

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 on 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.

Running Head: Cooperation in humans

Text Word Count: 25,450

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. 2007b). 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 2009a). In principle, Darwin's theory of natural selection provides a general framework that has the potential to unite 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, during this time, the leading evolutionary theoreticians have made a poor job of communicating these advances to other areas, such as 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), and which 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). This leads 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 paper is to provide an overview of evolutionary work on cooperation, in a way that is accessible to social scientists, emphasising common misconceptions. In the first part of our paper (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 and its central role in the major evolutionary transitions (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 literature. In the second part of the paper we discuss a number of common misconceptions and sources of confusion from the economics and social science literature, concerning social theory and the problem of cooperation (section 6). Finally, in the third part of the paper we focus on humans, discussing why they cooperate and if they are special (section 7).

2. Evolutionary Theory

Darwin's theory of natural selection explains both the process and the purpose of adaptation (see glossary in Table 1). The process of adaptation is the action of natural selection, which is mediated through the differential reproductive success of individual organisms. Those heritable characters that are associated with greater reproductive success are those that tend to accumulate in natural populations. Thus, Darwin argued that evolved characters will appear designed as if to maximise the individual's reproductive success. It is this that provides the purpose of adaptation, to maximise fitness. This is analogous to the idea in economics that individuals should be self-regarding utility maximisers – in both cases it is not required that individuals are consciously striving to maximise their fitness or utility, just that selection will have led to individuals that do so (Darwin 1859; Friedman 1953). The duality of process and purpose in evolution is captured in Darwin's suggestion that "natural selection can act only through and for the good of each being" (Darwin 1859).

These ideas were later formalised in mathematical terms by Fisher (1930, 1941), who united Darwinism with Mendelian genetics, and described 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, and hence the direct action of natural selection leads to an increase in the mean fitness of the population. This result is termed the 'fundamental theorem of natural selection', and was intended to capture the process (natural selection) and the purpose (maximisation of individual fitness) of adaptation.

Fisher's theorem was frequently misunderstood in the population genetics literature prior to the late 1980's (reviewed by Grafen 2003, 2007a), and this confusion also occurs in the social sciences (e.g. (Weibull 1995)). The first misunderstanding was that the fundamental theorem purports to describe total evolutionary change – it does not, and instead focuses upon the partial change in mean fitness, due to changes in gene frequency, that can be ascribed to the direct action of natural selection, neglecting other non-selective effects that are collectively termed 'deterioration of the environment' (Edwards 1994; Ewens 1989; Fisher 1930; Frank and Slatkin 1992; Grafen 2003, 2007a; Lessard 1997; Price 1972). The second misunderstanding was that the fundamental theorem concerns population fitness – it does not, and instead describes changes in individual fitness, which is expressed relative to the rest of the population (Fisher 1941; Grafen 2003, 2007a). Grafen (1999, 2002, 2007a) has suggested that a better way to capture the optimising action of natural selection is to demonstrate an isomorphism between gene frequency dynamics and an optimisation program, which he has done by forming an 'individual as maximising agent' (IMA) analogy to evolutionary genetics. Maynard Smith's well-known ESS approach is based upon the assumption that individuals behave so as to maximise fitness (Maynard Smith and Price 1973); this central axiom of evolutionary game theory is justified because it is a result that derives from population genetics theory.

Since Darwin, the only fundamental change in our understanding of how natural selection

works has been Hamilton's (Hamilton 1964) development of inclusive fitness theory. 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 (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 (Figure 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 usually referred to as 'kin selection' (Maynard Smith 1964).

The major point here is that inclusive fitness is not just a special case for interactions between relatives. Rather, it is our modern interpretation of Darwinian fitness in its most general form (Grafen 2007a). Specifically, individuals of any species should appear as if they have been designed to maximise their inclusive fitness. Inclusive fitness theory is therefore a theory that explains adaptation at the individual level. Alternative theoretical approaches, such as multilevel selection are not in competition with inclusive fitness theory – they are just different ways of looking at the dynamics by which inclusive fitness is maximised (Gardner and Grafen 2009; Grafen 2006a; Hamilton 1975).

There is a huge literature on the details of inclusive fitness theory, and its widespread application to empirical data, that we do not have space to go into here. 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 and 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 and Wilson 1998). 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, it is just that naive applications of inclusive fitness theory (especially Hamilton's rule) can lead to mistakes in such circumstances (Frank 1998; Gardner et al. 2007a). 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 and 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 (Brown 2001; West et al. 2007b).

The last 40+ years of research has seen inclusive fitness theory applied to a vast number of traits across a range of organisms, as can be seen in any animal behaviour textbook

(e.g. (Alcock 2005; Krebs and Davies 1993). Some of the areas which have attracted considerable attention include sex allocation and sex ratios (West 2009), policing and conflict resolution (Ratnieks et al. 2006), kin discrimination (Griffin and West 2003; Rousset and Roze 2007), parasite virulence (Frank 1996b), parent-offspring conflict (Trivers 1974), sibling conflict (Mock and Parker 1997), selfish genetic elements (Burt and Trivers 2006), cannibalism (Pfennig et al. 1999), dispersal (Hamilton and May 1977), alarm calls (Sherman 1977), and genomic imprinting (Haig 2002).

[Table 1 here]

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 they produce which survive until adulthood (Dawkins 1982; Grafen 2007a; 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 (Hamilton 1964) classified social behaviours according to whether the consequences they entail for the actor and recipient are beneficial (increase direct fitness) or costly (decrease direct 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, 1970; West et al. 2007a). Examples of altruistic and mutually beneficial cooperative behaviours are given in section 5.

