Review Article

Sixteen common misconceptions about the evolution of cooperation in humans

Stuart A. West*, Claire El Mouden, Andy Gardner

Department of Zoology, Oxford University, OX1 3PS Oxford, United Kingdom

Initial receipt 7 September 2009; final revision received 2 August 2010

Abstract

The occurrence of cooperation poses a problem for the biological and social sciences. However, many aspects of the biological and social science literatures on this subject have developed relatively independently, with a lack of interaction. This has led to a number of misunderstandings with regard to how natural selection operates and the conditions under which cooperation can be favoured. Our aim here is to provide an accessible overview of social evolution theory and the evolutionary work on cooperation, emphasising common misconceptions. © 2010 Elsevier Inc. All rights reserved.

Keywords: Altruism; Fitness; Inclusive fitness; Reciprocity

1. Introduction

One of the greatest problems for the biological and social sciences is to explain social behaviours such as cooperation (Darwin, 1871; Hamilton, 1996). In the biological sciences, the problem ranges from explaining cooperative helping behaviours in organisms such as bacteria or birds to the evolution of complex social insect societies (Sachs et al., 2004; West et al., 2007a). In the social sciences, the problem ranges from explaining human morality and aspects of our underlying psychology to the emergence of our institutions and societies (Binmore, 2005b; Gintis et al., 2005a; Nettle, 2009). In principle, Darwin's theory of natural selection provides a general framework that has the potential to unite aspects of research across these very different areas (Darwin, 1871).

However, there is relatively poor agreement between the social and biological sciences over the underlying evolutionary theory. Our understanding of social evolution theory has advanced hugely over the last 45 years, providing a unified framework that can be applied to all organisms, from microbes to vertebrates (see Section 2). Unfortunately, these advances have been communicated poorly to the social

sciences. Consequently, in many cases, the evolutionary theory being applied in the social sciences is based on secondary sources that were aimed at non-specialists (e.g. Dawkins, 1976; Wilson, 1975b), some of which contain fundamental errors (Grafen, 1982; Dawkins, 1979) and, do not reflect the current state of the field. At the same time, evolutionary biologists have generally remained unaware of many important developments in the social sciences, such as the vast theoretical literature on reciprocity (Binmore, 1998). These issues have led to many sources of confusion, such as the reinvention of old problems, the continuation of long-finished debates, and very different explanations being given to the same empirical observations or theoretical predictions.

Our overall aim in this article is to provide an overview of the evolutionary study of cooperation in a way that is accessible across disciplines, emphasising common misconceptions. In the first part of our study (Sections 2–5), we provide a brief summary of the relevant aspects of evolutionary theory. Specifically, we summarise the modern interpretation of Darwin's theory of natural selection (Section 2), the evolutionary classification of social traits such as altruism (Section 3), the problem of cooperation (Section 4) and the different ways in which the problem of cooperation can be solved (Section 5). We include a number of biological examples in Section 5, as this helps in the elucidation of general theoretical principles. Sections 2–5 could be skipped by readers familiar with the evolutionary

* Corresponding author.

E-mail address: stuart.west@zoo.ox.ac.uk (S.A. West).

literature. In the second part of the article, we discuss a number of common misconceptions and sources of confusion, concerning social theory and the problem of cooperation (Section 6). We hope that our purpose in this section does not come across as negative, as our aim is to facilitate progress at the interface of the biological and social sciences. Finally, in the third part of the article, we focus on humans, discussing why they cooperate and whether they are special (Section 7).

2. Evolutionary theory

2.1. Adaptation and natural selection

The cardinal problem for evolutionary biology is to explain adaptation (Leigh, 1971; Maynard Smith, 1995). The problem of adaptation is the need to explain the empirical fact that organisms appear designed (Paley, 1802). Within this problem, there are two distinct issues: process and purpose (Gardner, 2009). First, by what process (dynamics) does biological adaptation arise? Second, what is the purpose of adaptation: what is it that organisms appear designed to do?

Darwin's theory of natural selection explains both the process and the purpose of adaptation (see glossary in Table 1; Gardner, 2009). The process of adaptation occurs via the action of natural selection, which is driven by the differential reproductive success of individual organisms. Those heritable characters that are associated with greater reproductive success tend to accumulate in natural populations. Thus, Darwin explained the purpose of adaptation: he argued that evolved characters will appear designed as if to maximize the individual's reproductive success. This is analogous to the idea in economics that individuals should be self-regarding utility maximizers—in both cases, it is not required that individuals are consciously striving to maximize their fitness or utility, only that selection will have led to individuals that do so (Darwin, 1859; Friedman, 1953).

These ideas were later formalised in mathematical terms by Fisher (Fisher, 1930, 1941), who used population genetics theory to describe natural selection in terms of changes in gene frequencies. Specifically, Fisher showed that genes that are associated with greater individual fitness are predicted to increase in frequency; hence, natural selection acts to increase the mean fitness of individuals in a population. Fisher interpreted this result, which he termed *the fundamental theorem of natural selection* (see Supplementary material), as proof that organisms will appear increasingly designed so as to maximize their Darwinian fitness. At the time, the work of Fisher and others (Haldane, 1932; Wright, 1931) was celebrated for uniting Darwinism with Mendelian genetics, showing that they were not competing alternative explanations for evolution (Provine, 2001). However, Fisher's work also formalised both the process and the purpose of adaptation (Grafen 2002). The

Table 1
Glossary

Term	Definition
Actor	The focal individual performing a behaviour.
Adaptation	A trait that enhances fitness and that arose historically as a result of natural selection for its current role (Rose & Lauder, 1996). The problem of adaptation is the need to explain the empirical fact that organisms looked designed (Gardner, 2009).
Altruism	A behaviour that is costly to the actor and beneficial to the recipient or recipients. Costs and benefits are defined on the basis of the lifetime direct fitness consequences of a behaviour (Hamilton, 1964).
Cooperation	A behaviour that provides a benefit to another individual (recipient), and the evolution of which has been dependent on its beneficial effect for the recipient (West et al., 2007a).
Direct fitness	The component of fitness gained through the impact of an individual's behaviour on the production of its own offspring; the component of personal fitness due to one's own behaviour.
Inclusive fitness	The effect of one individual's actions on everybody's production of offspring, weighted by the relatedness; the sum of direct and indirect fitness; the quantity maximised by Darwinian individuals (Grafen, 2006a; Hamilton, 1964).
Indirect fitness	The component of fitness gained from aiding related individuals.
Kin selection	Process by which traits are favoured because of their effects on the fitness of related individuals; the way in which natural selection may be separated into direct and indirect components.
Neighbour-modulated fitness	The personal fitness of an individual, which may be dependent upon the behaviours of social partners.
Mutual benefit	A behaviour that is beneficial to both the actor and the recipient (West et al., 2007a).
Personal fitness	An individual's number of offspring, surviving to adulthood. In a class-structured population, each offspring is weighted by their reproductive value.
Recipient	An individual who is affected by the behaviour of the focal actor.
Relatedness	A measure of the genetic similarity of two individuals, relative to the average; the least-squares linear regression of the recipient's genetic breeding value for a trait on the breeding value of the actor (Box 1; Grafen, 1985; Hamilton, 1970).
Reproductive value	The expected, relative genetic contribution of an individual to generations in the distant future; the relative probability that a gene drawn at random from a generation in the distant future will trace back to the focal individual in the present generation (Fisher, 1930; Grafen, 2006b).
Selfishness	A behaviour which is beneficial to the actor and costly to the recipient.
Social behaviours	Behaviours which have a fitness consequence for both the individual that performs the behaviour (actor) and another individual (recipient).
Spite	A behaviour that is costly to both the actor and the recipient (Hamilton, 1970).

process is that natural selection leads to an increase in the frequency of genes associated with greater fitness. The purpose is that natural selection will lead to organisms which appear designed so as to maximize their individual fitness.

Since Darwin, the only fundamental change in our understanding of adaptation has been Hamilton's development of inclusive fitness theory (Hamilton, 1964). The traditional Darwinian view struggled to explain many cooperative social behaviours, with the most famous example being the sterile worker caste in eusocial insect species, the ants, bees, wasps, and termites. Fisher (1930) realised that genes can spread not only through their impact on their own direct transmission (direct fitness) but also through their impact on the transmission of copies of the same allele in other individuals (indirect fitness; see also Darwin, 1859, pp. 257–259), but he explicitly chose to neglect the latter effects in his derivation of the fundamental theorem. Hamilton (1964) incorporated indirect fitness effects into a genetical theory of social evolution and showed that the characters favoured by natural selection are those which improve the individual's "inclusive fitness," which is the sum of its direct and indirect fitness. Another way of thinking about this is that inclusive fitness represents the components of reproductive success of the actor and their social partners over which the actor has control (Fig. 1). The easiest and most common way in which indirect fitness benefits can occur is through helping close relatives, in which case genes are identical by descent (i.e., from a common ancestor), and so, this process is often referred to as "kin selection" (Maynard Smith, 1964).
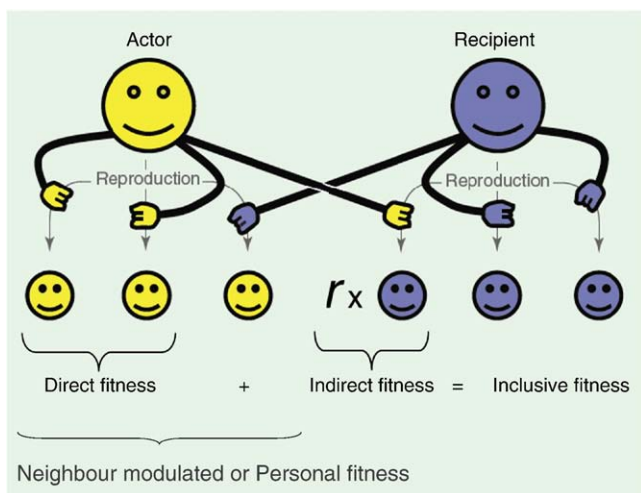

Fig. 1. Inclusive fitness is the sum of direct and indirect fitness (Hamilton, 1964). Social behaviours affect the reproductive success of self and others. The impact of the actor's behaviour (yellow hands) on its reproductive success (yellow offspring) is the direct fitness effect. The impact of the actor's behaviour (yellow hands) on the reproductive success of social partners (blue offspring), weighted by the relatedness of the actor to the recipient, is the indirect fitness effect. In particular, inclusive fitness does not include all of the reproductive success of relatives (blue offspring), only that which is due to the behaviour of the actor (yellow hands). Also, inclusive fitness does not include all of the reproductive success of the actor (yellow offspring), only that which is due to its own behaviour (yellow hands; adapted from West et al., 2007a). A key feature of inclusive fitness is that, as defined, it describes the components of reproductive success which an actor can influence, and therefore which they could be appearing to maximise.

Inclusive fitness is not just a special case for interactions between relatives. It is our modern interpretation of Darwinian fitness in its most general form, explaining both the process and purpose of adaptation (Grafen, 2007b, 2009). The process is that genes or traits which lead to an increase in inclusive fitness will be favoured, and that this increase can occur via direct or indirect routes. The purpose is that individuals should appear as if they have been designed to maximize their inclusive fitness. Grafen (1999, 2002, 2006a, 2007b) has formalised this link between the process and purpose of adaptation, by showing the mathematical equivalence between the dynamics of gene frequency change and the purpose represented by an optimisation program which uses an "individual as maximising agent" (IMA) analogy. This emphasises that inclusive fitness is not just an accounting method, it is the component of reproductive success an organism can influence and what organisms should appear to be maximising.

2.2. Uses and multiple methods

The idea that organisms can be viewed as maximizing agents has proven incredibly useful. This is because inclusive fitness theory takes the dynamics of gene frequency change (the gold standard of evolutionary theory) and turns them into predictions about how individuals should behave (which can be tested with relative ease). The use of this approach in explaining a vast number of traits across a range of organisms can be seen in any animal behaviour or evolutionary ecology textbook (Alcock, 2005; Krebs & Davies, 1993; Westneat & Fox, 2010). Some of the most successful areas include sex allocation (West, 2009), policing and conflict resolution (Ratnieks et al., 2006), cooperation (this paper), kin discrimination (Griffin & West, 2003; Rousset & Roze, 2007), parasite virulence (Frank, 1996b), parent-offspring conflict (Trivers, 1974), sibling conflict (Mock & Parker, 1997), selfish genetic elements (Burt & Trivers, 2006), cannibalism (Pfennig et al., 1999), dispersal (Hamilton & May, 1977), alarm calls (Sherman, 1977) and genomic imprinting (Haig, 2002).

The success of Maynard Smith's (1982) evolutionarily stable strategy (ESS) approach is also because it makes an IMA analogy and, hence, predicts the behaviour of individuals. Most ESS models are not concerned with social behaviours so can assume that indirect fitness effects are unimportant meaning that individuals should behave so as to maximize their personal fitness (Maynard Smith & Price, 1973). This is a special case of the more general inclusive fitness result, and has been formally justified with population genetics (Fisher, 1930; Grafen, 1999, 2002). Empirical success stories in this area include research on foraging, competing for resources and the evolution of mating systems (Krebs & Davies, 1993).

Inclusive fitness theory has well-developed links with all the other areas of evolutionary theory, especially quantitative and population genetics (Frank, 1998; Gardner et al., 2007a;

Grafen, 2006a; Queller, 1992a; Rousset, 2004; Taylor, 1990, 1996; Taylor & Frank, 1996; Wolf et al., 1999). As Hamilton (1964) originally showed, an advantage of inclusive fitness theory is that it can be applied at the genetic or phenotypic level [contra O'Gorman et al., 2008; Sober & Wilson, 1998]. Put another way, it is a genetic theory to explain individual level adaptations. Modern techniques for the development of inclusive fitness theory, termed the *direct* or *neighbour-modulated* fitness method, provide very general, powerful and simple methods for analysing the evolution of all forms of social behaviour (Frank, 1997, 1998; Rousset, 2004; Taylor, 1996; Taylor & Frank, 1996; Taylor et al., 2007b). Importantly, these methods allow the biology to lead the maths, rather than forcing the biology to fit the assumptions of stylized games such as the Prisoner's dilemma (Brown, 2001; West et al., 2007a). An introduction to the mathematics and methods of kin selection theory is provided elsewhere (Gardner et al., Submitted).

While the theoretical overview that we have given above is the framework within which the majority of evolutionary biologists work, it is not accepted by all researchers in the discipline. There are two issues here. First, of course we are not suggesting that inclusive fitness is the only way to model social evolution. A variety of methods exist, which each have pros and cons. Many researchers mix methods, by using the neighbour-modulated fitness method to construct models, which they then interpret with inclusive fitness theory (Taylor & Frank, 1996). In some cases, other methods are more useful. For example, if the co-evolutionary dynamics between traits is key, as is the case with punishment of non-cooperators, then multi-locus population genetic methods offer many advantages (Gardner et al., 2007a). Multi-level selection theory offers another method-ology, although this tends to be used relatively little, because it: (a) can can be hard or impossible to incorporate many important biological complexities (Queller, 2004), especially those that arise when populations are structured into classes (e.g., sexes, or age groups; Frank, 1998); (b) seems to lead to many sources of confusion (Misconceptions 9–13).

A general point with all these alternative methods is that they analyse the dynamics of natural selection differently and, so, do not constitute competing hypotheses as to how adaptation occurs or what it is for. Whatever way you do the maths, this does not change that organisms are predicted to maximize inclusive fitness (Gardner & Grafen, 2009; Grafen, 2006a, 2007b; Hamilton, 1975). The IMA analogy does not work with alternatives such as neighbour-modulated fitness or group fitness, because an individual cannot completely control these measures of fitness, except in special cases (Hamilton, 1964). In contrast, inclusive fitness looks at natural selection from the perspective of the elements of fitness over which the individual has control (Fig. 1). Consequently, we favour the inclusive fitness approach because (a) inclusive fitness provides a single theory that describes both the process and purpose of adaptation (no other theory has been shown to do this) and

Table 2
Social behaviours

Effect on actor	Effect on recipient	
	+	−
+	Mutually beneficial	Selfish
−	Altruistic	Spiteful

A Hamiltonian classification scheme for social behaviours that have been selected for by natural selection (West et al., 2007b). These classifications are based on the average consequences of a behaviour, which is what matters for natural selection.

(b) approaches such as multilevel selection, which focus on the process of adaptation, can lead to confusion over the purpose of adaptation (Misconception 12).

Second, it is sometimes assumed that inclusive fitness theory cannot be applied under certain conditions, such as when there is frequency dependence, strong selection (mutations of large effect) or multiplicative fitness effects. However, this is not the case, as such assumptions are not required by inclusive fitness theory (Hamilton, 1970; Queller, 1992c; Gardner et al., Submitted). Instead, it is that naive applications of inclusive fitness theory (especially Hamilton's rule) can lead to mistakes in such circumstances (Frank, 1998; Gardner et al., Submitted, 2007a).

3. Social traits

Within evolutionary biology, social behaviours are defined according to their personal fitness consequences for the actor and recipient. An individual's personal fitness is defined as the number of offspring that she produces that survive to adulthood (Dawkins, 1982; Grafen, 2007b; Hamilton, 1964; Maynard Smith, 1983; also termed *neighbour-modulated* fitness). From an evolutionary point of view, a behaviour (or action) is social if it has fitness consequences for both the individual that performs that behaviour (the actor) and another individual (the recipient).

Hamilton (1964) classified social behaviours according to whether the consequences they entail for the actor and recipient are beneficial (increase personal fitness) or costly (decrease personal fitness) (Table 2). A behaviour which is beneficial to the actor and costly to the recipient (+/−) is selfish, a behaviour which is beneficial to both the actor and the recipient (+/+) is mutually beneficial, a behaviour which is costly to the actor and beneficial to the recipient (−/+) is altruistic, and a behaviour which is costly to both the actor and the recipient (−/−) is spiteful (Hamilton, 1964; Hamilton, 1970; West et al., 2007b).

Social behaviours are defined in this way for two reasons. First, the adaptationist approach provides a formal justifica-tion for the use of intentional language (Grafen, 1999). As described in Section 2, there is a mathematical correspon-dence (isomorphism) between the dynamics of natural selection and the idea that the individual organism is striving

to maximize her lifetime reproductive success. Consequently, whether a behaviour is beneficial or costly is defined on the basis of: (i) the lifetime fitness consequences of the behaviour and (ii) the fitness of individuals relative to the whole population. Alternative evolutionary definitions of terms such as altruism that rely upon only the short-term fitness consequences (e.g., "reciprocal altruism"), or relative to only a fraction of a population (e.g., the local group, as in the group selection literature) lack formal justification, because there is no corresponding maximizing agent view that supports them.

Second, these intentional terms do not provide a superficial gloss, but are defined in ways that convey important information about gene frequency dynamics. In particular, altruistic and spiteful behaviours could not be explained by the Darwinian view, formalized by Fisher (1930), that individuals strive to maximize their personal fitness and, hence, required consideration of indirect fitness consequences (Hamilton, 1964). It is for these two reasons that Hamilton's definitions have proven so useful in fields such as animal behaviour (Krebs & Davies, 1993).

Altruistic behaviour is favoured when it is directed towards individuals who share a genetic predisposition for altruism (positive relatedness), such as when they share the same genes for altruism. In Misconceptions 1 and 2, we will discuss some of the confusion that has come about through researchers redefining altruism (Hamilton, 1964; West et al., 2007b). Spiteful behaviour is favoured when it is directed towards individuals who are genetically less similar than average (negative relatedness; Hamilton, 1970). One way of conceptualizing this is that the reduced fitness of the recipient reduces competition for other individuals who are more related to the actor than the recipient, i.e., spite is a form of indirect altruism (Gardner et al., 2007b). This requires very restrictive conditions, and there are only a couple of clear examples in the natural world, such as chemical warfare in bacteria and the sterile soldiers in polyembryonic wasps (Gardner et al., 2004; Gardner et al., 2007b). It seems extremely unlikely that these conditions would be met in humans, where apparently spiteful behaviours are more likely to provide a direct benefit and hence be selfish (West & Gardner, 2010).

