Sociobiology*, 9, 85–100.
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.
- Williams, G. C. and Williams, D. C. (1957). Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution*, 11, 32–9.
- Wilson, D. S. (2002). *Darwin's cathedral*. Chicago: University of Chicago Press.
- Wilson, D. S. (2007). Social semantics: Toward a genuine pluralism in the study of social behaviour. *Journal of Evolutionary Biology*, 21, 368–73.
- Wilson, D. S. and Dugatkin, L. A. (1997). Group selection and assortative interactions. *American Naturalist*, 149, 336–51.
- Wilson, D. S. and Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, 17, 585–654.
- Wilson, D. S. and Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology*, 82, 327–48.
- Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: Belknap Press.
- Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. Edinburgh: Oliver and Boyd.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.
- Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, 53, 205–14.
- Zahavi, A. (1977). The cost of honesty (Further remarks on the handicap principle). *Journal of Theoretical Biology*, 67, 603–5.