[Table 2 here]

A key point here is that whether a behaviour is beneficial or costly is defined on the basis of: (i) the lifetime fitness consequences of the behaviour (i.e. not just the short-term consequences); and (ii) the fitness of individuals relative to the whole population (i.e. not just relative to the individuals or social group with which the actor directly interacts). Hamilton's (Hamilton 1964, 1970) point here was that altruistic and spiteful behaviours could not be explained by the Darwinian view, formalized by Fisher (1930), that individuals strive to maximise their personal fitness. This is because altruistic and spiteful behaviours would lead to a decrease in the fitness of the individual that performs them. Instead, in order to explain such behaviours, the indirect fitness consequences had to be considered, which could be done with inclusive fitness theory.

Altruistic behaviour is favoured when it is directed towards individuals who share the same genes (positive relatedness). 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. 2007a). Spiteful behaviour is favoured when it is directed towards individuals who are less genetically 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. In misconceptions 6-9, we return to the point that inclusive fitness theory is a general encapsulation of how selection works, not just a special case for dealing with cooperation between close family members.

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. 2007a). 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 and 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 which are selectively favoured due to direct fitness consequences 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 (which they don't). 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.

The above definition of cooperation does not require that a behaviour initially evolved because of its benefit to others, or that it is completely explained by its benefit to others. A behaviour could initially be selected for because it provides a direct benefit, with the effects on others coming later. For example, many bacteria release factors that benefit both themselves and their neighbouring cells (West et al. 2006a). In order to be defined as cooperative traits, all that is required is that the rate at which these factors are produced is maintained at least partially by the social fitness consequences. This can be shown experimentally, without recourse to evolutionary history, by showing that mutants that do not produce these factors are able to exploit cells that do, and hence increase in frequency within populations (Diggle et al. 2007; Griffin et al. 2004). In this case, selection on the amount of factor produced will depend upon the relatedness of interacting cells (see section 5.1). This does not rule out the possibility that the factor originally evolved purely for the direct benefit of the cell that produces it – it just demonstrates the importance of social interactions for the current level at which it is maintained.

4. The problem of cooperation and the major evolutionary transitions

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, relatively selfish cheater arises through mutation or migration. In the absence of any mechanism to punish non-cooperators, the cheater free-rides by benefiting from the cooperative behaviour of its social partners, without paying any cost. Consequently, genes for cheating 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 (Figure 2). 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 and Raiffa 1957; Rapoport and 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. Although most attention on the problem of cooperation has focused on animals, such as insects, birds and mammals, it is increasingly being realised that cooperation is important throughout the tree of life (Sachs et al. 2004; West et al. 2007b). For example, the growth and survival of microorganisms, such as bacteria, appears to depend upon the cooperative excretion of products (public goods) that perform a variety of functions, such as scavenging nutrients, communication, defence and movement (Crespi 2001; West et al. 2006a). Furthermore, the same problem occurs at all levels of biological organization (Leigh 1991; Maynard Smith and Szathmary 1995). The very existence of multicellular organisms depends upon stable cooperation between the eukaryotic cells that make up their tissues. Likewise, the mitochondria or chloroplasts upon which these eukaryotic cells depend for energy production were once free-living bacterial prokaryotic cells but are now cooperative organelles within these larger cells. The genes that make up the genome of every cell also cooperate with each other, in what has been termed the ‘parliament of the genes’ (Leigh 1971). The general point here is that almost all of these major evolutionary transitions from replicating molecules to complex animal societies have relied upon solving the problem of cooperation (Table 3). This illustrates that not only is the problem of cooperation solvable, but it is fundamental to our understanding of evolution (Queller 2000).

[Table 3 here]

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. Contrary to popular understanding, evolutionary biology has discovered multiple mechanisms that can explain cooperation and the repression of selfish behaviour, and hence we have a good understanding of how the major transitions occurred (see misconception 1). In addition, as we hope to illustrate, there is now a good understanding of how the same evolutionary mechanisms can operate at different levels of complexity or in different types of organism.

Theoretical explanations for the evolution of cooperation (or any behaviour) are broadly classified into two categories: direct fitness benefits or indirect fitness benefits (Figure 3). 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 behaviour are termed ‘mutually beneficial’ – i.e. although they may appear altruistic, they are not (West et al. 2007a); see misconceptions 1,2 & 4). 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. 2007a); see misconceptions 1 & 2). Note that natural selection always operates upon inclusive fitness –it is just that with some mutually beneficial behaviours, the indirect component is insignificant (see misconception 8).

Evolutionary research in this area provides a unified body of work that can explain cooperation at all levels of biological complexity. There has been a huge amount of theoretical research, directed at both developing models for specific systems, and working out how these models relate to each other. Vast progress has been made in this area over the last 40 years, since Hamilton’s groundbreaking work, with overviews of the area produced by different researchers showing relatively general agreement (Bergmüller et al. 2007; Bshary and Bergmüller 2008; Clutton-Brock 2002; Foster and Wenseleers 2006; Frank 1995b, 1998, 2003; Gardner et al. 2007a; Grafen 1984, 1985, 2006a; Lehmann and Keller 2006; Queller 1992a; Rousset 2004; Sachs et al. 2004; Taylor 1996; West et al. 2007b; West et al. 2006a). An exciting feature of research in this area is that we have a single body of theory that can be applied to explain cooperation in everything, from the simplest replicating molecules to complex human societies. The overall aim is to see how the relative importance of the various mechanisms varies across different organisms (Bergmüller et al. 2007; Clutton-Brock 2002; Maynard Smith and Szathmary 1995; Sachs et al. 2004; West et al. 2007b). An additional possibility, cultural evolution, will be discussed in section 6.8, after several misunderstandings have been dispelled.