Cooperation is defined as a behaviour which provides a benefit to another individual (recipient) and which is selected for because of its beneficial effect on the recipient (West et al., 2007b). This definition of cooperation therefore includes all altruistic (−/+) and some mutually beneficial (+/+) behaviours. The latter clause in this definition relates to the standard text book definition of adaptation (Rose & Lauder, 1996), and focuses our attention upon behaviours that are selected for because of their social consequences [see also Scott-Phillips, 2008]. Therefore, we do not include any behaviours that only incidentally produce a one-way byproduct benefit to others. For example, when an elephant produces dung, this is beneficial to the elephant (emptying waste) and also beneficial to a dung beetle that comes along and uses that dung, but it is not useful to call this cooperation. We would only call this cooperation if the elephant were selected to increase its rate of dung production because it gained some benefit from the byproducts of the dung beetle using their dung. More generally, we could refer to "social adaptations" if we wanted to consider social behaviours (Table 2) whose selection has been influenced by the fitness consequences for the recipient.

4. The problem of cooperation

The problem of cooperation is to explain why an individual should carry out a cooperative behaviour that benefits other individuals (Hamilton, 1963, 1964). All else being equal (i.e., in the absence of one of the mechanisms we discuss below), cooperation would reduce the relative fitness of the performer of that behaviour and hence be selected against. To illustrate this, consider a population of unconditional cooperators in which an uncooperative free rider (cheat) arises through mutation or migration. In the absence of any mechanism to punish noncooperators, the free rider benefits from the cooperative behaviour of its social partners, without paying any cost. Consequently, genes for free riding have greater fitness than the genes for cooperation, and the former spread through the population, despite the fact that this will lead to a decline in population fitness. The problem of cooperation is often illustrated within the fields of economics and human morality, as the "tragedy of the commons" (Hardin, 1968) or the prisoner's dilemma (Luce & Raiffa, 1957; Rapoport & Chammah, 1965), but a variety of other games have also been used (Binmore, 1994, 1998, 2005b). Explaining the apparent paradox of cooperation is one of the central problems of biology because almost all of the major evolutionary transitions from replicating molecules to complex animal societies have relied upon solving this problem [see supplementary material; Leigh, 1991; Maynard Smith & Szathmary, 1995].

5. The solutions to the problem of cooperation

As cooperation is in evidence at all levels throughout the natural world, there must be one or many solutions to the problem. In this section, we shall give a brief overview of the potential solutions. Further details can be found in the supplementary material or elsewhere (Lehmann & Keller, 2006; Sachs et al., 2004; West et al., 2007a).

Theoretical explanations for the evolution of cooperation (or any behaviour) are broadly classified into two categories: direct fitness benefits or indirect fitness benefits (Fig. 2). A cooperative behaviour yields direct fitness benefits when the reproductive success of the actor, who performs the cooperative behaviour, is increased. Cooperative behaviours that benefit both the actor and the recipient(s) of the
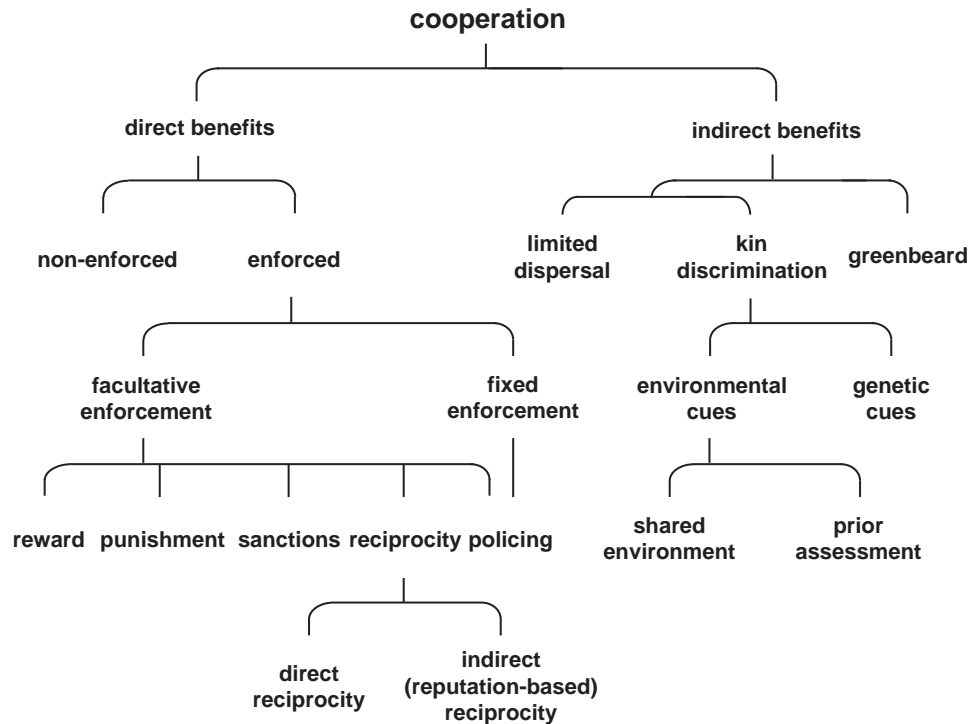
Fig. 2. A classification of the explanations for cooperation. Direct benefits explain mutually beneficial cooperation, whereas indirect benefits explain altruistic cooperation (Hamilton, 1964). Within these two fundamental categories, the different mechanisms can be classified in various ways (Bergmüller et al., 2007; Frank, 2003; Lehmann & Keller, 2006; Sachs et al., 2004; West et al., 2007a). These possibilities are not mutually exclusive—for example, a single act of cooperation could have both direct and indirect fitness benefits, or interactions with relatives could be maintained by both limited dispersal and kin discrimination. Our dividing up of conditional enforcement strategies is for illustration only, as a detailed discussion is beyond the scope of this paper, and provided elsewhere (Bergmüller et al., 2007) (adapted from West et al., 2007a).

behaviour are termed *mutually beneficial*—although they may appear altruistic, they are not (West et al., 2007b; see Misconceptions 1 and 2). These "self-interested" behaviours are readily studied using standard economics models. A cooperative behaviour can be explained by indirect fitness benefits if it is directed towards other individuals who carry genes for cooperation (Hamilton, 1964). As mentioned above, this is usually referred to as "kin selection" (Maynard Smith, 1964) because the simplest and most common way indirect benefits can occur is if cooperation is directed at genealogical relatives (kin), who share genes from a common ancestor (Frank, 1998). By helping a close relative reproduce, an individual is still passing copies of its genes on to the next generation, albeit indirectly. Cooperative behaviours that are costly to the actor and beneficial to the recipient are termed *altruistic* (Hamilton, 1964; West et al., 2007b; see Misconceptions 1 and 2).

Before describing the mechanisms that can explain cooperation, a general point about the differences between evolutionary mechanisms and rational choice theory is that evolutionary mechanisms only explain the average consequences of a behaviour. Therefore, it is quite normal in nature to observe seemingly "irrational" behaviour where an observed cooperative behaviour provides no direct or indirect fitness benefit, such as when: dolphins help an exhausted swimmer, enslaved ants rear the brood of the slave-making species that captured them, or a reed warbler feeds a cuckoo chick that is bigger than itself.

However, these "irrational" or seemingly maladaptive behaviours can be trivially explained by considering the average fitness consequences of such an evolved response. Specifically, the underlying mechanism that leads to such behaviours will have only been selected for if they, on average, provide a direct or indirect fitness benefit. For example, the behaviour of dolphins may be a byproduct of selection for helping within dolphin groups, the rearing behaviour of the enslaved ants is favoured because it is usually directed towards related brood, and the reed warbler feeds the cuckoo because the chicks in its nest will usually be its own offspring. The general point here, that we shall return to in Misconceptions 4 and 14, is that maximisation of fitness does not lead to an expectation for perfect fitness-maximising behaviour in every real-time situation. Behaviour should be studied within the context of the environment in which it was selected for and is being maintained (Herre, 1987). The possibility for such irrational mistakes arises even before we start considering the time that it takes for selection to "catch up" with environmental change (e.g., the time taken for bird species to evolve the ability to spot and avoid cuckoos).

5.1. Kin selection and indirect fitness benefits

Hamilton's inclusive fitness (kin selection) theory explains how altruistic cooperation can be favoured between relatives. This is encapsulated in a pleasingly simple form by Hamilton's (1963, 1964, 1970) rule, which states that a behaviour or trait will be favoured by selection, when $rb-c>0$, where c is the fitness cost to the actor, b is the fitness benefit to the recipient, and r is their genetic relatedness. The coefficient of relatedness (r) is a statistical concept, describing the genetic similarity between two individuals, relative to the average similarity of all individuals in the population (Grafen, 1985; Hamilton, 1970; Box 1). Putting this inequality into words, altruistic cooperation can therefore be favoured if the benefits to the recipient (b), weighted by the genetic relatedness of the recipient to the actor (r), outweigh the costs to the actor (c). All the terms (b, c and r) can be positive or negative, and so Hamilton's rule can be applied to all forms of social behaviour. The generality of Hamilton's rule as a complete description of the dynamics of natural selection is discussed elsewhere (Frank, 1998; Gardner et al., Submitted).

Box 1
What is relatedness?

An individual's phenotype can be separated into its genetic (i.e., heritable) component and its non-genetic (i.e., environmental) component (Fisher, 1918). The former component is termed *the individual's genetic value for the phenotypic character of interest*, and the action of natural selection is formally defined with respect to the change in the average of this quantity (Fisher, 1930). Fisher (1930) separated the action of natural selection into direct fitness effects (impact on personal reproductive success) and indirect fitness effects (impact on the reproductive success of relatives), and Hamilton (1963, 1964, 1970) showed how the latter are mediated by coefficients of relatedness between social partners.

The coefficient of relatedness is defined statistically, as measure of the genetical similarity between social partners, relative to the rest of the population (Grafen, 1985; Hamilton, 1970). Specifically, it is given by:

$$r = \mathrm{cov}(g, g')/\mathrm{cov}(g, g),$$

where g is the genetic value of a focal individual for the phenotypic character of interest, g' is the genetic value of the social partner of this individual, and cov denotes a statistical covariance taken over all individuals in the population (Frank, 1998; Gardner et al., Submitted; Orlove and Wood, 1978). This covariance formulation has a useful interpretation: if we make a scatter plot of the genetic values of social partners (g'; y-axis) against an individual's own genetic value (g; x-axis), then the coefficient of relatedness is equal to the slope of the straight line fitted through these data by means of least-squares regression (Gardner et al., Submitted; Grafen, 1985).

The statistic r can be positive or negative like any statistical correlation but will have a mean value within a population of zero (i.e., if the y-axis of the scatter plot represents the average genetic value of all individuals in the population, this will be a constant and the corresponding regression line will have slope zero). Grafen (1991) defines "the relatedness of a potential actor A to the potential recipient R [as] the extent to which A helping R is like A helping itself." In other words, the important measure of genetic similarity when considering the "r" in Hamilton's inequality, is the genetic similarity between two individuals relative to that between random individuals in the population as a whole. This stresses that it is genetic similarity not kinship *per se* which drives indirect fitness benefits—kinship just happens to be by far the most important reason by which genetic similarity arises (greenbeard genes being the other possibility).

To give a biological example of the statistical definition of relatedness, we use a thought experiment borrowed from Bourke & Franks (1995). Consider a wildebeest carrying a gene that causes it to eat less grass. This would be altruistic, because eating less food is costly to that actor, and leaving more grass to neighbours provides them with a benefit. When would this gene spread through the population? If herds are formed randomly, such that individuals are not surrounded by relatives, then the gene will not spread. This is because the neighbours and hence beneficiaries of the abstaining wildebeest are just a random subset of the population, and so are just as likely to have the altruistic gene as the rest of the population ($r=0$). Consequently, the altruism does not help the altruistic gene increase in frequency. In contrast, if wildebeest live in family groups, then the altruistic gene can spread through the population. This is because the beneficiaries of the altruism would be a non-random section of the population. Specifically, kinship would lead to shared inheritance of genes from ancestors, and hence an above average probability of sharing the altruistic gene with altruistic wildebeest ($r>0$). In this case, the gene for altruistic restraint could be positively selected for, because the extra young that are produced as a result would bear the altruistic gene with an above average frequency, and hence the altruistic gene would increase in frequency in the population.

Explanations for cooperation based on indirect fitness benefits require a sufficiently high genetic relatedness (r) between interacting individuals. Hamilton (1964) suggested two possible mechanisms through which a high relatedness could arise between social partners: kin discrimination and limited dispersal.

5.1.1. Kin discrimination

The first mechanism for generating sufficiently high relatedness to make indirect fitness benefits important is kin discrimination, when an individual can distinguish relatives from non-relatives and preferentially direct aid towards them (nepotism) (Hamilton, 1964). This has been demonstrated in a range of organisms, from fungi to birds to humans (see supplementary material). A clear example is provided by Britain's only cooperative breeding bird, the long-tailed tit, where individuals that fail to breed independently, preferentially help at the nest of close relatives (Russell & Hatchwell, 2001).

Kin discrimination can occur through the use of environmental or genetic cues (Grafen, 1990b). Environmental cues, such as prior association or shared environment, appear to be the most common mechanism of kin discrimination and have been found in organisms ranging from ants to humans (Helanterä & Sundström, 2007; Lieberman et al., 2003), for example, in long-tailed tits, where individuals distinguish between relatives and non-relatives on the basis of vocal contact cues, which are learned from related adults during the nesting period (associative learning) (Sharp et al., 2005). Genetic cues include examples such as the odour produced by scent glands in a mammal (Grafen, 1990b). This has been demonstrated in a range of organisms, including social amoebae, ants, and mammals (Boomsma et al., 2003; Mateo, 2002). There are a number of studies on potential mechanisms for kin discrimination in humans (see supplementary material).

5.1.2. Limited dispersal

The second mechanism for generating sufficiently high relatedness to make indirect fitness benefits important is limited dispersal (Hamilton, 1964). Limited dispersal (population viscosity) can generate high degrees of relatedness between interacting individuals because it will tend to keep relatives together (Hamilton, 1964). In this case, unconditional cooperation directed indiscriminately at other group members (neighbours) could be favoured, because group members (those neighbours) are more likely to be relatives (have a coefficient of relatedness above the population average). This mechanism has the potential to be important in a wide range of cases, from the simplest replicating molecules to humans and other vertebrates, because it does not require the evolution of any potentially costly mechanism of kin discrimination to work (West et al., 2002a). Instead, all that is required is that the level of cooperation evolves in response to the average relatedness between individuals who tend to interact by chance. Direct experimental evidence for a role of limited dispersal has come from observational field data and laboratory experimental evolution on social amoebae and bacteria (Brockhurst et al., 2007; Diggle et al., 2007; Gilbert et al., 2007; Griffin et al., 2004; Kümmerli et al., 2009) and field data on cooperative breeding vertebrates (Cornwallis et al, 2009, 2010).

5.2. Direct fitness benefits

The evolution of cooperation does not only depend upon kin selection and indirect fitness benefits—cooperation can also provide a direct fitness benefit to the cooperating individual (Trivers, 1971). In this case, cooperation is mutually beneficial, not altruistic, and hence would be favoured by "self interested" or "selfish" agents (West et al., 2007b). We divide the direct fitness explanations for cooperation into two categories: byproduct benefits and enforcement (Fig. 1).

5.2.1. By-product benefits

First, the direct benefits of cooperating may flow automatically (passively) as a by-product of helping another individual (Darwin, 1871; Chapter 3). Coordinated foraging in groups appears to be an example of this, where everyone gains an immediate benefit from increased acquisition of food, such as in African wild dogs. A more complicated example, where the benefits can be in the future, rather than immediate, is if cooperation leads to an increase in group size, which increases the fitness of everyone in the group, including the individual who performs the cooperative behaviour (Kokko et al., 2001; Wiley & Rabenold, 1984; Woolfenden, 1975). This process, termed *group augmentation*, has been argued to be important in many cooperatively breeding vertebrates, such as meerkats, where a larger group size can provide a benefit to all the members of the group through an increase in survival, foraging success and the likelihood of winning conflicts with other groups (Clutton-Brock, 2002). Similar arguments can explain cases of helping between unrelated individuals in wasps, where high mortality rates mean that there is an appreciable chance that a subordinate individual can inherit the dominant position and, hence, also inherit any workers that they helped produce (Queller et al., 2000).

5.2.2. Enforcement

The second way in which cooperation can provide direct fitness benefits is if there is some mechanism for enforcing cooperation by rewarding cooperators or punishing cheaters. Trivers (1971) emphasised that cooperation could be favoured in reciprocal interactions with individuals preferentially aiding those that have helped them in the past, as encapsulated by the well known phrase "you scratch my back and I'll scratch yours." This idea dates back to Hume (1739) and had already been analysed in detail in the economics literature before Trivers rediscovered it (reviewed by Aumann, 1981; Aumann & Maschler, 1995; Binmore,

1994, 1998, 2005b, 2007; Fudenberg & Maskin, 1986; Kandori, 1992; Luce & Raiffa, 1957; Mailah & Samuelson, 2006). Reciprocal helping is sometimes referred to as direct reciprocity (help those who help you), to distinguish it from indirect reciprocity, where cooperation is directed at those who are known to cooperate with others, via some method of "image scoring" (help those who help others; Alexander, 1987; Nowak & Sigmund, 1998).

The possibility for cooperation via reciprocity has attracted much enthusiasm, with a huge theoretical literature investigating its possibility. In addition, both direct and indirect reciprocity appear to be important in the evolution and maintenance of cooperation in humans (Alexander, 1987; Binmore, 1994, 1998, 2005b; Gächter & Herrmann, 2009; Henrich & Henrich, 2007; Milinski & Wedekind, 1998; Milinski et al., 2002; Nowak & Sigmund, 2005; Palameta & Brown, 1999; Seabright, 2004; Trivers, 1971; Wedekind & Milinski, 2000). However, reciprocity is thought to be generally unimportant in other organisms, which lack the cognitive capacity of humans (Bergmüller et al., 2007; Clutton-Brock, 2002, 2009; Hammerstein, 2003; Russell & Wright, 2008; Stevens & Hauser, 2004; Whitlock et al., 2007). Even classical text book examples such as blood sharing in vampire bates (Wilkinson, 1984) can be explained more simply without the need for reciprocity by mechanisms such as by-product benefit (Clutton-Brock, 2009). Overall, after 40 years of enthusiasm, there is a lack of a clear example of reciprocity in a non-human species, and so, it is clearly not a major force outside of humans.

In contrast, there is increasing empirical support for a range of other mechanisms that enforce cooperation (see supplementary material). These other possibilities have been termed *punishment*, policing, sanctions, partner switching and partner choice (Bergmüller et al., 2007; Frank, 2003; Sachs et al., 2004; West et al., 2007a). Empirical examples include dominant female meerkats evicting subordinates that try to breed (Young et al., 2006), Superb Fairy Wrens punishing subordinates that don't help (Mulder & Langmore, 1993), cleaner fish clients punishing and avoid cleaners who take a bite of their tissue (Bshary, 2002; Bshary & Grutter, 2002; Bshary & Schäffer, 2002), soybeans cutting off the supply of oxygen to rhizobia bacteria that fail to supply them with Nitrogen (Kiers et al., 2003), a range of pollinator mutualisms where the plants abort overexploited flowers (Goto et al., 2010; Jander & Herre, 2010; Pellmyr & Huth, 1994), and the policing of worker laid eggs in the social insects (Ratnieks et al., 2006).

5.2.3. Why enforce?

While it is clear that enforcing behaviours such as punishment or policing favour cooperation, it is sometimes less obvious why the actual punishment or policing will be favoured by selection. If behaviours such as punishment are costly, then they themselves represent a second-order public good, and so individuals could be selected to avoid the cost of punishment. A possible solution to this is the punishment of individuals who refuse to punish cheats, but this just moves the problem up another level because punishment of nonpunishers represents a third-order public good (Henrich & Boyd, 2001; Sober & Wilson, 1998).

This problem has been solved by a number of theoretical and empirical studies showing how enforcing behaviours can provide a direct or indirect benefit. The simplest way in which punishment could provide a direct fitness advantage is if it led to the termination of interactions with relatively uncooperative individuals (ostracism) and, hence, allowed interactions to be focused on more cooperative individuals (Frank, 2003; Murray, 1985; Schuessler, 1989; West et al., 2002b). This mechanism appears to be operating in cases discussed above such as the cleaner fish, pollinator mutualisms and soybeans. In meerkats, pregnant subordinates will kill other young, even those of the dominant, and so, the dominant increases the survival of her offspring by harassing and evicting pregnant subordinates (Young & Clutton-Brock, 2006). A more complicated possibility is that the punished individuals change their behaviour in response to punishment and are more likely to cooperate with the punisher in future interactions (Clutton-Brock & Parker, 1995). This mechanism is at work in cleaner fish, as described above, and could be important in species such as cooperative breeding vertebrates or humans (Gächter et al., 2008). Enforcement could also be favoured if it provides an indirect fitness benefit (El Mouden et al., 2010; Frank, 1995a; Gardner & West, 2004; Lehmann et al., 2007c; Ratnieks, 1988). An example of this is provided by species of ant, bee, and wasp, where workers selectively cannibalize or "police" eggs laid by workers so that resources can instead be invested into the offspring of the queen, to whom they are more related (Ratnieks et al., 2006).

5.3. Interactions and the origins of cooperation

Although we have emphasised how the different mechanisms favouring cooperation can be divided up, there is considerable scope for interactions between them. In particular, many of the direct fitness benefits can also provide an indirect benefit if directed at relatives. Byproduct mechanisms such as group augmentation involve individuals gaining a direct benefit from larger group size; however, they will also gain an indirect benefit if their group includes relatives, as will often be the case. Enforcement mechanisms can be selected for on the basis of either direct or indirect fitness benefits. Indeed, such mechanisms of enforcement cut across the direct/indirect fitness distinction because they can alter the relative cost and benefit of cooperating—the b and c terms of Hamilton's rule (Lehmann & Keller, 2006).

Different selective forces may be involved in the origin and then subsequent elaboration/maintenance of a trait. In

many cases where there could eventually be a direct fitness benefit to cooperation, it can be hard or impossible for cooperation to spread initially, because to not cooperate (defection) is also an ESS. This is for instance the case with direct reciprocity (Axelrod & Hamilton, 1981), indirect reciprocity (Panchanathan & Boyd, 2004), punishment (Gardner & West, 2004; Henrich & Boyd, 2001), group augmentation (Kokko et al., 2001), and costly signalling (Gintis et al., 2001). In cases where these processes are invoked, it is therefore likely that cooperation initially arose due to factors such as indirect fitness benefits or shared interests, and that only after this do mechanisms such as reciprocity or punishment select for higher levels of cooperation, even when relatedness falls to zero. So, for humans, it may be unnecessary to prove how cooperation can arise de novo in unrelated populations if it originated in a hominid that lived in groups of relatives.

6. Common Misconceptions

In this section, we briefly run through sixteen common misconceptions about social evolution theory, which are summarised in Table 3. There is some overlap and repetition between sections, partly because multiple misconceptions are made in the same areas of research and partly because we wish that each can be read relatively independently. Further misconceptions on the issue of whether and why humans are special are discussed in Section 7.1. The interested reader is also directed towards the "Twelve

Table 3
Sixteen common misconceptions about social evolution theory

Misconception	Reality
1. The various redefinitions of altruism.	Many behaviours that have been described as altruism actually involve a net direct fitness benefit, and so are mutually beneficial, not altruistic. The jargon associated with redefining altruism often obscures the underlying selective forces.
2. Kin selection and reciprocity are the major competing explanations for altruism in biological theory.	In the context of reciprocity, cooperation is not altruistic, and there are many other mechanisms by which cooperation can be favoured due to direct fitness benefits (Fig. 2).
3. Mutually beneficial cooperation is less interesting.	Mechanisms to provide direct fitness benefits to (mutually beneficial) cooperation can often be much more complicated, from both a theoretical and empirical perspective, than indirect benefits, which can arise through relatively simple processes such as limited dispersal or kin discrimination.
4. Proximate and ultimate explanations.	Proximate answers cannot provide a solution to ultimate problems.
5. Kin selection requires kin discrimination.	A sufficiently high relatedness can also arise through limited dispersal.
6. Relatedness is only high between members of the nuclear family.	If there is population structuring (viscous populations or limited dispersal), then relatedness can be relatively high between group members who are not close kin.
7. Kin selection only applies to interactions between relatives and greenbeard genes can explain cooperation in humans.	Indirect fitness benefits can accrue if cooperation is directed towards non-relatives who share the same cooperative gene. Such "greenbeard" mechanisms are unlikely to be important in humans.
8. Greenbeards are a type of costly signaling.	Greenbeards and costly signalling are two different things.
9. Group selection is a formal theory with one meaning.	Group selection is used to mean at least four different things.
10. Group selection can apply in situations when kin selection cannot explain cooperation.	Group selection and kin selection are simply different approaches to describing the same biological process.
11. Kin selection is a subset of group selection.	No group selection model has ever been constructed where the same result cannot be found with kin selection theory. The reverse is not necessarily true.
12. Group selection leads to group adaptations.	Group selection will only lead to group adaptations in the special circumstances where either: (a) the group is composed of genetically identical individuals (clonal groups, $r=1$), or (b) there is complete repression of competition between groups (i.e., no conflict within groups).
13. Most evolutionary biologists view group selection as completely wrong, or that there is some ulterior motive for the lack of attention given to it.	The reason that most evolutionary biologists, both theoretical and empirical, do not use the group selection approach is simply that it is less useful, and if they express negative views, it is because it has generated more confusion than insight.
14. Human cooperation in economic games requires the novel evolutionary force of strong reciprocity.	The simplest explanations for cooperating and punishing in one-shot encounters are individuals making mistakes and/or that it is a byproduct of selection for cooperation in other conditions.
15. The theoretical models on strong reciprocity provide a novel solution to the problem of cooperation, that are outside of the usual inclusive fitness explanations.	The theoretical models of strong reciprocity work upon standard direct and indirect fitness benefits.
16. The claims made in the empirical and the theoretical strong reciprocity literature are compatible.	The work on strong reciprocity can be divided into four areas – what the empirical data show, what it is argued the empirical data show, what the theoretical models show, and what it is argued the theoretical models show. All of these four areas are in disagreement with each other.

misunderstandings of kin selection," of Dawkins (1979), many of which are still pertinent today.

6.1. Kin Selection, Reciprocity, and Altruism

6.1.1. Misconception 1: The various redefinitions of altruism (Baschetti, 2007; Becker, 1974; Bergstrom, 1995, 2002; Bowles, 2006, 2009; Bowles & Gintis, 2004, 2008; Boyd et al., 2003; Fehr & Fischbacher, 2003; Gintis, 2000; Sober & Wilson, 1998; Trivers, 1971; Wilson, 1975a)

In Section 3 we emphasized how terms such as altruism have very specific meanings, that have formal justification and convey useful information. If these terms are misused, or redefined, the result is confusion. This has been a particularly large problem with the term altruism (West et al., 2007b; p. 419–423), which has been redefined in evolutionary models in many ways, including: (a) a decrease in fitness over the short term, so that reciprocity is "reciprocal altruism" (Becker, 1974; Fehr & Fischbacher, 2003; Trivers, 1971); (b) a decrease in the fitness of the focal individual, relative to the other members of its group (relatively costly to individual, relatively beneficial to the group; sometimes termed *weak altruism*) (Baschetti, 2007; Bergstrom, 1995; Bowles, 2006; Bowles & Gintis, 2004; Boyd et al., 2003; Gintis, 2000; Sober & Wilson, 1998; Wilson, 1975a); (c) playing cooperate in a prisoners' dilemma game (Bergstrom, 2002); (d) a failure to harm others (Field, 2001); (f) giving up resources in order to benefit others (Pradel et al., 2009); (g) the mechanism by which one individual is motivated to help others (Axelrod, 1984); (h) the willingness to take mortal risks as a fighter (Bowles, 2009).

The first problem with these redefinitions is that they lack a formal justification to use intentional language from an evolutionary or ultimate perspective. This is because they require the costs and benefits to be defined in different ways and not with respect to lifetime reproductive success. As discussed in Sections 2 and 3, natural selection produces organisms that behave intentionally, as maximizing agents, at the level of lifetime reproductive success.

The second problem is that these redefinitions include scenarios where cooperation could provide a direct fitness benefit, and hence be either mutually beneficial (+/+) or altruistic (−/+). Considering a specific case, Gintis (2000) compared the relative fitness of two different strategies: "self-interested agents" who do not punish or cooperate, and altruistic "strong reciprocators" who cooperate and punish noncooperators. He labels strong reciprocators as altruistic because they "increase the fitness of unrelated individuals at a cost to themselves." However, in this and related models, cooperation is individually costly within the social group but provides a benefit to all the members of the group through mechanisms such as increased productivity or reducing the rate of group extinction (Bowles & Gintis, 2004; Bowles et al., 2003; Boyd et al., 2003; Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003;

Henrich & Boyd, 2001). Consequently, any individual that behaves cooperatively also gains this (direct) benefit, which can outweigh the cost of performing the behaviour (Binmore, 2005b; Burnham & Johnson, 2005; Lehmann et al., 2007c; West et al., 2007b). This leads to the confusing situation where: (a) cooperation can be favoured because it provides a direct benefit to the cooperator because it increases the chance they and the rest of their group survive, but this is defined as altruistic rather than in their self interest (West et al., 2007b); (b) a "selfish agent" (Bowles & Gintis, 2004) can have a lower direct fitness than an altruist.

A general issue here is that redefinitions of altruism obscure the fundamental distinction between when direct or indirect fitness benefits are required to explain the observed cooperation (Dawkins, 1979; Smuts, 1999; West et al., 2007b). This can lead to the situation where a behaviour is described as altruistic but can be explained by direct fitness benefits (i.e., by self-interested or self-regarding behaviours). This also clouds the relation to other research. For example, the models discussed in the above paragraph are related to models of group augmentation (Section 5.2.1), where cooperation has been argued to provide both direct and indirect benefits. An analogous example from the economics literature is the confusion that has arisen from the multiple redefinitions of the term "social capital" (Binmore, 2005b; Manski, 2000).

We appreciate that terms can have different meanings in different fields, such as the motivational definition of altruism in the psychology literature, and we would not like to give the impression that an evolutionary definition is the only valid one. However, in all the cases discussed above, the authors are considering the evolution and maintenance of cooperation or altruism, with reference to the evolutionary literature. An even greater problem is when papers mix up definitions, starting with a statement of how altruism (or spite) poses a problem for evolutionary theory (which is true based an evolutionary definition) but then actually focus on altruistic behaviours as defined by motivational or mechanistic definition and, so, where the evolutionary problem doesn't necessarily apply (Misconception 4; West & Gardner, 2010).

Finally, some confusion over terminology may also have arisen from the Dawkins title "The Selfish Gene" (Dawkins, 1976) because he defined terms at a different level to which had been done before (i.e., the gene rather than the individual). As discussed in Sections 2 and 3, Hamilton's (1964) use of intentional language (Table 2) followed from the idea that individuals should appear as maximizing agents and, hence, defined behaviours such as altruism and selfishness according to their direct consequences for individuals (Grafen, 1999, 2007a). If this same logic is applied to genes, then selection could favour genes that are selfish or altruistic or mutually beneficial or spiteful. However, Dawkins defined genes as selfish not from the perspective of a single copy of a gene found in an

individual, but from the perspective of all copies of that gene. In this case, as selection only favours genes that increase in frequency, it can only favour genes that are selfish (at the level of every copy of that gene) (Burt & Trivers, 2006). This would be analogous to the situation that would arise had Hamilton defined terms such as altruism at the level of the inclusive fitness of the individual, in which case, because natural selection favours traits that lead to an increase in inclusive fitness, these traits would always be defined as selfish (at the level of inclusive fitness). For social scientists in the 1970s, a misconceived view that "selfish genes" referred to an individual's gene copies appeared to support the economists description of individuals as purely "self-interested." As the selfishness axiom was effectively challenged in economics, so it was assumed that evolutionary theory too was unable to explain human sociality. This was the origin of many of the misconceptions and "new" evolutionary explanations for human behaviour we discuss.

6.1.2. Misconception 2: Kin selection and reciprocity are the major competing explanations for altruism in biological theory (e.g., Boyd & Richerson, 2005; Boyd et al., 2003; de Waal, 2008; Fehr & Gächter, 2002; Fehr & Rockenbach, 2003; Fehr & Fischbacher, 2003; Fehr & Rockenbach, 2004; Gintis et al., 2005b; Henrich & Boyd, 2001; Richerson & Boyd, 1999; Schloss, 2002; Silk, 2002)

This is wrong on two counts. First, reciprocity is not altruistic—it provides a direct fitness advantage to cooperating. If an individual does not pay the cost of cooperation in the short term then it will not gain the benefit of cooperation in the long term. Consequently, cooperation is only favoured (between nonrelatives) if it leads to an overall benefit, in which case it is mutually beneficial (+/+). Much of the confusion here is due to the term "reciprocal altruism" of Trivers (1971, 1985) the introduction of which was accompanied by multiple redefinitions of altruism (West et al., 2007b, p. 420). It was for these reasons that Hamilton (1996, p. 263) thought that reciprocal altruism was misnamed, and several authors have used less confusing alternatives such as "reciprocity" or "reciprocal cooperation" (Alexander, 1974; Axelrod & Hamilton, 1981; Binmore, 1994, 1998; West et al., 2007b).

Second, when considering explanations for cooperation, the major competing hypotheses are not kin selection and reciprocity. Reciprocity is only one of the many ways in which cooperation can lead to direct fitness benefits (Fig. 2), and while it appears to be important in humans, it is relatively unimportant in other species. In some cases, this misconception appears to arise from only considering the evolutionary literature up to approximately the late 1970s and, hence, missing the huge advances that have been made since then (sometimes referred to as the "disco problem"). As well as in the papers cited above from the primary literature, Misconception 2 or a close approximation occurs in a scarily large number of undergraduate textbooks.

6.2. Mutually beneficial cooperation

Misconception 3: Mutually beneficial cooperation is less interesting

Misconception 1 illustrated the point that altruism is often redefined so that it will include a particular case of cooperation that is being examined. Furthermore, researchers are often disappointed to discover particular cases fit into the mutually beneficial category (+/+) and are not altruistic (−/+). Indeed, altruism may be redefined so frequently because researchers prefer their research problem to be altruism. This reflects the common feeling that mutually beneficial behaviours are somehow less interesting. We strongly disagree. Indeed, mechanisms to provide direct fitness benefits to cooperation can often be much more complicated, from both a theoretical and empirical perspective, than indirect benefits, which can arise through relatively simple processes such as limited dispersal or kin discrimination. Determining the relative importance of direct and indirect benefits remains a key problem and has long been a major topic of debate in areas such as the evolution of helping in cooperative breeding vertebrates (Clutton-Brock, 2002; Cockburn, 1998; Griffin & West, 2002; Jennions & Macdonald, 1994). A contributing factor here may be the often quoted statement from the sociobiology book of Wilson (1975b, p.31) that: "the central theoretical problem of sociobiology [is]: how can altruism, which by definition reduces personal fitness, possibly evolve by natural selection?" (Becker, 1974). This is misleading because, as emphasized in Section 4, it is actually the evolution of cooperation that is the central problem of sociobiology (see supplementary material).

6.3. Proximate and ultimate explanations

Misconception 4: Proximate explanations provide a solution to the ultimate problem of cooperation.

It is useful to distinguish between ultimate and proximate explanations of traits or behaviours (Mayr, 1961; Tinbergen, 1963). Proximate explanations are concerned with the causal mechanisms underlying a behaviour (how questions). Ultimate explanations are concerned with the fitness consequences of a behaviour (why questions). Evolutionary biology attempts to explain features of an organism from an ultimate perspective—why are organisms the way they are? The key point is that these different methodologies are complementary and not competing alternatives.

The Nobel Prize winner Niko Tinbergen (1963) famously clarified the distinction between ultimate and proximate explanations for animal behaviour, in the most influential paper of his career (Kruuk, 2003); less well known to many biologists is that Niko's brother Jan won the 1969 Nobel memorial prize in Economics. One of Tinbergen's classic studies to illustrate this distinction was on the removal of eggshells from their nests by black-headed gulls. The mechanistic (proximate) explanation for this is that individuals are more likely to remove objects from their nest if

they are white- or egg-coloured, have frilly edges, and if they are feather-light. The evolutionary (ultimate) explanation for this is that it makes aerial predators such as herring gulls less likely to find their brood. These explanations are clearly not competing (each answer cannot provide a solution to the other problem), and a fuller understanding is gained by considering both.

A clear example of the confusion that may be caused by conflating ultimate and proximate factors is provided by work on "strong reciprocity," which is defined proximately but then given as a solution to an ultimate problem (Bowles & Gintis, 2004; Fehr & Gächter, 2002; Fehr & Rockenbach, 2003, 2004; Fehr & Fischbacher, 2003, 2004; Fehr et al., 2002; Gintis et al., 2003). A strong reciprocator has been defined as a combination of "a predisposition to reward others for cooperative, norm-abiding behaviours" and "a propensity to impose sanctions on others for norm violations" (Fehr & Fischbacher, 2003). This is a description of a proximate mechanism. However, it is then given as a solution to an ultimate problem—for example, "Strong reciprocity thus constitutes a powerful incentive for cooperation even in nonrepeated interactions when reputation gains are absent" (Fehr & Fischbacher, 2003), or "cooperation is maintained because many humans have a predisposition to punish those who violate group-beneficial norms" (Bowles & Gintis, 2004).

This is illustrated even more clearly with a discussion of neurological work, where it is suggested that an explanation for the punishment of individuals who do not cooperate is that such punishment leads to "satisfaction" (Fehr & Rockenbach, 2004; Quervain et al., 2004). For example, in two adjoining sentences, Quervain et al. (2004, p. 1254) follow an ultimate question "Why do people punish violators of widely approved norms although they reap no offsetting material benefits themselves?" with a proximate answer "We hypothesize that individuals derive satisfaction from the punishment of norm violators." This does not solve the ultimate problem because it does not answer why evolution should have produced a psychology or nervous system that mechanistically encourages (rewards) such punishment.

This approach mixes up two different questions (how and why, or process and product). Claiming that cooperation is favoured because individuals have a predisposition to cooperate, and punish those that do not, is circular, as it does not explain why individuals should have a predisposition to cooperate and punish in the first place. The proximate question is how is cooperation maintained? The answer to this is a predisposition to cooperate and avoid punishment, i.e., what has been termed *strong reciprocity*. The ultimate question is why is cooperation maintained, or more specifically, why are cooperation and punishment (strong reciprocity) maintained? The possible answer to this is because it provides either a direct and/or an indirect fitness benefit (Gardner & West, 2004). We are not arguing that proximate questions are not interesting, and we appreciate that they are, with good reason, the focus of much human research. Instead, our point is that it is very misleading to mix and match by posing and justifying a problem from an ultimate perspective and then providing a proximate answer.

Similar confusion over proximate and ultimate factors occurs in numerous other places. One example is "social institution" models, where selection for cooperation is increased by "the commonly observed human practices of resource sharing among group members" (Bowles, 2006; Bowles et al., 2003). However, as an institution is a form of cooperation itself, it just provides a proximate answer (cooperation is explained by cooperation) that avoids the ultimate problem of why would the social institution of cooperative resource sharing ever evolve? This question can be addressed with models which assume that mechanisms for repressing competition within groups are potentially costly traits under selection (El Mouden et al., 2010; Frank, 1995a, 1996a, 2003; Leigh, 1971; Ratnieks, 1988). Another example is the suggestion that "adults may support their parents in order to imprint a corresponding behavior pattern on their own children" (Bergstrom, 1996). This is a proximate answer, and does not answer why such imprinting would be favoured. Similar mixing up of proximate and ultimate factors occur in the literature on the evolution of language (Scott-Phillips, 2007), the group selection literature (Smuts, 1999) and at the interface of the primate and human literature (de Waal, 2008).

6.4. Inclusive Fitness, Kin Selection, Relatedness and Greenbeards

There are three related misconceptions about how a significant relatedness and indirect fitness benefits (kin selection) can occur.

6.4.1. Misconception 5: Kin selection requires kin discrimination

In his original papers on inclusive fitness theory, Hamilton pointed out a sufficiently high relatedness to favour altruistic behaviours could accrue in two ways—kin discrimination or limited dispersal (Hamilton, 1964, 1971, 1972, 1975). There is a huge theoretical literature on the possible role of limited dispersal (reviewed by (Platt & Bever, 2009; West et al., 2002a), as well as experimental evolution tests of these models (Diggle et al., 2007; Griffin et al., 2004; Kümmerli et al., 2009). However, despite this, it is still sometimes claimed that kin selection requires kin discrimination (Oates & Wilson, 2001; Silk, 2002). Furthermore, a large number of authors appear to have implicitly or explicitly assumed that kin discrimination is the only mechanism by which altruistic behaviours can be directed towards relatives and have reinvented the role of limited dispersal, usually calling it something else, and claiming that indirect fitness, kin selection or relatedness is not important (Table 4).