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 a female gorilla protects human children that fall into her pen, when dolphins help an exhausted swimmer, or when enslaved ants rear the brood of the slave making species that captured them. However, these ‘irrational’ or seemingly maladaptive behaviours can

be trivially explained by considering the average fitness consequences of such an evolved “rule-of-thumb”. 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 the female gorilla may be a consequence of selection for maternal care, the behaviour of dolphins may be a by product of selection for helping within dolphin groups, and the rearing behaviour of the enslaved ants is favoured because it is usually directed towards related brood. The general point here, that we shall return to in misconceptions 5 & 11, in relation to humans, is that maximisation of fitness does not lead to an expectation for perfect fitness-maximising behaviour in every real-time situation. This stresses the importance of studying behaviour 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. natural selection has not adapted gorillas to live in zoos).

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 (Hamilton 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. As we shall discuss in greater detail in misconception 7, the coefficient of relatedness (r) is not simply the average genetic relatedness (e.g. brothers=1/2, cousins=1/8), but 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). 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). Explanations for cooperation based on indirect fitness benefits require a sufficiently high genetic relatedness (r) between interacting individuals. Hamilton (Hamilton 1964) suggested two possible mechanisms through which a high relatedness could arise between social partners: kin discrimination and limited dispersal.

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. A clear example is provided by Britain’s only cooperative breeding bird, the long-tailed tit, where individuals that fail to breed independently, preferentially go and help at the nest of closer relatives (Russell and Hatchwell 2001).

Kin discrimination can occur through the use of environmental or genetic cues (Grafen 1990b). The most common mechanism for kin discrimination appears to involve environmental cues, such as prior association or shared environment, as demonstrated in a range of organisms from ants to humans (Helanterä and Sundström 2007; Lieberman et al. 2003). This is also the case with 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). Another possible mechanism for kin discrimination is via some cue that is genetically determined, such as the odour produced by scent glands in a mammal (Grafen 1990b); also termed 'genetic similarity detection', 'matching' or 'tags'). This has been demonstrated in a range of organisms, including ants and mammals (Boomsma et al. 2003; Mateo 2002). In Belding's ground squirrels, individuals discriminate kin on the basis of odours from oral and dorsal glands (Mateo 2002), and use alarm calls to preferentially warn closer relatives of the approach of predators, despite the fact that this increases their own visibility to predators (Sherman 1977).

There are a number of studies on potential mechanisms for kin discrimination in humans. Considering environmental cues, individuals are treated as closer relatives if there was a longer period of association during their childhood, which would provide a reasonable rule of thumb for family relationships (Lieberman et al. 2003, 2007). A role for odour cues has been supported by the observations that they allow mothers to recognise their newborns (Porter and Cernoch 1983; Russel et al. 1983), newborns to recognise their mothers (Cernoch and Porter 1985; Russel 1976), and adult siblings to recognize each other (Porter et al. 1986). The use of kin discriminatory cues is important for fathers, if attempting to overcome paternity uncertainty. Paternal investment towards a child is determined not only by perceived mate fidelity (Apicella and Marlowe 2004) but also by physical and psychological similarities between him and the child (Hauber and Sherman 2001; Lieberman et al. 2007). Furthermore, (Platek et al. 2005; Platek et al. 2004) have shown the brain area involved in the detection of child facial resemblance is more active in men than women. In order to reinforce the father's paternity belief, the mother and maternal relatives are more inclined to publicly ascribe facial resemblance of a newborn to the father (Alvergne et al. 2007; Daly and Wilson 1982a; McLain et al. 2000; Regalski and Gaulin 1993). Patterns of investment in children reflects relatedness certainty, with maternal grandmothers investing most in grandchildren, followed by maternal grandfathers, paternal grandmothers and then paternal grandfathers (Eisenberg 1988; Kahana and Kahana 1970; Pollet et al. 2007; Rossi and Rossi 1990). This pattern is also reflected by aunts and uncles, who invest more in the matriline (McBurney et al. 2002). Infanticide data follow similar patterns (Daly and Wilson 1982b).

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.

A confusing aspect of both the biological and social sciences literature is that this idea seems to be particularly prone to being reinvented, as a “new” explanation for altruism or cooperation. In all cases the key role of relatedness and indirect fitness benefits can be shown by more rigorous analyses (Grafen 2007b, c; Lehmann and Keller 2006; Lehmann et al. 2007a; Lehmann et al. 2007b; Lehmann et al. 2007c; Taylor et al. 2007a). Examples of indirect benefits via limited dispersal being reinvented in the biological literature include work on the prisoner’s dilemma and other games in a spatial setting (e.g. (Boyd and Richerson 2002; Nowak and May 1992), group or multi-level selection (e.g. (Traulsen and Nowak 2006; Wilson 1975a), population structure (e.g. (Killingback et al. 2006) and network reciprocity via games on graphs (e.g. (Lieberman et al. 2005). In misconceptions 6, 7 & 16, we shall discuss an example from the economics literature - strong reciprocity. The mistakes in these areas seem to stem from the incorrect assumption that kin selection or indirect fitness benefits require kin discrimination (misconception 6), despite the fact that Hamilton pointed out the potential role of limited dispersal in his earliest papers on inclusive fitness theory (Hamilton 1964; Hamilton 1971; Hamilton 1972; Hamilton 1975); misconceptions 6 & 7). The level of confusion that this reinvention problem can create is illustrated by a recent review which suggested five mechanisms for the evolution of cooperation (Nowak 2006), but where it turns out that three of those are just the same thing – ‘network reciprocity’ and group selection are just different ways of analyzing special cases of kin selection (Grafen 2007b; Lehmann et al. 2007a; Lehmann et al. 2007c; Taylor et al. 2007a; West et al. 2007b).