Table 4
Some examples of the reinvention of how limited dispersal increases relatedness between interacting individuals, and can hence provide an indirect fitness benefits (kin selection) for cooperation

Suggested explanation for cooperation	Authors suggesting explanation	Authors showing that the suggestion is a reinvention of kin selection via limited dispersal
Games in spatial settings	Nowak et al., 2010	Lehmann & Keller, 2006
Spatial structure	MacLean & Gudelj, 2006; Pfeiffer et al., 2001	Frank, 1998; Frank, 2010
Group or multilevel selection	Nowak, 2006; Traulsen & Nowak, 2006; Wilson, 1975a	Frank, 1986; Grafen, 1984; Lehmann et al., 2007b; Queller, 1992a
Population structure	Killingback et al., 2006	Grafen, 2007c
Network reciprocity via games on graphs	Lieberman et al., 2005; Nowak, 2006	Grafen, 2007a; Grafen & Archetti, 2008; Lehmann et al, 2007a; Taylor et al., 2007a
Strong reciprocity	Bowles & Gintis, 2004; Gintis, 2000	Lehmann & Keller, 2006; Lehmann et al., 2007c

6.4.2. Misconception 6: Relatedness is only high between close family members (Bowles & Gintis, 2004; Boyd & Richerson, 2005; Gintis, 2000)

It is sometimes implicitly assumed in the theoretical literature that relatedness can only be high between close family relatives. One example is the various strong reciprocity theoretical models where it is argued that kin selection is not important (e.g., Bowles & Gintis, 2004; Gintis, 2000), but then limited dispersal is assumed of a form that can lead to a substantial relatedness between interacting individuals (Lehmann et al., 2007c; West et al., 2007b) (see also Misconception 15). Another example, is provided by the claim that group selection is an alternative mechanism that explains cooperation between non-relatives but that it only works when "groups are small and migration infrequent" (Boyd et al., 2005, p. 215), without realising that this is when relatedness is high (see also Misconceptions 9–13).

These conclusions appear to be based on the well-known approximation that relatedness is approximately $r=1/2$ between full siblings, $r=1/4$ between half siblings, $r=1/8$ between cousins, etc. However, these are only approximations for large well-mixed populations, and the formal definition of relatedness is a statistical measure of genetic similarity (Box 1). If there is population structuring with limited migration (viscous populations or limited dispersal), then relatedness between group members can be relatively high because it will tend to increase the genetic similarity between interacting individuals (Hamilton, 1964, 1970, 1971, 1972, 1975). To give a specific example, consider a population split into groups of size 100, and where 1% of individuals disperse from their natal patch before breeding. In this case, the increased genetic similarity that results from population structuring will lead to the average relatedness of group mates being approximately one third (see supplementary material). Hence, the relatedness between first cousins will be more than one third, and not the commonly assumed one eighth. Clear quantitative support for the effects of population structure on relatedness have been provided by experimental evolution studies with bacteria (Brockhurst et al., 2007; Griffin et al., 2004; Kümmerli et al., 2009).

The above discussion for Misconceptions 5 and 6 rest upon the understanding that relatedness is a statistical measure of genetic similarity (Box 1). It is sometimes argued that relatedness was originally a simple measure of genealogical relationship and that evolutionary theoreticians later reinvented it as a more general measure of genetic similarity, either in the 1980s (e.g., by Grafen (1985) or later (e.g., by Lehmann & Keller, 2006). However, this is completely incorrect. In his original papers, Hamilton made clear that what mattered was genetic similarity per se, discussing relatedness in terms of a regression coefficient (Hamilton 1963, p. 355) and possible green beard effects among genealogically unrelated individuals (Hamilton, 1964, p. 24–25). He then went on to formalise this in his 1970 *Nature* paper (Hamilton, 1970, 1975; Michod & Hamilton, 1980), providing the regression definition of relatedness that is at the centre of modern social evolution theory (Frank, 1998; Gardner et al., Submitted; Grafen, 1985, 2006a; Taylor & Frank, 1996). As well as the huge primary literature on this issue, the fact that it is genetic similarity that matters was also made clear in the popularisations of inclusive fitness theory of Dawkins (1976, 1982). Relatedness and inclusive fitness theory have not been reinvented—the modern interpretation is that developed by Hamilton in the 1960's.

Two other points are worth considering here. First, how do empirical biologists approach the concept of relatedness? Is the statistical (regression) definition of relatedness purely a theoretical concept, with empirical biologists using coancestry to measure relatedness in natural populations? No. The most common method by which empirical biologists measure relatedness is to use molecular markers such as microsatellites, and then plug the data from those into programmes such as *Kinship*, which estimates relatedness with the statistical definition (Queller & Goodnight, 1989). The extent to which the statistical measure of relatedness is used by empirical biologists is clear from the fact that the Queller and Goodnight (1989) methods paper has been cited >1200 times (*Web of Science* search, September 2010). Second, it is true that introductory animal behaviour textbooks such as Krebs and Davies

(1993) and Alcock (2005) define relatedness through co-ancestry and not statistically. However, the coancestry definition is a useful approximation for teaching certain age groups of undergraduates. The primary literature needs to build upon and relate to the primary literature, not to introductory textbooks.

6.4.3. Misconception 7: Inclusive fitness only applies to interactions between relatives, and greenbeard genes can explain cooperation in humans (Bergstrom, 1995, 1996, 2002; Bowles & Gintis, 2004, 2008; Boyd & Richerson, 2005; Frank, 1987; Gintis, 2000; Robson, 1990)

This follows on from the previous two misconceptions, and is wrong on three counts. First, as discussed in Section 2, inclusive fitness is a very general encapsulation of evolutionary theory, not a special case; it applies equally well to social and nonsocial characters. Second, as discussed in Misconception 6, relatedness can be high between individuals who are not close family members.

Third, as pointed out by Hamilton in his original formulation of inclusive fitness, indirect fitness benefits can accrue if cooperation is directed towards non-relatives who share the same cooperative gene (Hamilton, 1964; p. 24–25). Dawkins (1976, 1982) illustrated this with a hypothetical example of a gene that causes its bearer to grow a green beard and also to preferentially direct cooperation towards other green-bearded individuals. This mechanism can also occur without a visible tag, for example, if the cooperative gene also causes some effect on habitat preference that leads individuals who carry the gene to settle close together (Hamilton, 1964, 1975). Consequently, although this mechanism is usually termed a *greenbeard*, it more generally represents an assortment mechanism, requiring a single gene—or a number of tightly linked genes (e.g., physically close on a chromosome and so not separated during sexual reproduction by recombination)—that encodes both the cooperative behaviour and causes cooperators to associate (Gardner & West, 2010; Lehmann & Keller, 2006). One way of conceptualising greenbeards is that they are an extreme end point on the genetic kin discrimination continuum, with no recombination between the tag and helping loci (Rousset & Roze, 2007).

Greenbeard genes are likely to be extremely rare in the real world (Gardner & West, 2010; West & Gardner, 2010). The idea of greenbeards was initially developed as a thought experiment to illustrate that what matters for inclusive fitness is genetic similarity at the locus (or loci) being considered, rather than genealogical relationship *per se* (Hamilton, 1964, 1970, 1971, 1975). It was assumed that that greenbeards would be unimportant in the real world because cheaters, which display the green beard or assorting behaviour without also performing the cooperative behaviour, could invade and overrun the population (Dawkins, 1976, 1982). To date, only five examples of possible greenbeard genes have been found in nature, three cooperative and two spiteful, four in

microbes and one in an ant (Gardner & West, 2010). The feasibility of greenbeard genes is greatest in simpler organisms, such as bacteria, where there can be a relatively simple link between genotype and phenotype and, hence, the possibility that a single gene could have the required multiple (pleiotropic) effects.

Models for the evolution of cooperation that rely upon greenbeards are unlikely to be important in humans (Fehr & Fischbacher, 2005a, 2005b; Gardner & West, 2010; Henrich, 2004). This is because the polygenic nature of behaviours would readily allow the evolution of cheats who displayed a tag or performed the assortative behaviour, but did not cooperate. Despite this, two classes of models of cooperation in humans have been proposed which rely upon a greenbeard mechanism, and which are therefore based upon an unlikely and evolutionary unstable assumption. In both cases the assumption of a greenbeard mechanism was implicit and not realised by the original authors. First, it has been suggested that individuals who cooperate differ from individuals who cheat in "some observable characteristic" other than the cooperation phenotype itself (Amann & Yang, 1998; Frank, 1987; Robson, 1990). This represents the original green beard scenario, which is unlikely to work in humans, as described above. Owren & Bachorowski (2001) provide a more specific version of this scenario, where the observable characteristic is smiling and laughter. However, there is no reason to expect genes for cooperative behaviours to be tightly linked to, or the same as genes that control smiling and laughter.

Second, some (but not all) models of "strong reciprocity" assume that helping and punishment are completely linked traits (Bowles & Gintis, 2004; Gintis, 2000). In these strong reciprocity models, the benefit of helping has no influence on selection for strong reciprocity because it is cancelled out by the increased kin competition that is generated by the act of helping (Lehmann et al., 2007c). Instead, strong reciprocity is selected for, because helping acts as a tag of who is carrying the punishment allele, and so, punishment can be directed at individuals who do not carry that allele, reducing competition for individuals who do carry this allele. Consequently, in contrast to the verbal claim that these models are examining the evolution of cooperation (Bowles & Gintis, 2004; Gintis, 2000), they are actually examining the evolution of spiteful greenbeards (Lehmann et al., 2007c)! Furthermore, not only is selection driven by indirect fitness consequences, but the trait is costly to the group — this is the exact opposite of what is claimed verbally in the original papers. The confusion that the can be caused by a such a mismatch between how a model works, and how it is claimed to work, is nicely illustrated by the fact that Fehr & Fischbacher (2005a) cite Gintis (2000) as showing how strong reciprocity can favour cooperation in humans in a paper where the main focus was to argue that greenbeards cannot explain cooperation in humans.

6.4.4. Misconception 9: Greenbeards are a type of costly signaling (Henrich, 2004; Owren & Bachorowski, 2001)

Greenbeards and costly signalling are two different things. As discussed above, the greenbeard mechanism involves a trait and a tag being encoded by the same gene, or tightly linked genes (i.e., genetic linkage prevents lying). In contrast, costly (or honest) signalling is the idea that signalling can be evolutionary stable if the signal is costly and cannot be faked (i.e., lying is too costly; Grafen, 1990a; Spence, 1973). For example, if cooperative behaviours are costly, then cooperation could function as a signal of quality because individuals in better condition would be able to behave more cooperatively (even though, in principle, anyone could perform cooperative behaviours; Gintis et al., 2001). This is further illustrated by considering the smiling and laughing example discussed above (Owren & Bachorowski, 2001). In order for laughing and smiling to be favoured as a signal of cooperative behaviour via a greenbeard mechanism, we would require that laughing and smiling be controlled by the same gene(s) (or tightly linked genes) as cooperative behaviours. In contrast, for smiling and laughing to be favoured as a signal of cooperative behaviour via a costly signalling mechanism, it would require that laughing and smiling are too costly for individuals who have chosen not to cooperate. This also seems unlikely—given that laughing and smiling are likely to be relatively cost free, it seems more likely that laughing and smiling act as a signal or bond between individuals with a shared interest. A similar argument can be made about blushing, which is more easily described as mutually beneficial signal of appeasement (Crozier, 2001). An analogous problem occurs in the evolution of language literature when it is suggested that traits such as politeness are costly honest signals, but where the costs and signal are dissociated, and arise from later behaviours such as reciprocity (Knight, 1998, 2008; van Rooy, 2003; see also Scott-Phillips, 2007; Scott-Phillips, 2008).

Howard's (1971) metagames with "transparent disposition" and Gauthier's (1986) theory of "constrained maximization" are also relevant here. In these cases, it is assumed that the second player in a one shot PD can choose a fixed disposition (e.g., always defect, always cooperate, play tit-for-tat, etc.) that can be detected by the first player and that the first player can adjust their strategy accordingly. Given that the second player can predict what the first player will do depending upon their chosen disposition, the second player can choose the disposition that will lead to the maximum payoff (backward induction). The assumption here is that disposition can be chosen facultatively, and so, in order for this to work, disposition must be a costly honest signal, which seems very unlikely (at least to good politicians and poker players; Binmore, 1994, pp. 174–186). A greenbeard version of this hypothesis could also be constructed, but this would require that the outward appearance of disposition be controlled by (or strongly linked to) the genes that control cooperation, which seems even more unlikely.

6.5. Group selection

In this section, we summarise the five misconceptions generated by the group selection literature—the interested reader is directed towards more detailed reviews elsewhere (Gardner & Grafen, 2009; West et al., 2007b, 2008).

6.5.1. Misconception 9. Group selection is a formal theory with one meaning

A major part of the confusion surrounding group selection stems from the fact that the term has been used to mean at least three or four different things (Fig. 3; Okasha, 2004, 2006; West et al., 2007b, 2008).

6.5.1.1. Old group selection and group adaptations. During the 1960s, Wynne-Edwards (1962) argued for the importance of group selection in its original or "old" form. He argued that in groups consisting of selfish individuals, resources would be over exploited, and the group would go extinct. In contrast, groups consisting of cooperative individuals would not over exploit their resources and, so, avoid extinction. Hence, by a process of differential survival of groups, behaviours evolved that were for the good of the group. Another way of looking at this is that selection would favour traits that maximize group success, termed *group adaptations*.

During the 1960s and 1970s, a large amount of theoretical and empirical evidence was piled up against this idea. Theory showed that this type of group selection would only work under extremely restrictive conditions, and so, its importance would be rare or nonexistent (Leigh, 1983; Maynard Smith, 1964, 1976; Williams, 1966). For example, selection will produce behaviours that maximize group success if all the individuals within a group are genetically identical clones, or if there is complete repression of competition within groups, such that the reproductive success of members of the group cannot differ (Gardner & Grafen, 2009). These correspond to the extreme cases where maximizing group success is the same as maximizing inclusive fitness (Fig. 4). Empirical work supported these theoretical conclusions by showing that individuals were reproducing at the rate that maximized their inclusive fitness and were not adapted to maximize group fitness (Krebs & Davies, 1987, 1993; Lack, 1966; West et al., 2008).

It is this old form of group selection that leads people to the false conclusion that individuals behave for the good of the population or species or ecosystem, or that human societies can be viewed as superorganisms in the same way as certain social insect colonies (Kohn, 2008; Shennan, 2002; Wilson et al., 2008; see also the review of the anthropological literature by Soltis et al., 1995). For example, as summed up by quotes such as "the concept of social groups as like single organisms" (Wilson & O'Brien 2009) and "Our species is the primate equivalent of a beehive or a single organism" (Kohn, 2008). Similar
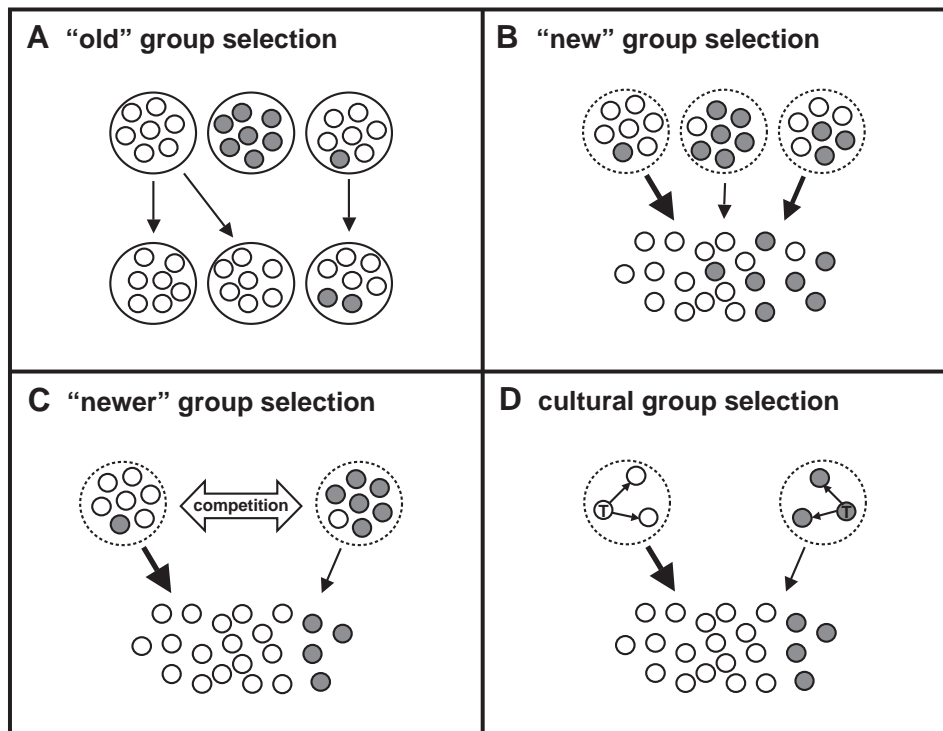
Fig. 3. The different types of group selection. The white circles represent cooperators, whereas the grey circles represent relatively selfish individuals who do not cooperate. Panel A shows the "old" group selection, with well-defined groups with little gene flow between them (solid outline). Competition and reproduction is between groups. The groups with more cooperators do better, but selfish individuals can spread within groups. Panel B shows the "new" group selection, with arbitrarily defined groups (dashed lines), and the potential for more gene flow between them. The different groups make different contributions to the same reproductive pool (although there is also the possibility of factors such as limited dispersal leading to more structuring), from which new groups are formed. Panel C shows the "newer" group selection, which emphasises the more proximate mechanism of inter-group competition as a factor shaping the evolution of social behaviours. Panel D shows cultural group selection, in which social behaviours can be horizontally transmitted between group mates, for example with all individuals in the group imitating the behaviour of one "teacher" (T).

confusion surrounds some discussions of punctuated equilibrium, where it seems to be assumed that this would lead to group-level or species-level adaptations (Arnold, 1993; Shennan, 2002; Zeder, 2009).

6.5.1.2. New group selection. In the 1970s and 1980s, a "new" form of group selection was championed by Wilson and others, which examined the consequences of interactions in small structured populations (Colwell, 1981; Hamilton, 1975; Wilson, 1975a, 1977). These models
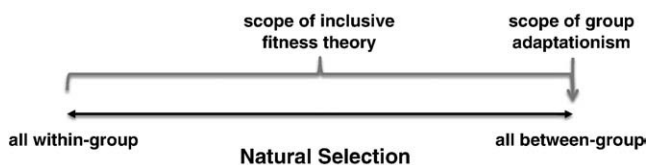

Fig. 4. The scope of inclusive fitness theory and group adaptation. Irrespective of the extent to which selection is within or between groups, natural selection will lead to organisms that appear to be maximising their inclusive fitness (Frank, 1986; Grafen, 2006a; Hamilton, 1975). In contrast, individuals will only be selected to maximise group fitness in the extreme scenario where there is negligible within group selection (Gardner & Grafen, 2009).

assumed that there are multiple levels of selection, which can vary in their importance, and showed that cooperation could be favoured if the benefits at the group level (between-group) outweighed the benefits at the individual level (within-group). It was suggested that this new group selection approach provided an alternative explanation to cooperation or altruism, in situations where kin selection or inclusive fitness could not. However, it has since been realized that group selection and kin selection were just different ways of conceptualizing the same evolutionary process. For example, while the earliest group selection models (e.g., Colwell, 1981; Traulsen & Nowak, 2006; Wilson, 1975a, 1977) were reinventing how indirect fitness benefits (kin selection) can work via limited dispersal, later models (e.g., Wilson & Dugatkin, 1997; Wilson & Hölldobler, 2005) were reinventions of the green beard process (Dawkins, 1979; Foster et al., 2006; Frank, 1986; Grafen, 1984; Hamilton, 1975; Harvey et al., 1985; Lehmann & Keller, 2006; Lehmann et al., 2007b; Maynard Smith, 1976).