The potential role of limited dispersal has recently been much discussed in work on cooperation in bacteria and other microorganisms, where clonal reproduction means that neighbouring cells can be highly related (Diggle et al. 2007; Gilbert et al. 2007; Griffin et al. 2004). Experimental support for the role of limited dispersal was provided by an experimental evolution study on how bacteria release siderophore molecules to scavenge for iron. These siderophores represent a cooperative public good: they are costly to the individual to produce, but iron bound to siderophores can be taken up by any cell, providing a benefit to other individuals in the locality. When populations of the bacterium *Pseudomonas aeruginosa* containing a mixture of a wild-type strain that produces siderophores and a cheater mutant that does not were maintained in conditions that led to relatively high or low relatedness, the cooperative wild-type strain outcompeted the cheater mutant strain under conditions of relatively high relatedness, but not under conditions of relatively low relatedness (Griffin et al. 2004). More generally, bacteria release a huge range of extracellular products that appear to be public goods (West et al. 2007c), and a high relatedness also favours signaling between bacterial cells to coordinate the production of these products (Brown and Johnstone 2001; Diggle et al. 2007).

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. 2007a). We divide the direct fitness explanations for cooperation into two categories (Figure 2). First, the direct benefits of cooperating may flow automatically (passively) as a by-product of helping another individual (Darwin 1871), chapter III). Coordinated foraging in groups appears to be an example of this, where everyone gains an immediate benefit from increased acquisition of food, in animals such as 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 and 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). Another theoretical possibility is that cooperation is a costly and honest signal of quality (Gintis et al. 2001).

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 (Trivers 1971) emphasised that cooperation could be favoured in reciprocal interactions with individuals preferentially aiding those that have helped them in the past. 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 and Maschler 1995; Binmore 1994, 1998, 2005b, 2007; Fudenberg and Maskin 1986; Kandori 1992; Luce and Raiffa 1957; Mailah and Samuelson 2006). Within the economics literature, the possibility for reciprocity or punishment to solve the problem of cooperation in repeated interactions is usually referred to as the 'Folk Theorem'. It was termed this, because everyone seemed to already appreciate this possibility, when it was formalized in the 1950's by (Aumann 1959). This form of 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 and Sigmund 1998). Both direct and indirect reciprocity have been argued to be important in the evolution and maintenance of cooperation in humans (Alexander 1987; Binmore 1994, 1998, 2005b; Gächter and Herrmann 2009; Henrich and Henrich 2007; Milinski and Wedekind 1998; Milinski et al. 2002; Nowak and Sigmund 2005; Palameta and Brown 1999; Seabright 2004; Trivers 1971; Wedekind and Milinski 2000).

In contrast to the huge body of research investigating the theoretical plausibility of reciprocity, statements in standard animal behaviour textbooks, and its empirical importance in humans, reciprocity is thought to be generally unimportant in other organisms (Bergmüller et al. 2007; Clutton-Brock 2002, In press; Hammerstein 2003;

Russell and Wright 2008; Stevens and Hauser 2004). Even classical examples such as blood sharing in vampire bats (Wilkinson 1984), can be explained more simply without the need for reciprocity (Clutton-Brock In press). Furthermore, it is now widely recognised that reciprocity is only one of the mechanisms for enforcing cooperation – other possibilities, for which there are greater empirical support, have been termed punishment, policing, sanctions, partner switching and partner choice (Bergmüller et al. 2007; Frank 2003; Sachs et al. 2004; West et al. 2007b). One example of punishment is found in meerkats, where the dominant female suppresses reproduction in her subordinates (Young et al. 2006). If a subordinate female becomes pregnant when the dominant is also pregnant, then the dominant is likely to subject the subordinate to aggressive attack and temporarily evict her from the group, which usually leads to abortion of the subordinate's litter. Another example is provided by Superb Fairy Wrens, where subordinates are punished if they are removed (so cannot help) during the breeding season, but not outside it (Mulder and Langmore 1993). Punishment may also be important in humans, as we shall discuss below.

Enforcement mechanisms have been suggested to be important in explaining cooperation between species. An elegant example is provided on coral reefs where the cleaner fish *Labroides dimidiatus* removes and eats ectoparasites from its 'clients', which refrain from consuming this potential prey while it performs the service. Although parasite removal and food acquisition are clearly beneficial to the client and cleaner, respectively, there is a conflict, because the cleaners would prefer to eat the tissue or mucus of their hosts, which is costly to the host (Bshary and Grutter 2002). The clients use three mechanisms to suppress this conflict and enforce cooperative feeding on ectoparasites only: avoiding cleaners that they have observed cheating (reputation effects or partner choice), leaving for another cleaner (partner switching), and aggressively chasing the cleaner (punishment) (Bshary 2002; Bshary and Schäffer 2002; Bshary and Grutter 2002). Observational and experimental data suggest that cleaner fish are more cooperative and less likely to feed on mucus after punishment (Bshary and Grutter 2005).