The key point here is that this new group selection (multilevel selection) is just a different way of looking at the dynamics by which inclusive fitness is maximized. They are mathematically identical (Frank, 1986, 1995b;

Table 5
Kin selection and new group section

Area	Papers claiming that a result requires group selection	Papers showing equivalent result can be obtained with kin selection / inclusive fitness
Sex ratios with local mate competition	Colwell, 1981; Sober & Wilson, 1998; Wilson & Colwell, 1981	Frank, 1986; Grafen, 1984; Harvey et al., 1985
Sex ratios with budding dispersal	Avilés, 1993	Gardner et al., 2009
Eusociality	Wilson & Wilson, 2007; Wilson & Hölldobler, 2005	Foster et al., 2006; Helanterä & Bargum, 2007; Hughes et al., 2008
Strong reciprocity	Bowles & Gintis, 2004; Bowles et al., 2003; Boyd & Richerson, 2002; Boyd et al., 2005; Fehr & Fischbacher, 2003; Gintis, 2000; Gintis et al., 2003; Henrich, 2004; Traulsen & Nowak, 2006	Gardner & West, 2004; Lehmann et al., 2007c
Cooperation	Bowles, 2006; Taylor & Nowak, 2007; Traulsen & Nowak, 2006	Hamilton, 1975; Lehmann et al., 2007b
Virulence	Kohn, 2008; Sober & Wilson, 1998; Wilson, 2008; Wilson & Wilson, 2007	Frank, 1996b; Wild et al., 2009
Policing	Sober & Wilson, 1998	Ratnieks, 1988; Wenseleers et al., 2004

There is no theoretical or empirical example of group selection that cannot be explained with kin selection. Here, we provide examples of situations where it has been argued that group selection gives a result that cannot be explained by kin selection, but where it was then shown that it can. More general theoretical overviews are provided elsewhere (Frank, 1986; Gardner et al., 2007a; Grafen, 1984; Hamilton, 1975; Queller, 1992a).

Gardner, 2008; Gardner & Grafen, 2009; Gardner et al., 2007a; Grafen, 1984, 2006a; Hamilton, 1975; Lehmann et al., 2007b; Queller, 1992a; Wade, 1985). New group selection models show that cooperation is favoured when the response to between-group selection outweighs the response to within-group selection, but it is straightforward to recover Hamilton's rule from this. Both approaches tell us that increasing the group benefits and reducing the individual cost favours cooperation. Similarly, group selection tells us that cooperation is favoured if we increase the proportion of genetic variance that is between-group as opposed to within-group, but that is exactly equivalent to saying that the kin selection coefficient of relatedness is increased (Frank, 1995a). In all cases, where both methods have been used to look at the same problem, they give identical results (Table 5). This is not surprising given how they are both formalized with the Price equation (Frank, 1986; Gardner, 2008; Gardner et al., 2007a, submitted). As we shall discuss in further detail in Misconception 13, the reason that most biologists focus on the inclusive fitness or kin selection approach is that it is much easier to develop models and apply them to real organisms (West et al., 2008).

6.5.1.3. Newer group selection. More recently, over the last decade, group selection has been used in a third "newer" way. In these models, it is argued that a key factor favouring cooperation is direct competition between groups, and this is referred to as group selection (Binmore, 2005a; Bowles, 2006, 2009; Bowles et al., 2003; Boyd & Richerson, 1990, 2002; Boyd et al., 2003; Gintis, 2003; Gintis et al., 2003; Henrich, 2004). For example, as discussed in Misconception 2, when groups compete for territories, and territories are won by the groups with the most cooperators. However, these models do not provide

an alternative to inclusive fitness or kin selection — individuals gain a direct fitness benefit through cooperating because they increase the success of their group (including themselves), and an indirect fitness benefit in the cases where the models also assume limited dispersal, which leads to significant relatedness between the individuals in a group (see Misconceptions 5, 6 ,and 15). Another distinction is that kin selection, old group selection and new group selection are examining the level at which ultimate selective forces act, whereas the newer group selection is more proximate, saying that group competition plays a causal role in mediating the fitness consequences of cooperative behaviors. Brewer and Caporael (1990) define group selection to mean that the group is the selection environment for human evolution at the individual level, which is analogous but not exactly equivalent to newer group selection.

6.5.1.4. Cultural group selection. The term group selection is also used when discussing "cultural group selection" or "gene-culture coevolutionary multilevel selection." Cultural group selection is used to label situations when differential group success results from the expression of different cultural traits (Boyd & Richerson, 2005; Fehr & Fischbacher, 2005b; Fehr et al., 2002; Gintis, 2003; Gintis et al., 2003; Henrich, 2004; Henrich & Boyd, 2001; Lehmann et al., 2008; McElreath & Henrich, 2006; Richerson & Boyd, 2005). This is analogous to the third use described above, in that it is used to mean that competition occurs between groups. However, it differs in that it refers to selection on a cultural trait, rather than a genetically determined trait. As with genetic group selection, just because competition is occurring between groups, this does not mean that group level adaptations are expected to evolve (Gardner & Grafen, 2009). Consequently, while it is

often argued that the group is the fundamental unit of cultural evolution, or that cultural evolution is a group-level process (Boyd & Richerson, 1985), there is no formal basis for this. Finally, we also note that it has been suggested that there are even three different types of cultural group selection (Henrich, 2004)!

6.5.1.5. The various group selections. The above discussion shows how the term group selection has been used to mean three to six different things (Fig. 3), specifically, that (1) selection produces traits that maximize group fitness (old), (2) selection acts at multiple levels (new), or (3) competition occurs between groups (newer). The various forms of cultural group selection could be either subsumed under newer, or form a new category ("even newer") or categories. This variable use of group selection has been possible because there is no formal theory of group selection (West et al., 2008, p.380–381; Gardner & Grafen, 2009), which leads to authors confusingly switching between different meanings (Palmer et al., 1997; Trivers, 1998a, 1998b; West et al., 2007b, 2008). For example, several authors switch between the old and new group selection, using the new to justify the old (e.g., O'Gorman et al., 2008; Robson, 2008; Sober & Wilson, 1998; Wilson et al., 2008), while Bergstrom (2002) discusses all three types as if they are the same thing (old: pp. 85–86; new: pp. 71–72, 76–77, 80; newer: pp. 81, 85–86).

A lack of an appreciation of the different types of group selection has led to numerous sources of confusion. These include (A) The new group selection approach has been used to justify old group selection thinking (e.g., Sober & Wilson, 1998; Wilson, 2008; Wilson & Wilson, 2007). (B) A new group selection approach is used to produce an equation that it is a form of Hamilton's rule, and so of general importance, but then, on the basis of old group selection thinking, it is suggested that this is unlikely to be important for genetic traits or outside of humans (Bowles et al., 2003, pp. 136-140; Boyd & Richerson, 1990, p. 340; Henrich, 2004, pp. 15–16). This is analogous to saying that indirect fitness effects are thought to be generally unimportant, which is clearly incorrect. (C) The group selection jargon hides links with other areas of evolutionary theory. For example: (i) how the various group selection models with limited dispersal (e.g., Bowles et al., 2003; Boyd & Richerson, 2002; Boyd et al., 2005; Traulsen & Nowak, 2006) relate to the inclusive fitness literature on the same issues (reviewed by Lehmann et al., 2007b; Queller, 1992b; West et al., 2002a, 2008); (ii) that some models (e.g., Gintis, 2000; Wilson & Dugatkin, 1997) rely on greenbeard effects and, so, are unlikely to be of general importance, especially in humans (see Misconception 7), and (iii) it can obscure the various mechanisms by which within group competition can be repressed, such as reciprocity, punishment, ostracism etc (e.g., O'Gorman et al., 2008).

6.5.2. Misconception 10: Group selection can apply in situations when inclusive fitness cannot explain cooperation (e.g., Arrow, 2007; Baschetti, 2007; Bergstrom, 2002; Boyd et al., 2003; Fehr et al., 2002; Gintis et al., 2001, 2003; Henrich, 2004; Richerson & Boyd, Manuscript (1999)

This is incorrect. The old group selection ideas only work in the extreme scenarios where there is negligible within group selection, which can occur via high relatedness or repression of competition (Fig. 4; Gardner & Grafen, 2009). In contrast, individuals are expected to maximise their inclusive fitness irrespective of the relative strengths of within-group versus between-group selection (Grafen, 2006a; Hamilton, 1975). New group selection is not an alternative to inclusive fitness—it is just a different way of looking at the dynamics of natural selection. Finally, the newer group selection is also not in conflict with inclusive fitness—it is a mechanism for providing direct and/or indirect fitness benefits.

A recent example of the confusion that can arise here is provided by two quotes from the same paragraph of Boyd et al. (2005, p.215). It is first claimed that group selection works when interactions are not between relatives (this misconception), but then, stated that group selection only favours altruism when groups are small and migration rare (i.e., which is when limited dispersal means interacting individuals will be highly related—see Misconceptions 5 and 6): "Cooperation among nonkin is commonly explained by one of two mechanisms: repeated interactions (Axelrod & Hamilton 1981; Trivers 1971; Clutton-Brock & Parker 1995) or group selection (Sober & Wilson 1998)" and "Group selection can lead to the evolution of altruism only when groups are small and migration infrequent (Eshel 1972; Aoki 1982; Rogers 1990)" (see also Boyd & Richerson, 2002).

6.5.3. Misconception 11: Inclusive fitness or kin selection is a subset of group selection

This is incorrect. Kin selection and group selection are just different ways of carving up the dynamics by which inclusive fitness maximisation is reached. Consequently, it is no surprise that no group selection model has ever been constructed where the same result cannot be found with kin selection theory (Table 5). Although, while it is possible to translate all group selection models into corresponding kin selection models, the reverse may not be true. One reason for this is that it can be hard or impossible to incorporate many important biological complexities into group selection models (Queller, 2004). It is for this reason that group selection models have focused on the simplest possible cases, whereas the inclusive fitness approach is also used to develop specific models and provide predictions that can be tested with empirical work (West et al., 2008). Another reason is that the inclusive fitness approach has successfully integrated fundamental issues that have not been tackled in the group selection literature, such as the theory of reproductive value and gene-frequency change in class-structured populations (Frank, 1997, 1998; Taylor, 1990,

1996; Taylor & Frank, 1996; Taylor et al., 2007b). This has proven particularly useful for dealing with issues such as different forms of dispersal in spatially structured populations (West et al., 2002a).

6.5.4. Misconception 12: Group selection leads to group adaptations (Reeve & Hölldobler, 2007; Sober & Wilson, 1998; Wilson & Wilson, 2007; Wilson & Hölldobler, 2005; Wynne-Edwards, 1962)

As discussed in Section 2, Darwinism is a theory of the process and purpose of adaptation. The purpose is that natural selection should lead to individuals appearing as if they were designed to maximize their fitness, and that this fitness is inclusive fitness (Grafen, 2006a; Hamilton, 1964). In contrast, since Wynne-Edwards, a number of workers have argued that group selection will lead to "group adaptations" that have been selected for because of their benefit for the good of the group, and that groups can be viewed as adaptive individuals (superorganisms) in their own right (Reeve & Hölldobler, 2007; Sober & Wilson, 1998; Wilson, 2008; Wilson & Wilson, 2007; Wilson & O'Brien, 2009; Wilson & Hölldobler, 2005; Wynne-Edwards, 1962). However, formal analysis has shown that selection for group adaptations requires special circumstances, with negligible within group selection (Fig. 4), such as when (a) the group is composed of genetically identical individuals (clonal groups, r=1), or (b) there is complete repression of competition between groups (i.e., no conflict within groups; Gardner & Grafen, 2009).

It is useful here to distinguish adaptation and design from dynamics of how selection leads to design. The dynamics of selection can be examined with either an individual (inclusive fitness or kin selection) or group selection approach. However, only the individual level approach provides a general model of adaptation. The idea that individuals strive to maximise their inclusive fitness holds irrespective of the intensity of selection operating within and between groups (Section 2; Fig. 4). In contrast, as discussed above, group adaptations or maximization of fitness at the group level are only expected in the extreme case where there is no within group selection.

6.5.5. Misconception 13: Most evolutionary biologists view group selection as hotly debated, completely wrong, or that there is some ulterior motive for the lack of attention given to it (Baschetti, 2007; Sober & Wilson, 1998; Traulsen & Nowak, 2006; Wilson & Wilson, 2007)

This misconception is encapsulated in phrases such as "I believe that this is a hold-over of US ideologies, which have been strongly individualist and anti-collectivist" (Baschetti, 2007), or "vigorous criticism and a general denial of such ideas" (Traulsen & Nowak, 2006). We cannot stress enough that this is incorrect. While the old group selection idea does not hold (selection does not maximize fitness at the group level except under the very special circumstances described in Misconception 12), the new or newer ideas are able to

explain the dynamics of natural selection. Indeed, many researchers who normally focus on the kin selection approach, including ourselves, use multi level selection methods when they are the most appropriate tool for solving the problem (Frank, 1998; Gardner et al., 2007a; Gardner & Grafen, 2009). The reason that most evolutionary biologists, both theoretical and empirical, do not use the group selection approach, or use it very little, is that they find it less useful (Frank, 1998; Queller, 2004), and if they express negative views, it is because it has generated more confusion than insight (reviewed in detail by West et al., 2007b, 2008). Put another way, the method isn't wrong *per se*, it is more that it is often misused and misinterpreted.

The inclusive fitness approach has received more attention because it is easier to develop general models and apply them to real biological situations. It is for this reason that (a) the group selection debate only takes place over simple models and has not stimulated empirical work, and (b) all the major developments in social evolution theory have been pioneered and led by the inclusive fitness approach, and not group selection (Section 2.2; West et al., 2008). In contrast to this empirical progress spurred by the inclusive fitness approach, group selection thinking appears to be easy to misapply, leading to incorrect statements about how natural selection operates, as shown by research in many areas such as animal behaviour (reviewed by Dawkins, 1976), microbiology (reviewed by West et al., 2006a), parasitology (reviewed by Herre, 1993) and agriculture (reviewed by Denison et al., 2003). While inherently attractive, it is highly misleading to portray multilevel selection as a means to unify the economic and social sciences by suggesting that our self-regarding preferences (broadly matching the predictions of classical rational choice theory) are explained by biological individualist selection while population-level (principally cultural) evolutionary processes explain why we have pro-social preferences (e.g., Shennan, 2002).

6.6. Strong Reciprocity

In recent years, there has been much attention to the suggestion that cooperation in humans can be explained by "strong reciprocity," which is defined as a predisposition to help others and to punish those that are not helping (Bowles & Gintis, 2004, 2008; Boyd et al., 2003; Fehr & Gächter, 2002; Fehr & Fischbacher, 2003; Fehr & Rockenbach, 2004; Fehr et al., 2002; Gintis, 2000, 2003; Gintis et al., 2003, 2005a). This literature has contributed to 10 misconceptions, numbers 1, 2, 4–8 and 14–16. It is useful here to divide the work on strong reciprocity into four areas—what the empirical data show, what it is argued the empirical data show, what the theoretical models show, and what it is argued the theoretical models show. A major source of confusion is that all of these four areas are in disagreement with each other, and that there are several inconsistencies between the different papers on this topic.

A number of elegant economic experiments have suggested that people have a propensity to cooperate and punish individuals who do not cooperate (Burnham & Johnson, 2005; Fehr & Fischbacher, 2003; Gächter & Herrmann, 2009). Importantly, this includes one-shot games, without the possibility for repeated interactions, where individuals would gain a greater financial reward from not cooperating or punishing. This is a clear demonstration that people do not always behave in ways that maximise their economic payoffs, even if they are given perfect knowledge. It has been argued that strong reciprocity provides an explanation for this behaviour (see Misconception 4 for quotations).

6.6.1. Misconception 14: Human cooperation in economic games requires the novel evolutionary force of strong reciprocity

There is a large empirical literature showing that when humans play anonymous one-shot economic games, they cooperate more than would be expected if they were purely self-interested (Ledyard, 1995). From a proximate perspective, it has been argued that this is because individuals value the success of others as well as their own, showing prosocial preferences (Fehr and Schmidt, 1999). From an evolutionary perspective, it has been argued that this proximate mechanism cannot be explained by standard evolutionary explanations of cooperation, such as kin selection and reciprocity, and requires the a novel explanation of strong reciprocity (Fehr & Gächter, 2002; Fehr & Fischbacher, 2003; Fehr & Rockenbach, 2004; Fehr et al., 2002; Gintis et al., 2003, 2005b). For example, it has been claimed that human behaviour "cannot be rationalized as an adaptive trait by the leading evolutionary theories" (Fehr et al., 2002).

However, the empirical data are open to multiple explanations and do not support this claim. First, in some cases, an equally valid explanation for the data is that humans are antisocial, rather than prosocial. In the ultimatum game, the expected strategy is for individuals to make minimal offers and for these to be accepted. If there is a chance that minimal offers will be rejected (punished) then individuals are expected to make larger offers (Gale et al., 1995). Consequently, the larger than minimal offers that are observed in experiments may just reflect the fact that individuals expect small offers to be punished. In this case, the unexpected behaviour is the rejection of small offers, and so, we might conclude that the data show that humans have a tendency to punish at a level greater than that expected from selfish interests. Note that our purpose here is not to argue that humans are particularly pro- or antisocial, just that it is easy to give multiple explanations for the data.

Second, higher than expected levels of cooperation can be explained by individuals making mistakes in laboratory settings. The previous interpretation of economic games is based upon the implicit assumption that if individuals do not play perfectly, then this does not lead to a systematic bias in the level of cooperation (Kümmerli et al., 2010). This is a problem, because when the predicted behaviour is to not cooperate at all (e.g., in standard public goods games), then any deviations from perfection would automatically be perceived as greater than expected cooperation (Houser & Kurzban, 2002; Kümmerli et al. 2010) tested this possibility, by allowing individuals to play modified versions of public goods games, where 100% cooperation was the strategy that would maximise their personal financial game. They found that while this led to an increased level of cooperation, it did not lead to full cooperation (see also Houser & Kurzban, 2002; Laury & Holt, 2008; Saijo & Nakamura, 1995). If the logic from previous studies (e.g., Fehr & Schmidt, 1999) was applied to this result, then it would give a utility function that is negatively influenced by the success of others (an antisocial preference). Given that a simultaneous positive and negative regard to others is not possible, these data instead suggest that individuals have a tendency to avoid both full defection and full cooperation (Haselton & Nettle, 2006).

Third, another possible explanation is that higher levels of cooperation are normally favoured and that this leads to a psychology that results in cooperation in one-shot experiments. The idea here is that, even if they are given perfect information, individuals find it hard to disassociate themselves from the real world, and so, cooperation occurs as a byproduct of the fact that is normally favoured (Bateson et al., 2006; Binmore, 2006; Burnham & Johnson, 2005; Hagen & Hammerstein, 2006; Haley & Fessler, 2005; Levitt & List, 2007; Nowak et al., 2000; Trivers, 2004; West et al., 2007b). Experimental support for this suggestion comes from a number of experiments which show that players taking part in one shot games, in which there are no future interactions, still adjust their level of cooperation in response to artificial cues, such as the presence of eye-spot pictures on computer desktops (Bateson et al., 2006; Burnham & Johnson, 2005; Haley & Fessler, 2005) or interactions with individuals which do not influence the game (Houser & Kurzban, 2002; Kurzban et al., 2007). The idea here is that these cues trigger responses that have arisen in response to situations outside of the laboratory, where whether or not they are being observed will matter. Further support comes from cultural differences in experimental games (Gächter & Herrmann, 2006; Henrich et al., 2006; Henrich et al., 2005), which appear to reflect differences in how the game is perceived to relate to everyday events (Binmore, 2006). To put it another way, "Experimental play often reflects patterns of interaction found in everyday life" (Henrich et al., 2005, p. 798) and not just the game set up imposed by the experimenter. Furthermore, even in laboratory settings, behaviours such as punishment can provide a direct benefit if longer periods of interactions are allowed for (Gächter et al., 2008).

The data discussed in the previous two paragraphs suggest that humans have a psychology which can "misfire" in laboratory settings. While it might be argued that the possibility that humans don't always behave perfectly is no

surprise, the more important point is that such imperfect behaviour can lead to a systematic bias towards higher than expected levels of cooperation. Future work must address this issue, through the use of appropriate controls and by exercising greater caution when interpreting the absolute level of cooperation in particular treatments (Kümmerli et al., 2010). It would be useful to test whether there is a bias to accept evidence for humans being "extra cooperative," without sufficient basis, due to a bias towards positive evidence or because this is a nice result to get.