Conditional enforcement may be extremely important in explaining cooperation between species, where kin selection cannot be the driving force. Other examples include how *Yucca* plants selectively abort flowers which have been over exploited by their pollinator moth (which lays its eggs inside the flowers) (Pellmyr and Huth 1994), and how legume plants (beans and peas) cut off the O₂ supply to (sanction) the rhizobia bacteria that live in nodules in their roots, if they do not provide them with nitrogen (needed for plant growth) at a sufficiently high rate (Kiers et al. 2003).

5.3 Why enforce?

Whilst 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 non-punishers represents a third-order public good (Henrich and Boyd 2001; Sober and Wilson 1998). Clearly, punishment will only be favoured if it provides a net direct or an indirect fitness benefit (Gardner and West 2004; Lehmann et al. 2007b). 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, *Yucca* and legume-rhizobia interaction. 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 and Clutton-Brock 2006). Alternatively, individuals could preferentially initiate interactions with relatively more cooperative individuals (Hauk 2001), which appears to be happening in cleaner fish (Bshary and Schaffer 2002). 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 and 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.

Enforcement could also be favoured if it provides an indirect fitness benefit (Frank 1995b; Gardner and West 2004; Lehmann et al. 2007b; Ratnieks 1988). The simplest way this could occur is by reducing the fitness of individuals who are competing with relatives and hence freeing up resources for relatives. This occurs in some ants, bees and wasps where a fraction of the workers lay their own eggs (Ratnieks and Visscher 1989). Other workers frequently do not tolerate such behaviour and selectively cannibalise or ‘police’ eggs laid by workers. This behaviour is selected for because the policing workers can be more related to the sons of the queen than to the sons of the other workers and because cheating workers raising their own sons can reduce the colony’s overall productivity (Ratnieks 1988; Wenseleers et al. 2004). Across species, it has been shown that there are higher levels of worker cooperation in species where policing is more common and effective. Specifically, the proportion of workers who lay eggs is negatively correlated with the probability of worker laid eggs being killed (Wenseleers and Ratnieks 2006). One way of conceptualizing this is that policing reduces the fitness gains of cheating, which is the same as reducing the cost (c) of cooperating in Hamilton’s rule

5.4 Interactions between direct and indirect benefits

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 with animals such as meerkats and (ancestral) humans. 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 and Keller 2006).

5.5 Origin versus maintenance of cooperation

It is important to distinguish between the evolutionary forces that favoured the origin and then subsequent elaboration and maintenance of a trait. The selective force initially responsible for the emergence of cooperation may differ from that maintaining the observed level of cooperation. In particular, even when 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 or selfish behaviour) is also an ESS. This is for instance the case with direct reciprocity (Axelrod and Hamilton 1981), indirect reciprocity (Panchanathan and Boyd 2004), punishment (Gardner and West 2004; Henrich and 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. As we shall see in misconception 18, a distinction between initial origin and later elaboration is also key to our understanding of cultural selection.

6. Common Misconceptions

In this section we briefly run through eighteen common misconceptions about social evolution theory (sociobiology), which are summarised in table 4. 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 about if and why humans are special are discussed in section 7.1. The interested reader is also directed towards Dawkins' (1979) "Twelve misunderstandings of kin selection", many of which are still pertinent today.

[Table 4 here]

6.1. Kin Selection, Reciprocity and altruism

Misconception 1: Kin selection and reciprocity are the major competing explanations for altruism in biological theory (e.g. (Boyd and Richerson 2005; Boyd et al. 2003; de Waal 2008; Fehr and Gächter 2002; Fehr and Rockenbach 2003; Fehr and Fischbacher 2003; Fehr and Rockenbach 2004; Gintis et al. 2005b; Henrich and Boyd 2001; Richerson and Boyd 1999).

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 non-relatives) if it leads to an overall benefit, in which case it is mutually beneficial (+/+). Put another way, altruism is

defined according to the lifetime consequences of a behaviour (see section 3), and reciprocity is only altruistic in the short term. Much of the confusion here is due to Trivers' (1971, 1985) term 'reciprocal altruism', the introduction of which was accompanied by multiple redefinitions of altruism (West et al. 2007a), 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' (Alexander 1974; Binmore 1994, 1998; West et al. 2007a), or 'reciprocal cooperation' (Axelrod and Hamilton 1981). We discuss confusion of the term 'altruism' in more detail in misconception 2.

Second, even 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 (Figure 3), and whilst it may be particularly important in humans, it is relatively unimportant elsewhere. In some cases, this misconception appears to arise from only considering the evolutionary literature up until approximately the late 1970's, and hence missing the huge advances that have been made since then (the "disco problem"; see also misconception 4). As well as in the papers cited above from the primary literature, misconception 1 or a close approximation occurs in a scarily large number of undergraduate textbooks.

Misconception 2: The various redefinitions of altruism (Baschetti 2007; Becker 1974; Bergstrom 1995, 2002; Bowles and Gintis 2004, 2008; Boyd et al. 2003; Fehr and Fischbacher 2003; Gintis 2000; Sober and Wilson 1998; Trivers 1971; Wilson 1975a).