This "misfire" idea has been argued to be incorrect in several papers, where it is labelled the "big mistake" or maladaptation hypothesis (Boyd & Richerson, 2002; Fehr & Henrich, 2003; Gintis et al., 2003; Henrich, 2004). The implicit idea here is that humans should always behave perfectly. However, this hypothesis is clearly falsified by the numerous examples of how proximate mechanisms which have been previously favoured by natural selection lead to behaviours that do not maximise fitness under certain conditions. For example, the mismatch between real danger and our fear of snakes and spiders versus automobiles, various aspects of the porn industry, rises in obesity, or the decline in reproductive rate can be associated with better living conditions (Hagen & Hammerstein, 2006). It is even clearly falsified in the context of economic games, where, as discussed above, individuals show variation in behaviour in response to misleading "cues" of being observed, such as eye-spots on computers. It can be misleading to call such imperfect behaviour a maladaptation or a mistake, in the sense that it may be the optimal state, just that the benefits of improving a behaviour have to be balanced or traded off against the costs (Partridge & Sibley, 1991; Stearns, 1992). The general point here is that maximisation of fitness does not imply perfect behaviour in every possible situation, and that the selective regimen needs to be considered, as has been shown frequently in the animal behaviour literature (Davies, 1992; Herre, 1987; Herre et al., 2001; Krebs & McCleery, 1984; Pompilio et al., 2006; Wehner, 1987). Evolutionary theory does not predict that humans (or any other organism) should behave as perfect maximising agents in every situation in which they can be placed.

6.6.2. Misconception 15: The theoretical models on strong reciprocity provide a novel solution to the problem of cooperation, that are outside of the usual inclusive fitness explanations (Bowles & Gintis, 2004; Fehr & Rockenbach, 2003, 2004; Fehr & Fischbacher, 2005b; Gintis, 2000)

It has been claimed that the theoretical models of strong reciprocity do not rely on "explanatory power of inclusive fitness theory" and "cannot be explained by inclusive fitness" (Bowles & Gintis, 2004) and that they can explain the evolution of cooperation and punishment, even when they do "not yield future economic benefits for the altruist" (Fehr & Rockenbach, 2003) "it is implausible to expect that these costs will be repaid" or "even though as a result they receive lower payoffs than other group members" (Bowles &

Gintis, 2004). However, this is not possible—a trait will not be selected for unless it provides an inclusive fitness benefit (see Section 2). One source of confusion here is the jargon used in the strong reciprocity modeling literature, in that the strategies that are referred to as altruistic are not necessarily altruistic as they can lead to an increase in personal fitness (Misconception 1).

The other source of confusion is that while the impression is given that the strong reciprocity models do not rely upon standard direct and indirect fitness benefits, more formal analyses have shown that they do, it is just that this was not made explicit (Gardner & West, 2004; Lehmann et al., 2007c). Cooperation can provide a direct benefit because it provides a benefit to everyone in the group, including the focal cooperator through reducing the chance of group extinction or increasing the chance of success in between group competition (analogous to models of "group augmentation"). Cooperation can provide an indirect benefit because these models assume limited dispersal, which leads to a significant relatedness between the individuals interacting within the group (Misconception 6), for example, $r \approx 0.1$ in groups of size 50 if the migration rate is 0.1 (Lehmann et al., 2007c). This extent to which relatedness can build up appears to be frequently ignored in the strong reciprocity theoretical literature—for example, Bowles & Gintis (2004) assume group sizes of 20, where relatedness will be higher, but claim that "there are many unrelated individuals, so altruism cannot be explained by inclusive fitness" (Bowles & Gintis, 2004). Note that we are not saying that in their model strong reciprocity is always altruistic, as both direct and indirect benefits can occur and so whether it is mutually beneficial or altruistic will depend upon parameter values (Lehmann et al., 2007c). Similarly, punishment can provide a direct or indirect benefit by reducing competition for the actor or their relatives, respectively.

Overall, the relative importance of direct and indirect fitness benefits will depend upon the details and parameter values of a model (Gardner & West, 2004; Lehmann et al., 2007c). Specifically, whether cooperation and punishment are favoured as either mutually beneficial or altruistic behaviours depends upon parameters such as group size and the dispersal rate (Lehmann et al., 2007c). For example, decreasing group size makes cooperation and punishment more likely to provide a direct benefit because the actor gains a greater share of the group benefit from cooperation, and a greater benefit from the reduced competition that follows from punishment. A general point here is that the earlier models of strong reciprocity were analysed with a simulation approach and then explained with verbal arguments. Since then, multilocus population genetic methodology has been used to provide analytical solutions that allow the underlying selective forces to be formally analysed, showing that these earlier verbal arguments were incorrect (Lehmann et al., 2007c). Considering Fig. 2, the strong reciprocity models have involved selective forces that occur on multiple branches

[e.g., non-enforced direct benefits; enforced direct benefits (punishment); indirect benefits by limited dispersal], as well as a branch outside the tree that isn't even cooperation (spiteful green beards).

6.6.3. Misconception 16: The claims made in the empirical and the theoretical strong reciprocity literature are compatible

We return to our point that there are four contradictory aspects of strong reciprocity. First, the empirical results show that humans cooperate at higher levels than expected in some situations, and punish individuals who do not cooperate. That is a clear and repeatable result. Second, it has been claimed that this propensity can be explained by strong reciprocity. However, strong reciprocity is a proximate mechanism and not a solution to the ultimate problem of why humans cooperate (Misconception 4). Third, it has been claimed that the theoretical models of strong reciprocity can explain cooperation and punishment in one-shot encounters and that they provide a novel solution to the problem of cooperation that is outside of inclusive fitness theory. Fourth, the theoretical models of strong reciprocity actually show how competition between groups and limited dispersal can lead to direct and/or indirect benefits to cooperation (Misconceptions 1, 4–6, 8 and 15). These models therefore are easily understood from an inclusive fitness context and do not predict cooperation in one-shot encounters. In order to predict cooperation in one-shot encounters, it would be necessary to develop more mechanistic models, which allowed for factors such as a trade-off between the complexity of a strategy and its cost, and could hence predict misfiring (Misconception 14).

The potential confusion that can arise from these contradictions is illustrated in how two sentences from the abstract of a single paper can contradict each other (Gintis et al., 2003). Specifically, it is first claimed that strong reciprocity cannot be explained by standard evolutionary models, then soon followed by a second sentence that claims strong reciprocity is evolutionarily stable (which means it can be explained by evolutionary theory): "strong reciprocity is a predisposition to cooperate with others and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid." and "We show that under conditions plausibly characteristic of the early stages of human evolution, a small number of strong reciprocators could invade a population of self regarding types, and strong reciprocity is an evolutionary stable strategy." Confusion also arises because of inconsistencies between papers. For example, compare the first quote in this paragraph with "strong reciprocity must have promoted individual fitness, or it could not have evolved. Our contention is that strong reciprocity enhanced relative fitness because groups with a high frequency of altruism survived and prospered at a higher rate than groups with a low frequency of altruism" (Gintis et al., 2008, p. 248).

6.7. Cultural evolution

Up until now, we have focused on genetic evolution. However, humans are clearly unique in the extent to which behaviour can be transmitted culturally, and the possible role of cultural evolution also needs to be considered. Culture is information capable of affecting an individual's behaviour that is acquired from other members of their species through teaching, imitation, and other forms of social transmission or social learning (Boyd & Richerson, 1985). Cultural traits can therefore be transmitted horizontally between individuals of the same generation. This contrasts with genetically inherited traits that are generally only passed vertically from parent to offspring, with notable exceptions in bacteria (Smith, 2001; West et al., 2006a). It is often suggested that cultural evolution is able to explain cooperation in cases where genetic selection cannot (Bergstrom, 1995; Boyd & Richerson, 1985; Boyd & Richerson, 2002, 2005, 2006; Boyd et al., 2003; Fehr et al., 2002; Gintis, 2003; Henrich, 2004; Henrich & Boyd, 2001; McElreath & Henrich, 2006; Richerson & Boyd, 2005). One reason for this is that cultural traits can be transmitted horizontally within groups, which could lead to cultural relatedness r being higher than genetic r.

However, recent theory by Lehmann et al. (2007c, 2008; Lehmann & Feldman, 2008b) has questioned whether cultural evolution will automatically make it easier for cooperation to evolve. Consider the case of when imitation occurs through adaptive learning mechanisms such as "pairwise payoff comparison" or "prestige" or "success" bias, where individuals copy others, from either their own or other groups, on the basis of some arbitrary payoff or estimate of success (Bergstrom, 1995; Boyd & Richerson, 1985, 2002; Boyd et al., 2003; Henrich, 2004). In this case, in comparison with genetic selection, cultural selection is less favourable for cooperation, and even leads to selection for harming behaviours (Feldman et al., 1985; Lehmann et al., 2007c, 2008). The reason for this is that by helping neighbours and, hence, achieving a lower payoff, a helping individual makes it less likely that they will be imitated. Conversely, harming can be selected for because it decreases competition with neighbours, who will then be less likely to be chosen (Lehmann et al., 2008). Lehmann et al. (2008; 2007c) argue that earlier papers came to the different conclusion that such imitation could favour cooperation because: (i) Boyd et al. (2003) did not compare the situation with genetic evolution, they just claimed it would be less likely to favour cooperation; (ii) Boyd and Richerson (2002) made the additional assumption that there was some other mechanism driving the initial spread of the trait, so that it exceeded a certain threshold frequency at which it became beneficial (through avoidance of punishment) and, hence, was no longer altruistic.

Our aim here is not to argue whether cultural evolution makes it easier or harder for cooperation to evolve. This is an exciting and active area of research with much to be

done. Instead, we merely wish to emphasise that this provides another example of the need to formally determine how theoretical models are working, and their relation to existing theory. Considering the example given in the previous paragraph, Boyd et al. (Submitted) have argued that the difference between their and Lehmann's results are due to differences in whether small or large fitness consequences were allowed for, whereas Lehmann replies that both small and large effects were examined in Lehmann et al. (2007c), and no assumptions were made about the size of fitness effects in Lehmann & Feldman (2008b). The advantage of this debate is that it makes such assumptions explicit and so will clarify when cultural evolution either favours or disfavours cooperation, but also why. More specifically, what forms of cultural mechanisms would be favoured by genetic selection, and how would these influence selection for cooperation?

7. Discussion

In the preceding sections we have provided a general review of social evolution theory, the potential solutions to the problem of cooperation, and some common misconceptions. Here, we return to the specific questions surrounding cooperation in humans: (1) Why do humans cooperate? (2) Are humans special, and if so, why? Throughout, our focus is on why humans behave as they do, rather than what they ought to do, i.e., positive, not normative or regulative, science (Friedman, 1953).

7.1. Why do humans cooperate?

The discussion surrounding Misconceptions 1 and 6 make it clear that cooperation in humans could have originally evolved due to either (or both) direct and indirect fitness benefits. Direct benefits could have arisen for a number of reasons including more cooperative groups being more successful, through competition with other groups or avoiding group extinctions (group augmentation), all the usual reciprocity arguments, avoidance of punishment and other mechanisms. Indirect benefits are likely because reasonable estimates of migration rates and group sizes for early hominids suggest there would have been appreciable relatedness between interacting individuals (Lehmann et al., 2007c). Indeed, a synergy between direct and indirect benefits is also likely—as discussed in Section 5.3, direct benefits are often more likely to become important when cooperation is already favoured due to indirect benefits.

A possible question is what were/are the relative importance of direct and indirect fitness benefits in explaining cooperation in humans? However, we suggest that this question is so unanswerable to be almost pointless. The relative importance of direct and indirect fitness benefits depends upon the exact parameter values of theoretical models, with the same model being able to lead to mutually beneficial or altruistic cooperation depending on the values

taken by its parameters (Lehmann et al., 2007c). Researchers are unlikely to be able to obtain sufficiently good parameter estimates about ancestral humans to address this problem with sufficient confidence. This is clearly illustrated by the extent to which the last 40 years of research have been unable to resolve the relative importance of direct and indirect fitness in cooperative breeding vertebrates, where the empirical and experimental opportunities are much greater (Clutton-Brock, 2002; Cockburn, 1998; Griffin & West, 2002; Jennions & Macdonald, 1994).

We stress here that our aim when discussing the various misconceptions has not been to argue against the possible importance of factors such as punishment, between-group competition or cultural evolution. Instead, our main aim has been to point out that there is often a large disparity between what it is claimed is shown by a particular data set or theoretical model, and what is actually shown. Key examples have included claiming that (1) relatedness is not important in a particular model, but then, assuming a population structure that leads to an appreciable relatedness between interacting individuals, i.e., relatedness is there, just unacknowledged (Misconceptions 5–7, 15); (2) an altruistic group-beneficial trait is being modelled, when actually the trait can be mutually beneficial (Misconceptions 1, 2 and 15), or even spiteful and costly at the group level (Misconception 7); (3) proximate data provides an answer to an ultimate question (Misconception 4). Similar examples can be found elsewhere, such as discussions on how and when selection favours hostility between groups (compare Choi & Bowles, 2007 with Lehmann & Feldman, 2008a), or the debate over how and why cultural evolution models influence the evolution of helping (Section 6.7).

7.2. Are humans special?

It is frequently assumed that the form of cooperation in humans is special (Boyd & Richerson, 2002; Boyd et al., 2003; Fehr & Fischbacher, 2003, 2005b; Fehr & Rockenbach, 2004; Henrich, 2004). For example "The nature and level of cooperation in human societies is unmatched in the animal world" (Quervain et al., 2004) or "Human cooperation represents a spectacular outlier in the animal world" (Fehr & Rockenbach, 2004) or "Human altruism goes far beyond that which has been observed in the animal world" (Fehr & Fischbacher, 2003). Indeed, this assumption has even been taken as a starting point, that cooperation in humans requires different evolutionary (ultimate) forces, rather than something that must be demonstrated: "What are the ultimate origins behind the rich patterns of human altruism described above? It must be emphasized in the context of this question that a convincing explanation of the distinct features of human altruism should be based on capacities which are distinctly human—otherwise, there is the risk of merely explaining animal, not human, altruism." (Fehr & Fischbacher, 2003). In this section we critically assess the different ways in which

human cooperation may be special. We are not denying that humans could be special, but want to determine, from an evolutionary perspective, exactly why.

Do humans have especially high levels of altruism (Fehr & Fischbacher, 2003, 2005b; Warneken et al., 2007)? No, a number of organisms have higher levels of altruism than humans, ranging from social amoebae and bacteria to ants and cooperative breeding vertebrates. In both social amoebae and the social insects, a number of individuals completely forgo the chance to reproduce to help others, which represents the most extreme possible form of altruism. In social amoebae and bacteria, these are the stalk cells, which hold up spore cells so that they can be dispersed (Bonner, 1967; Gilbert et al., 2007; Velicer et al., 2000). In social insects these are the sterile workers that give up the chance to reproduce for themselves and instead help to raise the offspring of the queen or queens (Bourke & Franks, 1995; Hamilton, 1972). In cooperative vertebrates, helping is sometimes mutually beneficial, and sometimes altruistic, depending upon the species (Griffin & West, 2003). An extreme example at the altruistic end of the continuum is the long tailed tit, where helpers never reproduce and so cooperation has been favoured purely by indirect fitness benefits (MacColl & Hatchwell, 2004; Russell & Hatchwell, 2001). In contrast, in humans, direct fitness benefits are often likely to play a greater role, and cooperation is more likely to be mutually beneficial than altruistic.

Are humans special because cooperation occurs between nonrelatives (Boyd & Richerson, 2002; Boyd et al., 2003; Fehr & Fischbacher, 2003; Fehr & Rockenbach, 2004; Henrich, 2004)? No, cooperation between nonrelatives occurs in a range of organisms. Many forms of cooperation occur between nonrelatives in birds and mammals (Clutton-Brock, 2002). In cooperative breeding vertebrates, there are several examples where non-relatives cooperate, the indirect fitness benefits of cooperation appear to be negligible and it is thought that cooperation is driven by direct fitness benefits (Clutton-Brock, 2002; Cockburn, 1998; Emlen, 1997; Griffin & West, 2002; Jennions & Macdonald, 1994; Krebs & Davies, 1997). Even in social insects such as ants and wasps, there are some examples where nonrelatives come together for mutually beneficial cooperation (Bernasconi & Strassmann, 1999; Queller et al., 2000). However, perhaps the most extreme examples of cooperation between nonrelatives are the various examples of cooperation between species, termed *mutualisms* (Herre et al., 1999; Sachs et al., 2004; West et al., 2007a). For example, between cleaner fish and their clients on the tropical reef, fig trees and fig wasps, plants and their mycorrhizae or rhizobia root symbionts, or the various symbionts that live within animal hosts. Finally, we also note that cooperation between non-relatives has also played a key role in some of the major evolutionary transitions, such as the incorporation of symbiotic bacteria that became mitochondria, in the transition to eukaryotes (supplementary material; Queller, 2000).

Are humans special because we enforce cooperation with mechanisms such as punishment? No, enforcement occurs across a range of taxa from plants to animals (Section 5.2.2). For example, clients chase and attack cleaner fish that do not cooperate (Bshary & Grutter, 2002; Bshary & Grutter, 2005), soya bean plants cut off the oxygen supply to rhizobia that do not supply them with nitrogen (Kiers et al., 2003), dominant meerkats attack and evict subordinates who try to breed (Young et al., 2006), honey bees destroy (police) eggs laid by workers (Ratnieks & Visscher, 1989) and ineffective pollinators are punished by a range of plant species (Goto et al., 2010; Jander & Herre, 2010; Pellmyr & Huth, 1994).

In contrast, what appears to be special about cooperation in humans is the proximate factors involved. Humans are able to assess the local costs and benefits of cooperative behaviour, and adjust their behaviour accordingly (Fehr & Gächter, 2002; Fehr & Rockenbach, 2003; Fehr & Fischbacher, 2003, 2004; Henrich et al., 2005; Semmann et al., 2004; West et al., 2006b). Consequently, human cognitive abilities allow individuals to be highly flexible in the level of cooperation they perform in response to whether there is the possibility for punishment (Fehr & Gächter, 2002), cues of reciprocity (Bateson et al., 2006; Semmann et al., 2004), whether they are competing locally or globally for resources (West et al., 2006b), and competition between groups (Burton-Chellew et al., 2010; Puurtinen & Mappes, 2009). In many of these cases, human behaviour does appear to be special. For example, the importance of reciprocity in humans contrasts with the lack of evidence for it playing a role in explaining cooperation in other organisms (Section 5.2.2). Importantly, this fine-tuning of behaviour can be done in response to both previous experience (learning) and observations of others (social learning). Furthermore, this has allowed the extreme division of labour that is observed in human societies.

While many organisms have impressive proximate mechanisms for enforcing cooperation, humans can have both more complex and diverse systems. Mechanisms such as direct and indirect reciprocity can be important in humans, whereas they are thought to be beyond the cognitive abilities of most other animals (Stevens & Hauser, 2004; Stevens, Cushman, & Hauser, 2005). More complex and unique mechanisms to enforce cooperation have arisen in humans, such as contracts, laws, justice, trade and social norms, leading to incredible feats such as the extreme division of labour that keeps large cities or nations going (Binmore, 1994, 1998, 2005b; Boyd & Richerson, 1992; Seabright, 2004; Young, 2003). These mechanisms allow direct benefits to be obtained from cooperation in situations where cheating would otherwise be favoured. To put this into game theoretic terms, such mechanisms allow more efficient equilibria to be reached than would ever be possible in less cognitively advanced species. Cultural evolution allows a potential way in which

different mechanisms or strategies could be tested (Binmore, 2005b; Boyd & Richerson, 1985), and determining how this influences cooperation remains a major outstanding task (Section 6.7).

The above discussion suggests that humans are special because our cognitive abilities mean we are particularly efficient enforcers, which has expanded the range of situations in which cooperation can be favoured. However, we stress that we are not saying that humans have the best cognitive abilities for all behaviours related to cooperation. For example, considering indirect fitness benefits, while social amoebae and social insects are able to adjust their behaviour in response to direct cues of genetic relatedness (Boomsma et al., 2003; Mehdiabadi et al., 2006), humans must rely on indirect learnt cues such as childhood co-residence (Lieberman et al., 2003). Overall, the general point appears to be that, as with other aspects of the mental powers and moral sense, the difference in cooperative behaviours between humans and other animals is "one of degree and not of kind" (Darwin, 1871, pp. 104–106).

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.evolhumbehav.2010.08.001.

Acknowledgments

References

Alcock, J. (2005). *Animal Behavior Eighth Edition*. Sunderland, MA: Sinauer Associates.

Alexander, R. D. (1974). The evolution of social behavior. *Annual Review of Ecological Systematics*, *5*, 325–383.

Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.

Amann, E., & Yang, C. L. (1998). Sophisitication and the persistence of cooperation. *Journal of Economic Behavior & Organization*, *37*, 91–105.

Aoki, K. (1982). A condition for group selection to prevail over counteracting individual selection. *Evolution*, *36*, 832–842.

Arnold, J. E. (1993). Labor and the rise of complex hunter-gatherers. *Journal of Anthropological Archaeology*, *12*, 75–119.

Arrow, H. (2007). The sharp end of altruism. *Science*, *318*, 581–582.

Aumann, R. (1981). Survey of repeated games. In V. Bohm (Ed.), *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern* (pp. 11–42). Mannheim: Bibliographisches Institut.

Aumann, R., & Maschler, M. (1995). *Repeated Games with Incomplete Information*. Cambridge, Mass: MIT Press.

Avilés, L. (1993). Interdemic selection and the sex ratio: a social spider perspective. *American Naturalist*, *142*, 320–345.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Baschetti, R. (2007). Evolutionary, neurobiological, gene-based solution of the ideological "puzzle" of human altruism and cooperation. *Medical Hypotheses*, *69*, 241–249.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412–414.

Becker, G. S. (1974). Altruism, egoism, and genetic fitness: economics and sociobiology. *Journal of Economic Literature*, *14*, 817–826.

Bergmüller, R., Bshary, R., Johnstone, R. A., & Russell, A. F. (2007). Integrating coopeative breeding and cooperation theory. *Behavioural Processes*, *76*, 61–72.

Bergstrom, T. C. (1995). On the evolution of altruistic ethical rules for siblings. *The American Economic Review*, *85*, 58–81.

Bergstrom, T. C. (1996). Economics in a family way. *Journal of Economic Literature*, *34*, 1903–1934.

Bergstrom, T. C. (2002). Evolution of social behavior: individual and group selection. *Journal of Economic Perspectives*, *16*, 67–88.

Bernasconi, G., & Strassmann, J. E. (1999). Cooperation among unrelated individuals: the ant foundress case. *Trends in Ecology & Evolution*, *14*, 477–482.

Binmore, K. (1994). *Game Theory And The Social Contract Volume 1: Playing Fair*. Cambridge, MA: MIT Press.

Binmore, K. (1998). *Game Theory And The Social Contract Volume 2: Just Playing*. Cambridge, MA: MIT Press.

Binmore, K. (2005a). Economic man or straw man? *Behavioral and Brain Sciences*, *28*, 817–818.

Binmore, K. (2005b). *Natural Justice*. Oxford: Oxford University Press.

Binmore, K. (2006). Why do people cooperate? *Politics, Philosophy & Economics*, *5*, 81–96.

Binmore, K. (2007). *Game Theory: A Very Short Introduction*. Oxford: Oxford University Press.

Bonner, J. T. (1967). *The Cellular Slime Molds*. Princeton: Princeton University Press.

Boomsma, J. J., Nielsen, J., Sundstrom, L., Oldham, N. J., Tentschert, J., Petersen, H. C., et al. (2003). Informational constraints on optimal sex allocation in ants. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 8799–8804.

Bourke, A. F. G., & Franks, N. R. (1995). *Social Evolution in Ants. Princeton*. New Jersey: Princeton University Press.

Bowles, S. (2006). Group competition, reproductive levelling and the evolution of human altruism. *Science*, *314*, 1569–1572.

Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, *324*, 1293–1298.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology*, *65*, 17–28.

Bowles, S., & Gintis, H. (2008). Cooperation. In L. Blume, & S. Durlauf (Eds.), *The New Palgrave Dictionary of Economics*. MacMilan.

Bowles, S., Choi, J. K., & Hopfensitz, A. (2003). The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology*, *223*, 135–147.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R., & Richerson, P. J. (1990). Group selection among alternaive evolutionarily stable strategies. *Journal of Theoretical Biology*, *145*, 331–342.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171–195.

Boyd, R., & Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a stuctured population. *Journal of Theoretical Biology*, *215*, 287–296.

Boyd, R., & Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Boyd, R., & Richerson, P. J. (2006). Culture and the evolution of the human social instincts. In S. Levinson, & N. Enfield (Eds.), Oxford: Berg.

Boyd, R., Richerson, P. J., & Henrich, J. (Submitted). Rapid cultural adaptation can facilitate the evolution of large-scale cooperations.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 3531–3535.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2005). The evolution of altruistic punishment. In H. Gintis (Ed.), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (pp. 215–227). Cambridge, MA: MIT Press.

Brewer, M. B., & Caporael, L. R. (1990). Selfish genes vs. selfish people: sociobiology as origin myth. *Motivation and Emotion*, *14*, 237–243.

Brockhurst, M. A., Buckling, A., & Gardner, A. (2007). Cooperation peaks at intermediate disturbance. *Current Biology*, *17*, 761–765.

Brown, S. P. (2001). Collective action in an RNA virus. *Journal of Evolutionary Biology*, *14*, 821–828.

Bshary, R. (2002). Biting cleaner fish use altruism to deceive image-scoring client reef fish. *Proceedings of the Royal Society of London B*, *269*, 2087–2093.

Bshary, R., & Grutter, A. S. (2002). Asymetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, *63*, 547–555.

Bshary, R., & Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, *63*, 557–564.

Bshary, R., & Grutter, A. S. (2005). Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters*, *1*, 396–399.

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

Burt, A., & Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish genetic Elements*. Cambridge, Massachusetts: Harvard University Press.

Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Cooperation in humans: competition between groups and proximate emotions. *Evolution and Human behavior*, *31*, 104–108.

Choi, J. K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, *318*, 636–640.

Clutton-Brock, T. H. (2002). Breeding together: kin selection, reciprocity and mutualism in cooperative animal societies. *Science*, *296*, 69–72.

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

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

Cockburn, A. (1998). Evolution of helping behavior in cooperatively breeding birds. *Annual Review of Ecological Systematics*, *29*, 141–177.

Colwell, R. K. (1981). Group selection is implicated in the evolution of female-biased sex ratios. *Nature*, *290*, 401–404.

Cornwallis, C., West, S. A., & Griffin, A. S. (2009). Routes to cooperatively breeding vertebrates: kin discrimination and limited dispersal. *Journal of Evolutionary Biology*, *22*, 2245–2457.

Cornwallis, C., West, S. A., Davies, K. E., & Griffin, A. S. (2010). Promiscuity and the evolutionary transition to complex societies. *Nature*, *466*, 969–972.

Crozier, W. R. (2001). Blushing and the exposed self: Darwin revisited. *Journal for the Theory of Social Behaviour*, *31*, 61–72.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London, UK: John Murray.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Davies, N. B. (1992). *Dunnock Behaviour and Social Evolution*. Oxford: Oxford University Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1979). Twelve misunderstandings of kin selection. *Zeitschrift fur Tierpsychologie*, *51*, 184–200.

Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University Press.

de Waal, F. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, *59*, 279–300.

Denison, R. F., Kiers, E. T., & West, S. A. (2003). Darwinian agriculture: when can humans find solutions beyond the reach of natural selection. *Quarterly Review of Biology*, *78*, 145–168.

Diggle, S. P., Griffin, A. S., Campbell, G. S., & West, S. A. (2007). Cooperation and conflict in quorum-sensing bacterial populations. *Nature*, *450*, 411–414.

El Mouden, C., West, S. A., & Gardner, A. (2010). The enforcement of cooperation by policing. *Evolution*, *64*, 2139–2152.

Emlen, S. T. (1997). Predicting family dynamics in social vertebrates. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural Ecology* (pp. 228–253). Oxford, UK: Blackwell.

Eshel, I. (1972). On the neighbor effect and the evolution of altruistic traits. *Theoretical Population Biology*, *3*, 258–277.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, *422*, 137–140.

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

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation* (pp. 55–82). Cambridge, MA: MIT Press.

Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology*, *14*, 784–790.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63–87.

Fehr, E., & Fischbacher, U. (2005a). Altruists with green beards. *Analyse & Kritik*, *27*, 73–84.

Fehr, E., & Fischbacher, U. (2005b). Human altruism—proximate patterns and evolutionary origins. *Analyse & Kritik*, *27*, 6–47.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, *13*, 1–25.

Feldman, M. W., Cavalli-Sforza, L. L., & Peck, J. R. (1985). Gene-culture coevolution: models for the evolution of altruism with cultural transmission. *Proceedings of the National Academy of Sciences of the United States of America*, *82*, 5814–5818.

Field, A. J. (2001). *Altruistically inclined? The behavioral sciences, evolutionary theory, and the origins of reciprocity*. Michigan: University of Michigan Press.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon.

Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics*, *11*, 53–63.

Foster, K. R., Wenseleers, T., & Ratnieks, F. L. W. (2006). Kin selection is the key to altruism. *Trends in Ecology & Evolution*, *21*, 57–60.

Frank, R. H. (1987). If Homo economicus could choose hos own utility function, would he want one with a conscience? *The American Economic Review*, 593–604.

Frank, S. A. (1986). Hierarchical selection theory and sex ratios. I. General solutions for structured populations. *Theoretical Population Biology*, *29*, 312–342.

Frank, S. A. (1995a). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, *377*, 520–522.

Frank, S. A. (1995b). George Price's contributions to evolutionary genetics. *Journal of Theoretical Biology*, *175*, 373–388.

Frank, S. A. (1996a). Policing and group cohesion when resources vary. *Animal Behaviour*, *52*, 1163–1169.

Frank, S. A. (1996b). Models of parasite virulence. *Quarterly Review of Biology*, *71*, 37–78.

Frank, S. A. (1997). Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *Journal of Theoretical Biology*, *189*, 307–316.

Frank, S. A. (1998). *Foundations of Social Evolution*. Princeton: Princeton University Press.

Frank, S. A. (2003). Repression of competition and the evolution of cooperation. *Evolution*, *57*, 693–705.

Frank, S. A. (2010). A general model of the public goods dilemma. *Journal of Evolutionary Biology*, *23*, 1245–1250.

Friedman, M. (1953). The methodology of positive economics. In M. Friedman (Ed.), *Essays in Positive Economics* (pp. 3–43). Chicago: University of Chicago Press.

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*, 533–554.

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: previou insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society of London Series B*, *364*, 791–806.

Gächter, S., & Herrmann, B. (2006). The limits of self-governance when cooperators get punished: experimental evidence from urban and rural Russia. Centre for Decision Research and Experimental Economics. Discussion Paper 13. Available at: http://ideas.repec.org/p/cdx/dpaper/2007-11.html

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, *322*, 1510.

Gale, J., Binmore, K. G., & Samuelson, L. (1995). Learning to be imperfect: the ultimatum game. *Games and Economic Behavior*, *8*, 56–90.

Gardner, A. (2008). The Price equation. *Current Biology*, *18*, R198–R202.

Gardner, A. (2009). Adaptation as organism design. *Biology Letters*, *5*, 861–864.

Gardner, A., & West, S. A. (2004). Cooperation and punishment, especially in humans. *American Naturalist*, *164*, 753–764.

Gardner, A., & Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, *22*, 659–671.

Gardner, A., & West, S. A. (2010). Greenbeards. *Evolution*, *64*, 25–38.

Gardner, A., West, S. A., & Buckling, A. (2004). Bacteriocins, spite and virulence. *Proceedings of the Royal Society of London, Series B*, *271*, 1529–2535.

Gardner, A., West, S. A., & Barton, N. H. (2007a). The relation between multilocus population genetics and social evolution theory. *American Naturalist*, *169*, 207–226.

Gardner, A., Arce, A., & Alpedrin, J. (2009). Budding dispersal and the sex ratio. *Journal of Evolutionary Biology*, *22*, 1036–1045.

Gardner, A., West, S.A., and Wild, G. (Submitted). The genetical theory of kin selection. *Journal of Evolutionary Biology*.

Gardner, A., Hardy, I. C. W., Taylor, P. D., & West, S. A. (2007b). Spiteful soldiers and sex ratio conflict in polyembryonic parasitoid wasps. *American Naturalist*, *169*, 519–533.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Clarendon Press.

Gilbert, O. M., Foster, K. R., Mehdiabadi, N. J., Strassmann, J. E., & Queller, D. C. (2007). High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 8913–8917.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*, 169–179.

Gintis, H. (2003). The hitchhiker's guide to altruism: gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology*, *220*, 407–418.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signalling and cooperation. *Journal of Theoretical Biology*, *213*, 103–119.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, *24*, 153–172.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2005a). *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, Massachusetts: MIT Press.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2005b). Moral sentiments and material interests: origins, evidence, and consequences. In H. Gintis (Ed.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 3–39). Cambridge, Massachusetts: MIT Press.

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research*, *21*, 241–253.

Goto, R., Okamoto, T., Kiers, E. T., Kawakita, A., & Kato, M. (2010). Selective flower abortion maintains moth cooperation in a newly discovered pollination mutualism. *Ecology Letters*, *13*, 321–329.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, *298*, 425–426.

Grafen, A. (1984). Natural selection, kin selection and group selection. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 62–84). Oxford, UK: Blackwell Scientific Publications.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology*, *2*, 28–89.

Grafen, A. (1990a). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Grafen, A. (1990b). Do animals really recognise kin? *Animal Behaviour*, *39*, 42–54.

Grafen, A. (1991). Modelling in behavioural ecology. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology, an evolutionary approach* (pp. 5–31). Oxford: Blackwell.

Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proceedings of the Royal Society of London, Series B*, *266*, 799–803.

Grafen, A. (2002). A first formal link between the Price equation and an optimization program. *Journal of Theoretical Biology*, *217*, 75–91.

Grafen, A. (2006a). Optimisation of inclusive fitness. *Journal of Theoretical Biology*, *238*, 541–563.

Grafen, A. (2006b). A theory of Fisher's reproductive value. *Journal of Mathematical Biology*, *53*, 15–60.

Grafen, A. (2007a). An inclusive fitness analysis of altruism on a cyclical network. *Journal of Evolutionary Biology*, *20*, 2278–2283.

Grafen, A. (2007b). The formal Darwinism project: a mid-term report. *Journal of Evolutionary Biology*, *20*, 1243–1254.

Grafen, A. (2007c). Detecting kin selection at work using inclusive fitness. *Proceedings of the Royal Society of London, Series B*, *274*, 713–719.

Grafen, A. (2009). Formalizing Darwunism and inclusive fitness theory. *Philosophical Transactions of the Royal Society of London Series B*, *364*, 3135–3141.

Grafen, A., & Archetti, M. (2008). Natural selection of altruism in enelastic viscous homogeneous populations. *Journal of Theoretical Biology*, *252*, 694–710.

Griffin, A. S., & West, S. A. (2002). Kin selection: fact and fiction. *Trends in Ecology & Evolution*, *17*, 15–21.

Griffin, A. S., & West, S. A. (2003). Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science*, *302*, 634–636.

Griffin, A. S., West, S. A., & Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature*, *430*, 1024–1027.

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theoretical Population Biology*, *69*, 339–348.

Haig, D. (2002). *Genomic Imprinting and Kinship*. New Brunswick, NJ: Rutgers University.

Haldane, J. B. S. (1932). *The causes of evolution*. New York: Longmans.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, *26*, 245–256.

Hamilton, W. D. (1963). The evolution of altruistic behaviour. *American Naturalist*, *97*, 354–356.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, 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. (1971). Selection of selfish and altruistic behaviour in some extreme models. In J. F. Eisenberg, & W. S. Dillon (Eds.), *Man and beast: Comparative social behavior* (pp. 57–91). Washington, DC: Smithsonian Press.

Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of Ecological Systematics*, 3, 193–232.

Hamilton, W. D. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In R. Fox (Ed.), *Biosocial anthropology* (pp. 133–155). New York: Wiley.

Hamilton, W. D. (1996). *Narrow roads of gene land: I Evolution of social behaviour*. Oxford: W.H. Freeman.

Hamilton, W. D., & May, R. (1977). Dispersal in stable habitats. *Nature*, 269, 578–581.

Hammerstein, P. (2003). *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243–1248.

Harvey, P. H., Partridge, L., & Nunney, L. (1985). Group selection and the sex ratio. *Nature*, 313, 10–11.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: an integrative evolutionary model of cognitive biases. Personality and social. *Psychology Review*, 10, 47–66.

Helanterä, H., & Bargum, K. (2007). Pedigree relatedness, not greenbeard genes, explains eusociality. *Oikos*, 116, 217–220.

Helanterä, H., & Sundström, L. (2007). Worker policing and nest mate recognition in the ant Formica fusca. *Behavioral Ecology and Sociobiology*, 61, 1143–1149.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53, 3–35.

Henrich, J., & Boyd, R. (2001). Why people punish defectors. *Journal of Theoretical Biology*, 208, 79–89.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., & Bolyanatz, A, et al. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H., et al. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28, 795–855.

Henrich, N., & Henrich, J. (2007). *Why humans cooperate: a cultural and evolutionary explanation*. New York: Oxford University Press.

Herre, E. A. (1987). Optimality, plasticity and selective regime in fig wasp sex ratios. *Nature*, 329, 627–629.

Herre, E. A. (1993). Population structure and the evolution of virulence in nematode parasites of fig wasps. *Science*, 259, 1442–1445.

Herre, E. A., Machado, C. A., & West, S. A. (2001). Selective regime and fig wasp sex ratios: Towards sorting rigor from pseudo-rigor in tests of adaptation. In S. Orzack, & E. Sober (Eds.), *Adaptionism and optimality* (pp. 191–218). Cambridge: University Press.

Herre, E. A., Knowlton, N., Mueller, U. G., & Rehner, S. A. (1999). The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution*, 14, 49–53.

Houser, D., & Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *The American Economic Review*, 92, 1062–1069.

Howard, N. (1971). *Paradoxes of rationality: theory of metagames and political behavior*. Cambridge, Massachusetts: MIT Press.

Hughes, W. O. H., Oldroyd, B. P., Beekman, M., & Ratnieks, F. L. W. (2008). Ancestral monogamy shows kin seelction is the key to the evolution of eusociality. *Science*, 320, 1213–1216.

Jander, K. C., & Herre, E. A. (2010). Host sanctions and pollinator cheating in the fig tree-fig wasp mutualism. *Proceedings of the Royal Society of London, Series B, Proc R Soc B*, 277, 1481–1488.

Jennions, M. D., & Macdonald, D. W. (1994). Cooperative breeding in mammals. *Trends in Ecology & Evolution*, 9, 89–93.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies*, 59, 63–80.

Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume-rhizobium mutualism. *Nature*, 425, 78–81.

Killingback, T., Bieri, J., & Flatt, T. (2006). Evolution in group-structured populations can resolve the tragedy of the ccommons. *Proceedings of the Royal Society of London, Series B*, 273, 1477–1481.

Knight, C. (1998). Ritual/speech coevolution: a solution to the problem of deception. In J. R. Hurford (Ed.), *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press.

Knight, C. (2008). "Honest fakes" and language origins. *Journal of Consciousness Studies*, 15, 236–248.

Kohn, M. (2008). The needs of the many. *Nature*, 456, 296–299.

Kokko, H., Johnstone, R. A., & Clutton-Brock, T. H. (2001). The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society of London, Series B*, 268, 187–196.

Krebs, J. R., & McCleery, R. H. (1984). Optimization in behavioural ecology. In J. R. Krebs, & N. B. Davies (Ed.), *Behavioural ecology, 2nd edn.* Oxford: Blackwell.

Krebs, J. R., & Davies, N. B. (1987). *An introduction to behavioural ecology*. Second Edition. Oxford, England: Blackwell Scientific Publications.

Krebs, J. R., & Davies, N. B. (1993). *An introduction to behavioural ecology*. Third Edition. Oxford, England: Blackwell Scientific Publications.

Krebs, J. R., & Davies, N. B. (1997). *Behavioural ecology. An evolutionary approach*. 4th Ed. Oxford: Blackwell Scientific.

Kruuk, H. (2003). *Niko's nature*. Oxford: Oxford University Press.