In section 3 we emphasized how terms such as altruism have very specific meanings, that 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. 2007a), p. 419-423), which has been redefined in many ways, including: (a) as discussed in misconception 1 with reciprocal altruism (Becker 1974; Fehr and Fischbacher 2003); (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 and Gintis 2004; Boyd et al. 2003; Gintis 2000; Sober and Wilson 1998; Wilson 1975a); (c) a decrease in fitness over the short term (Becker 1974; Fehr and Fischbacher 2003); (d) playing cooperate in a prisoners' dilemma game (Bergstrom 2002); (e) a failure to harm others (Field 2001); (f) the mechanism by which one individual is motivated to help others (Axelrod 1984). We have already discussed another redefinition of altruism, to include reciprocity, in misconception 1.

The problem with all of these redefinitions is that they include situations 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 non-cooperators. 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 and Gintis 2004; Bowles et al. 2003; Boyd et al. 2003; Gintis 2000; Gintis et al. 2003; Henrich and 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 and Johnson 2005; Lehmann et al. 2007b; West et al. 2007a). 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. 2007a); (b) a ‘selfish agent’ (Bowles and Gintis 2004) can have a lower direct fitness than an altruist. The importance of direct fitness benefits from cooperation are further emphasized by the model of Boyd et al. (2003), where groups compete for territories in pairs. In their model, the territory is won by the group with the most cooperators, and so it is clear that a single individual could potentially gain a huge direct fitness advantage by cooperating, and hence making its group much more successful.

More generally, the mechanism favouring cooperation in the models discussed in the previous paragraph appear to be the same as that already proposed in models of group augmentation, discussed in section 5.2. The idea of group augmentation was developed to show how cooperation could provide direct fitness benefits, and hence could be favored between non-relatives. An important point here is that we are not saying that these models of cooperation in humans never consider situations that are altruistic. Instead, cooperation can be mutually beneficial or altruistic, depending upon the values of parameters such as the dispersal rate, group size, cost and effect of punishment etc, all of which influence the local competition for resources (Lehmann et al. 2007b). The potential for the evolution of altruistic cooperation, as defined by Hamilton (1964), arises because these models assume limited dispersal, which leads to a significant relatedness between the individuals interacting within a group, and hence indirect fitness benefits (see misconceptions 6,7 & 16). Cooperation can therefore provide both direct fitness benefits and indirect benefits via cooperating with relatives (Lehmann et al. 2007b).

The above points illustrate that the 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. 2007a). 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). More formally, these redefinitions are misleading because the spread of a gene is determined by its overall (lifetime) fitness consequences, relative to others in the breeding population, and not its consequences over some arbitrarily defined ‘short-term’, or relative to some subset of the population, such as with whom they happen to interact (Grafen 1984, 2002, 2006a; West et al. 2007a).

Some confusion over terminology may also have arisen from Dawkins’ (Dawkins 1976) title “The Selfish Gene”, 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 & 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, 2007b). 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, but from the perspective of all copies of the 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). 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, this misconceived view of "selfish genes" appears to have supported 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.

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, and therefore follow from Hamilton's (1964) original definitions. Furthermore, as discussed in section 2, natural selection leads to the appearance of design or purpose at the individual level, with individuals expected to behave as maximizing agents, and so justifies the use of such intentional language at the individual level. It is precisely for this reason that Hamilton's definitions have proved so useful in the field of animal behaviour (Krebs and Davies 1993). Another common source of confusion 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 which therefore do not necessarily fit the evolutionary definition. This semantic problem leads to confusion over the underlying selective forces, and clouds the problem being addressed! 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).

6.2 The Prisoners' Dilemma and Tit-For-Tat

Misconception 3: The evolution of cooperation is encapsulated by the Prisoners' Dilemma (PD) and how it can be solved by the behavioural strategy 'tit-for-tat'.

The PD and tit-for-tat have led to much confusion in both the biological and social sciences literatures. Following Axelrod's (Axelrod 1984; Axelrod and Hamilton 1981) hugely influential work in this area, it has commonly been assumed that the strategy 'tit-

for-tat' (cooperate, but then punish non-cooperation by also not cooperating) is the evolutionary stable strategy (ESS) in the iterated PD game, and that it is disadvantageous to be the first player to defect (e.g. (Axelrod 1984; Maynard Smith 1982; Poundstone 1988); reviewed by (Binmore 1994), pp. 173-175, 194-203; (Binmore 1998), pp. 186, 313-319). This is especially true in the biological literature, where researchers appear to be generally unaware of the huge literature on reciprocity in the fields of game theory and economics (Binmore 1998). However, the folk theorem has long shown that tit-for-tat is only one of an enormous number of strategies (Nash equilibria) that can be favoured, and that there are many more ways of supporting cooperation in repeated interactions than by naive pairwise reciprocation (Binmore 1994, 1998; Boyd and Lorberaum 1987).

A number of studies have supported the folk theorem by showing that a mixed Nash equilibrium is selected for, in which multiple strategies are maintained, with tit-for-tat not even at a particularly high frequency (Binmore 1994, 1998; Boyd and Lorberaum 1987). Axelrod (Axelrod 1984) carried out computer simulations that competed different strategies against each other, and whilst tit-for-tat was the most common strategy after competition, the 'best' strategy was actually a mixed strategy that performs tit-for-tat only approximately 1/6th of the time. Since then, it has been shown that the results are extremely dependent upon the different strategies that are competed, their initial starting frequencies, and how long the simulations are run for (Binmore 1994, 1998; Young and Foster 1991). For example, in some cases, a greater success is enjoyed by the more nasty reciprocator 'tat-for-tit' (see also 'Pavlov' in (Nowak and Sigmund 1993), which starts by not cooperating, and only switches to cooperating in response to the cooperation of others. Furthermore, many of the famous anecdotes of tit-for-tat in humans, such as ceasefire arrangements between soldiers in the trenches during the First World War, are at least as consistent with 'nastier' strategies such as tat-for-tit. (Binmore 1994, 1998) Although, the point here, is not that we should argue over what is the best strategy, just that lots of strategies can be maintained at equilibrium, results depend hugely upon starting conditions, and that tit-for-tat won't even necessarily be the most common.

Second, it is sometimes assumed that the evolution of cooperation is usually studied with the PD (Bergstrom 2002; Nowak et al. 2004). To put this more bluntly, "A whole generation of scholars swallowed the line that the Prisoners' Dilemma embodies the essence of the problem of human cooperation" (Binmore 2007), p. 18). In the early days of modern biological social evolution research (1960's-80's), it was relatively hard to convince people that altruism and cooperation were problematic, and needed a solution (Hamilton 1996). The beauty of the PD was that it allowed a clear abstraction of the problem of cooperation, with a 2x2 payoff matrix, and a quick back-story, to help visualisation. The main use of the PD was therefore in convincing people that cooperation was a problem, rather than actually giving general insights into how this problem is solved. This is because it represents the situation in which selection against cooperation is as strong as possible (Binmore 2007). In the economics literature, this problem has been addressed with elegant extended analyses of the iterated PD (Benoit and Krishna 1985; Binmore 1994, 1998; Fudenberg and Maskin 1986, 1990; Fudenberg et al. 1994; Kandori 1992), but also by looking at a range of other games, for situations that may be particularly appropriate in humans, such as bargaining (Binmore 1994,

1998).

In the biological literature, the focus on the PD, and reciprocity, has actually hindered progress (Bergmüller et al. 2007; Boyd and Richerson 1988; Clutton-Brock 2002; Hammerstein 2003; Leimar and Hammerstein 2006; Stevens and Hauser 2004; Stevens et al. 2005; West et al. 2007b). The PD makes a large number of extremely specific and often unrealistic assumptions, which are rarely stated or justified, such as discrete strategies (cooperation is all or nothing), confounding cooperation with punishment (Axelrod (1984, p. 120-121) argued that this is why complex strategies don't do well), no partner choice (interactions at random), a limited payoff structure, simultaneous moves, etc. Whilst abstraction can be useful for ease of analysis or tractability, the various assumptions of games such as the PD often make analysis more complicated, leading to highly technical mathematics or heavy reliance upon numerical simulations. From an empirical perspective, a focus on the PD has led to the common assumption that reciprocity is important in animals (see any animal behaviour textbook), whereas in reality, there is a lack of evidence that it is important in any non-human animal (Clutton-Brock In press; Hammerstein 2003).

Consequently, whilst they are of interest from a mathematical or aesthetic perspective (Nowak et al. 1994), these approaches lack transparency, and so it can be hard to determine the underlying processes, or see how the results of different studies relate to each other (Lehmann and Keller 2006). This has led to much confusion because it leads to: (a) repeated reinvention of the 'wheel' – apparently novel solutions to the problem of cooperation that simply turn out to be new names for old ideas (Lehmann and Keller 2006; West et al. 2007b); (b) a lack of attention to all the other ways in which direct fitness benefits can favour cooperation (see figure 3 and misconception 1), but which cannot be analysed with the PD; (c) it results in empirical workers making things seem more complicated, in an attempt to contort real systems into the assumption of the PD (Brown 2001; West et al. 2007b). Theory should facilitate predictions and empirical testing of these predictions - if empiricists have to bend-over-backwards to get the real world to fit the theory, then something has gone badly wrong. The kin selection theory approach, and especially its 'direct' or neighbour-modulated fitness implementation (Frank 1998; Taylor and Frank 1996) provides a solution to these problems, because it allows the biology to lead the mathematics, and hence facilitates the empirical application and testing of theory

6.3 Mutually beneficial cooperation

Misconception 4: Mutually beneficial cooperation is less interesting.

Misconception 2 illustrated the point that altruism is often redefined so that it will include a particular case of cooperation that is being examined. Furthermore, discussion and correspondence suggests that researchers are usually disappointed to discover that a case under discussion fits into the mutually beneficial category (+/+) and is not altruistic (-/+). Indeed, altruism may be redefined so frequently simply 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 major 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 and West 2002; Jennions and Macdonald 1994).

A contributing factor to misconception 4 may be the often quoted statement from EO Wilson's (Wilson 1975b), p.31) sociobiology book 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 it is actually cooperation that is the central problem of sociobiology: how can behaviours which benefit others evolve by natural selection? Indeed, as discussed in section 4, understanding cooperation is actually one of the central problems of the whole field of evolutionary biology, because of its role in the major evolutionary transitions. In some cases these transitions have involved mutually beneficial cooperation, and in others, altruistic cooperation (Queller 2000). A general problem here is that secondary sources aimed at a less specialist audience can give an incorrect impression of the primary literature. For example, Wilson's (1975) book gave a weak representation of social evolutionary theory, even at the time, and is accepted to be misleading on several accounts including incorrect descriptions of inclusive fitness, kin selection, group selection, altruism and spite (Dawkins 1979; Foster et al. 2006; Gardner et al. 2007b; Grafen 1982; Helanterä and Bargum 2007; Lehmann et al. 2006; West et al. 2008).