Kümmerli, R., Gardner, A., West, S. A., & Griffin, A. S. (2009). Limited dispersal, budding dispersal and cooperation: an experimental study. *Evolution*, 63, 939–949.

Kümmerli, R., Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games. *Proceedings of the National Academy of Sciences USA*, 107, 10125–10130.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human behavior*, 28, 75–84.

Lack, D. (1966). *Population Studies of Birds*. Oxford: Clarendon Press.

Laury, S. K., & Holt, C. A. (2008). Voluntary provision of public goods: experimental results with interior Nash equilibria. *Handbook of Experimental Economics Results*, Volume I (pp. 792–80). Amsterdam: Elsevier.

Ledyard, J. O. (1995). Public goods: a survey of experimental research. In J. H. Kagel, & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.

Lehmann, L., & Keller, L. (2006). The evolution of cooperation and altruism. A general framework and classification of models. *Journal of Evolutionary Biology*, 19, 1365–1378.

Lehmann, L., & Feldman, M. W. (2008a). War and the evolution of belligerance and bravery. *Proceedings of the Royal Society of London, Series B Proc Biol Sci*, 2008, 2877–2885.

Lehmann, L., & Feldman, M. W. (2008b). The co-evolution of culturally inherited altruistic helping and cultural transmission under random group formation. *Theoretical Population Biology*, 73, 506–516.

Lehmann, L., Keller, L., & Sumpter, D. J. T. (2007a). Inclusive fitness explains the evolution of helping and harming behaviors on graphs. *Journal of Evolutionary Biology*, 20, 2284–2295.

Lehmann, L., Feldman, M. W., & Foster, K. R. (2008). Cultural transmission can inhibit the evolution of altruistic helping. *American Naturalist*, 172, 12–24.

Lehmann, L., Keller, L., West, S. A., & Roze, D. (2007b). Group selection and kin selection. Two concepts but one process. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6736–6739.

Lehmann, L., Rousset, F., Roze, D., & Keller, L. (2007c). Strong-reciprocity or strong-ferocity? A population genetic view of the evolution of altruistic punishment. *American Naturalist*, 170, 21–36.

Leigh, E. G. (1971). *Adaptation and Diversity*. San Francisco: Freeman, Cooper and Company.

Leigh, E. G. (1983). When does the good of the group override the advantage of the individual? *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 2985–2989.

Leigh, E. G. (1991). Genes, bees and ecosystems: the evolution of a common interest among individuals. *Trends in Ecology & Evolution*, *6*, 257–262.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*, 153–174.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society of London, Series B*, *270*, 819–826.

Lieberman, E., Hauert, C., & Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, *433*, 312–316.

Luce, R., & Raiffa, H. (1957). *Games and Decisions*. New York: Wiley.

MacColl, A. D. C., & Hatchwell, B. J. (2004). Determinants of lifetime fitness in a cooperative breeder, the long tailed tit *Aegithalos caudatus*. *Journal of Animal Ecology*, *73*, 1137–1148.

MacLean, R. C., & Gudelj, I. (2006). Resource competition and social conflict in experimental populations of yeast. *Nature*, *441*, 498–501.

Mailah, G. J., & Samuelson, L. (2006). *Repeated games and reputations: Long-run relationships*. New York: Oxford University Press.

Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economic Perspectives*, *14*, 115–136.

Mateo, J. M. (2002). Kin-recognition abilities and nepotism as a function of sociality. *Proceedings of the Royal Society of London, Series B*, *269*, 721–727.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, *201*, 1145–1147.

Maynard Smith, J. (1976). Group selection. *Quarterly Review of Biology*, *51*, 277–283.

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

Maynard Smith, J. (1983). Models of evolution. *Proceedings of the Royal Society of London, Series B*, *219*, 315–325.

Maynard Smith, J. (1995). Life at the edge of chaos? *The New York review*, 28–30.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*, 15–18.

Maynard Smith, J., & Szathmary, E. (1995). *The major transitions in evolution*. Oxford: W.H. Freeman.

Mayr, E. (1961). Cause and effect in biology. *Science*, *134*, 1501–1506.

McElreath, R., & Henrich, J. (2006). Modeling cultural evolution. In R. I. M. Dunvar, & Barret (Eds.), *Oxford Handbook of Evolutionary Biology*.

Mehdiabadi, N. J., Jack, C. N., Farnham, T. T., Platt, T. G., Kalla, S. E., Shaulsky, G., et al. (2006). Kin preference in a social microbe. *Nature*, *442*, 881–882.

Michod, R. E., & Hamilton, W. D. (1980). Coefficients of relatedness in sociobiology. *Nature*, *288*, 694–697.

Milinski, M., & Wedekind, C. (1998). Working memory constrains human cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 13755–13758.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Donors to charity gain in both indirect reciprocity and political reputation. *Proceedings of the Royal Society of London, Series B*, *269*, 881–883.

Mock, D. W., & Parker, G. A. (1997). *The evolution of sibling rivalry*. Oxford: Oxford University Press.

Mulder, R. A., & Langmore, N. E. (1993). Dominant males punish helpers for temporary defectrion in superb fairy wrens. *Animal Behaviour*, *45*, 830–833.

Murray, M. G. (1985). Figs (*Ficus* spp.) and fig wasps (*Chalcidoidea*, *Agaonidae*): hypotheses for an ancient symbiosis. *Biological Journal of the Linnean Society*, *26*, 69–81.

Nettle, D. (2009). *Evolution and genetics for psychology*. Oxford: Oxford University Press.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560–1563.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573–577.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, *289*, 1773–1775.

Nowak, M. A., Tarnita, C. E., & Antal, T. (2010). Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society of London Series B*, *365*, 19–30.

O'Gorman, R., Sheldon, K. M., & Wilson, D. S. (2008). For the good of the group? Exploring group-level evolutionary adaptations using multilevel selection theory. *Group Dynamics: Theory, Research, and Practise*, *12*, 17–26.

Oates, K., & Wilson, M. (2001). Nominal kinship cues facilitate altruism. *Proceedings of the Royal Society of London, Series B*, *269*, 105–109.

Okasha, S. (2004). Multilevel selection and the partitioning of covariance: a comparison of three approaches. *Evolution*, *58*, 486–494.

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford, UK: Oxford University Press.

Orlove, M. J., & Wood, C. L. (1978). Coefficients of relationship and coefficients of relatedness in kin seelction: a covariance form for the rho formula. *Journal of Theoretical Biology*, *73*, 679–686.

Owren, M. J., & Bachorowski, J. A. (2006). The evolution of emotional expression: a "selfish-gene" account of smiling and laughter in early hominids and humans. In T. J. Mayne, & G. A. Bonanno (Eds.), *Emotions: Current Issues and Future Directions*. New York: The Guilford Press.

Palameta, B., & Brown, W. M. (1999). Human cooperation is more than by-product mutualism. *Animal Behaviour*, *57*, F1–F3.

Paley, W. (1802). *Natural Theology*. London: Wilks & Taylor.

Palmer, C. T., Fredrickson, B. E., & Tilley, C. F. (1997). Categories and gatherings: group selection and the mythology of cultural anthropology. *Evolution and Human Behavior*, *18*, 291–308.

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

Partridge, L., & Sibley, R. (1991). Constraints in the evolution of life histories. *Philosophical Transactions of the Royal Society of London Series B*, *332*, 3–13.

Pellmyr, O., & Huth, C. J. (1994). Evolutionary stability of mutualism between yuccas and yucca moths. *Nature*, *372*, 257–260.

Pfeiffer, T., Schuster, S., & Bonhoeffer, S. (2001). Cooperation and competition in the evolution of ATP-producing pathways. *Science*, *292*, 504–507.

Pfennig, D. W., Collins, J. P., & Ziemba, R. E. (1999). A test of alternative hypotheses for kin recognition in cannibalistic tiger salamanders. *Behavior Ecology*, *10*, 436–443.

Platt, T. G., & Bever, J. D. (2009). Kin competition and the evolution of cooperation. *Trends in Ecology & Evolution*, *24*, 370–377.

Pompilio, L., Kacelnik, A., & Behmer, S. T. (2006). State-dependent learned valuation drives choice in an invertebrate. *Science*, *311*, 1613–1615.

Pradel, J., Euler, H. A., & Fetchenhauer, D. (2009). Spotting altruistic dictator game players and mingling with them: the elective assortation of classmates. *Evolution and Human behavior*, *30*, 103–113.

Provine, W. B. (2001). *The origins of theoretical population genetics*. Chicago: Chicago University Press.

Puurtinen, M., & Mappes, T. (2009). Between-group competition and human cooperation. *Proceedings of the Royal Society of London, Series B*, *276*, 355–360.

Queller, D. C. (1992a). Quantitative genetics, inclusive fitness, and group selection. *American Naturalist*, *139*, 540–558.

Queller, D. C. (1992b). Does population viscosity promote kin selection? *Trends in Ecology & Evolution*, *7*, 322–324.

Queller, D. C. (1992c). A general model for kin selection. *Evolution*, *46*, 376–380.

Queller, D. C. (2000). Relatedness and the fraternal major transitions. *Philosophical Transactions of the Royal Society of London Series B*, *355*, 1647–1655.

Queller, D. C. (2004). Kinship is relative. *Nature*, *430*, 975–976.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*, 258–275.

Queller, D. C., Zacchi, F., Cervo, R., Turillazzi, S., Henshaw, M. T., Santorelli, L. A., et al. (2000). Unrelated helpers in a social insect. *Nature*, *405*, 784–787.

Quervain, J. F. d., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma*. Ann Arbor: University of Chicago Press.

Ratnieks, F. L. W. (1988). Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *American Naturalist*, *132*, 217–236.

Ratnieks, F. L. W., & Visscher, P. K. (1989). Worker policing in the honeybee. *Nature*, *342*, 796–797.

Ratnieks, F. L. W., Foster, K. R., & Wenseleers, T. (2006). Conflict resolution in insect societies. *Annual Review of Entomology*, *51*, 581–608.

Reeve, H. K., & Hölldobler, B. (2007). The emergence of a superorganism through intergroup competition. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 9736–9740.

Richerson, P. J., & Boyd, R. (1999). The evolutionary dynamics of a crude superorganism. *Human Nature*, *10*, 253–289.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone*. Chicago: University of Chicago Press.

Richerson, P. J., & Boyd, R. (1999). Complex societies: The evolutionary origins of a crude superorganism. *Human Nature*.

Robson, A. J. (1990). Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *Journal of Theoretical Biology*, *144*, 379–396.

Robson, A. J. (2008). Group selection. In S. Durlauf, & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics. Second Edition*. Palgrave Macmillan.

Rogers, A. R. (1990). Group selection by selective emigration: the effects of migration and kin structure. *American Naturalist*, *135*, 398–413.

Rose, M. R., & Lauder, G. V. (Eds.), (1996). *Adaptation.* San Diego: Academic Press.

Rousset, F. (2004). *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.

Rousset, F., & Roze, D. (2007). Constraints on the origin and maintenance of genetic kin recognition. *Evolution*, *61*, 2320–2330.

Russell, A. F., & Hatchwell, B. J. (2001). Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proceedings of the Royal Society of London, Series B*, *268*, 2169–2174.

Russell, A. F., & Wright, J. (2008). Avian mobbing: byproduct mutualism is not reciprocal altruism. *Trends in Ecology & Evolution.*

Sachs, J. L., Mueller, U. G., Wilcox, T. P., & Bull, J. J. (2004). The evolution of cooperation. *Quarterly Review of Biology*, *79*, 135–160.

Saijo, T., & Nakamura, H. (1995). The "spite" dilemma in voluntary contribution mechanism experiments. *Journal of Conflict Resolution*, *39*, 535–560.

Schloss. (2002). Emerging accounts of altruism. In S. G. Post (Ed.), *Altruism and Altruitsic Love* (pp. 212–242). Oxford: Oxford University Press.

Schuessler. (1989). Exit threats and cooperation under anonymity. *Journal of Conflict Resolution*, *33*, 728–749.

Scott-Phillips, T. C. (2007). The social evolution of language, and the language of social evolution. *Evolutionary Psychology*, *5*, 740–753.

Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of Evolutionary Biology*, 21.

Seabright, P. (2004). *The company of strangers: A natural history of economic life*. Princeton: Princeton University Press.

Semmann, D., Krambeck, H. J., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology & Sociobiology*, *56*, 248–252.

Sharp, S. P., McGowan, A., Wood, M. J., & Hatchwell, B. J. (2005). Learned kin recognition cues in a social bird. *Nature*, *434*, 1127–1130.

Shennan, S. (2002). *Genes, memes and human history*. London: Thames and Hudson.

Sherman, P. W. (1977). Nepotism and the evolution of alarm calls. *Science*, *197*, 1246–1253.

Silk, J. B. (2002). Kin selection in primate groups. *International Journal for Parasitology*, *23*, 849–875.

Smith, J. (2001). The social evolution of bacterial pathogenesis. *Proceedings of the Royal Society of London, Series B*, *268*, 61–69.

Smuts, B. (1999). Multilevel selection, cooperation, and altruism. *Human Nature*, *10*, 311–327.

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

Soltis, J., Boyd, R., & Richerson, P. J. (1995). Can group-functional behaviours evolve by cultural group-selection? An empirical test. *Current Anthropology*, *36*, 473–493.

Spence, A. M. (1973). Job market signalling. *Quarterly Journal of Economics*, *90*, 225–243.

Stearns, S. C. (1992). *Evolution of life histories*. Oxford: Oxford University Press.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*, 60–65.

Stevens, J. R., Cushman, F. A., & Hauser, M. D. (2005). Evolving the psychological mechanisms for cooperation. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 499–518.

Taylor, C., & Nowak, M. A. (2007). Transforming the dilemma. *Evolution*, *61*, 2281–2292.

Taylor, P. D. (1990). Allele-frequency change in a class structured population. *American Naturalist*, *135*, 95–106.

Taylor, P. D. (1996). Inclusive fitness arguments in genetic models of behaviour. *Journal of Mathematical Biology*, *34*, 654–674.

Taylor, P. D., & Frank, S. A. (1996). How to make a kin selection model. *Journal of Theoretical Biology*, *180*, 27–37.

Taylor, P. D., Day, T., & Wild, G. (2007a). Evolution of cooperation in a finite homogeneous graph. *Nature*, *447*, 469–472.

Taylor, P. D., Wild, G., & Gardner, A. (2007b). Direct fitness or inclusive fitness: how shall we model kin selection. *Journal of Evolutionary Biology*, *20*, 301–309.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift fur Tierpsychologie*, *20*, 410–433.

Traulsen, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10952–10955.

Trivers, R. (2004). Mutual benefits at all levels of life. *Science*, *304*, 964–965.

Trivers, R. (1985). Social Evolution. Benjamin/Cummings, Menlo Park, CA.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Trivers, R. L. (1974). Parent-offspring conflict. *American Zoologist*, *14*, 249–264.

Trivers, R. L. (1998a). Think for yourself. *Skeptic*, *6*, 86–87.

Trivers, R. L. (1998b). As they would do to you. *Skeptic*, *6*, 81–83.

van Rooy, R. (2003). Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. In M. Tennenholtz (Ed.), *Proc. of the 4th Conference on Theoretical Aspects of Rationality & Knowledge* (pp. 45–58).

Velicer, G. J., Kroos, L., & Lenski, R. E. (2000). Developmental cheating in the social bacterium *Myxococcus xanthus*. *Nature*, *404*, 598–601.

Wade, M. J. (1985). Soft selection, hard selection, kin selection, and group selection. *American Naturalist*, *125*, 61–73.

Warneken, F., Hare, B., Melis, A. P., Hanus, D., & Tomasello, M. (2007). Spontaneous altruism by chimpanzees and young children. *PLOS Biology*, *e184*, 5.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, *288*, 850–852.

Wehner, R. (1987). 'Matched filters'—neural models of the external world. *Journal of Comparative Physiology A*, *161*, 511–531.

Wenseleers, T., Helantera, H., Hart, A., & Ratnieks, F. L. W. (2004). Worker reproduction and policing in insect societies: an ESS analysis. *Journal of Evolutionary Biology*, *17*, 1035–1047.

West, S. A. (2009). *Sex allocation*. Princeton: Princeton University Press.

West, S. A., & Gardner, A. (2010). Altruism, spite and greenbeards. *Science*, *327*, 1341–1344.

West, S. A., Pen, I., & Griffin, A. S. (2002a). Cooperation and competition between relatives. *Science*, *296*, 72–75.

West, S. A., Griffin, A. S., & Gardner, A. (2007a). Evolutionary explanations for cooperation. *Current Biology*, *17*, R661–R672.

West, S. A., Griffin, A. S., & Gardner, A. (2007b). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*, 415–432.

West, S. A., Griffin, A. S., & Gardner, A. (2008). Social semantics: How useful has group selection been? *Journal of Evolutionary Biology*, *21*, 374–385.

West, S. A., Kiers, E. T., Pen, I., & Denison, R. F. (2002b). Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology*, *15*, 830–837.

West, S. A., Griffin, A. S., Gardner, A., & Diggle, S. P. (2006a). Social evolution theory for microbes. *Nature Reviews Microbiology*, *4*, 597–607.

West, S. A., Gardner, A., Shuker, D. M., Reynolds, T., Burton-Chellow, M., & Sykes, E. M., et al. (2006b). Cooperation and the scale of competition in humans. *Current Biology*, *16*, 1103–1106.

Westneat, D. F., & Fox, C. W. (2010). *Evolutionary Behavioral Ecology*. Oxford: Oxford University Press.

Whitlock, M. C., Davis, B. H., & Yeaman, S. (2007). The costs and beenfits of resource sharing: reciprocity requires resource heterogeneity. *Journal of Evolutionary Biology*, *20*, 1772–1782.

Wild, G., Gardner, A., & West, S. A. (2009). Adaptation and the evolution of parasite virulence in a connected world. *Nature*, *459*, 983–986.

Wiley, R. H., & Rabenold, K. N. (1984). The evolution of cooperative breeding by delayed reciprocity and queuing for favorable social positions. *Evolution*, *38*, 609–621.

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. 308, 181–184.

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

Wilson, D. S. (1975a). A theory of group selection. *Proceedings of the National Academy of Sciences of the United States of America*, *72*, 143–146.

Wilson, D. S. (1977). Structured demes and the evolution of group advantageous traits. *American Naturalist*, *111*, 157–185.

Wilson, D. S. (2008). Social semantics: towards a genuine pluralism in the study of social behaviour. *Journal of Evolutionary Biology*, *21*, 368–373.

Wilson, D. S., & Colwell, R. K. (1981). The evolution of sex ratio in structured demes. *Evolution*, *35*, 882–897.

Wilson, D. S., & Dugatkin, L. A. (1997). Group selection and assortative interactions. *American Naturalist*, *149*, 336–351.

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Quarterly Review of Biology* , 327–348.

Wilson, D. S., & O'Brien, D. T. (2009). Evolutionary theory and cooperation in everyday life. In S. A. Levin (Ed.), *Games, groups, and the global good*. Berlin: Springer-Verlag.

Wilson, D. S., Vugt, M. V., & O'Gorman, R. (2008). Multilevel selection theory and major evolutionary transitions: Implications for psychological science. *Current Directions in Psychological Science*, *17*, 6–9.

Wilson, E. O. (1975b). *Sociobiology*. Cambridge, Massachusetts: Harvard University Press.

Wilson, E. O., & Hölldobler, B. (2005). Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 13367–13371.

Wolf, J. B., Brodie, E. D. I., & Moore, A. J. (1999). Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions. *American Naturalist*, *153*, 254–266.

Woolfenden, G. E. (1975). Florida scrub jay helpers at the nest. *The Auk*, *92*, 1–15.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*, 97–159.

Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. Oliver and Boyd: Edinburgh.

Young, A. J., & Clutton-Brock, T. H. (2006). Infanticide by subordinates influences reproductive sharing in cooperatively breeding meerkats. *Biology Letters*, *2*, 385–387.

Young, A. J., Carlson, A. A., Monfort, S. L., Russell, A. F., Bennett, N. C., & Clutton-Brock, T. H. (2006). Stress and the suppression of subordinate reproduction in cooperatively breeding meerkats. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 12005–12010.

Young, H. P. (2003). The power of norms. In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation*. Cambridge: MIT Press.

Zeder, M. A. (2009). The Neolithic macro-(r)evolution: macroevolutionary theory and the study of culture change. *Journal of Archaeological Research*, *17*, 1–63.