6.4. Proximate and ultimate explanations

Misconception 5: 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 that 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 and Gintis 2004; Fehr and Gächter 2002; Fehr and Rockenbach 2003; Fehr and Fischbacher 2003, 2004; Fehr and Rockenbach 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 and 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 non-repeated interactions when reputation gains are absent” (Fehr and Fischbacher 2003), or “cooperation is maintained because many humans have a predisposition to punish those who violate group-beneficial norms” (Bowles and Gintis 2004), or that a solution to the problem of why punish is that it leads to neurological “satisfaction” (Fehr and Rockenbach 2004).

This approach mixes up two different questions (how and why, or process and product). 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 answers to this are because it provides either a direct and/or an indirect fitness benefit (Gardner and West 2004). The related theory, which we discuss in misconceptions 2, 6, 7 & 16, suggest that the answer is a mixture of direct and indirect fitness benefits whose relative importance varies depending upon the exact assumptions (Lehmann et al. 2007b; West et al. 2007a).

Using a proximate mechanism to answer an ultimate question is invalid. Specifically, 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. Our point here is not that there are no theoretical models of strong reciprocity which address the ultimate problem of cooperation, because there are (see misconceptions 2, 6, 7 & 16), but that proximate and evolutionary issues are mixed up and confused, as illustrated by the above (and below) quotes. 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 and 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.” Again, this does not solve the ultimate problem, because it does not answer why evolution should have produced a nervous system that mechanistically encourages (rewards) such punishment.

Confusion over proximate and ultimate factors occurs in numerous places. Other examples where proximate mechanism have been suggested as explanation for the ultimate problem of “pro-social” preferences include:

1. ‘Warm-glow’ theories of altruism, which predict that people behave altruistically as they have an internal mechanism that makes them feel good when they do so (Andreoni 1990). As discussed above, this does not answer the ultimate problem, because it does not explain why such a proximate mechanism would have been favoured; the same issue applies to the next two examples.
2. Inequity aversion models, which assume altruistic behaviour can be explained because individuals do not like inequality so will behave altruistically towards those who are worse off and punish those who are better off than they are (Fehr and Schmidt 1999)
3. Self-identity models, where people care not only about reputation but also about self image (Akerlof and Kranton 2000; Benabou and Tirole 2004; Bonder and Prelec 2003).
4. ‘Social institution’ models, on how selection for cooperation is increased by “the commonly observed human practices of resource sharing among group members” (Bowles et al. 2003). However, as an institution is a form of cooperation itself, it just moves the problem to why would the social institution of cooperative resource sharing evolve? A useful comparison here can be made with models on how repression of competition within groups can favour cooperation, through mechanisms such as randomization of reproduction, or policing (Frank 1995b, 1996a, 2003; Leigh 1971; Ratnieks 1988), but which assume that repression of competition is a potentially costly trait, under selection, rather than cost free trait subject to random drift, as assumed by Bowles et al.
5. 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.
6. Similar mixing up of proximate and ultimate factors occurs at the interface of primate and human literature (de Waal 2008).
7. Confusion often arises with the use of the term ‘utility’ (Falk and Fischbacher 2001 (MS)) – saying that individuals behave so as to optimise utility is a different issue from determining why utility is of the assumed form.
8. Similar problems are discussed by (Scott-Phillips 2007) in the context of the evolution of language literature, and by (Smuts 1999) in the context of how Sober & Wilson (1998) mix up ultimate and proximate explanations of altruism.

A general issue here is that care must be taken to not over interpret proximate patterns from an ultimate perspective. Ultimate and proximate factors can operate at different time scales. Natural selection works over evolutionary time, with time steps of generations compared to economic settings where the time step can be a few seconds and the

expectation is that utility will be maximised on that time frame. Natural selection will not result in behaviours that are perfect in every conceivable situation - instead it favours proximate mechanisms that maximise fitness in the situations that are encountered (the 'selective regime'). Consequently, if we try to assign ultimate explanations to proximate patterns, outside of the context of the selective regime in which they were favoured, we run the risk of adaptationist story telling. We stress that the possible limits to adaptation is not an idea invented to explain patterns in humans - it has long been accepted in the field of animal behaviour (Herre 1987; Krebs and McCleery 1984; Wehner 1987). To give a specific example in the case of humans, it is often argued that indirect benefits of cooperation can be ruled out from the start in humans, because cooperation still occurs when we "know the players are not relatives". However, the key point here is not the present situation, but whether relatedness was significant when the underlying proximate mechanisms evolved (Binmore 1998; Seabright 2004). Note that, we are not necessarily saying that relatedness was important, just that proximate mechanisms (or observations that cooperation occurs between non-relatives) do not necessarily rule it out. We return to the general issue of over interpreting proximate patterns in misconception 15.

6.5 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.

Misconception 6: 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; Hamilton 1971; Hamilton 1972; Hamilton 1975). There is a huge theoretical literature on this (reviewed by (West et al. 2002a), as well as experimental evolution tests (Diggle et al. 2007; Griffin et al. 2004; Kümmerli et al. 2009). Since then, 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 / relatedness is not important. Examples from the biological literature include work on the prisoner's dilemma in a spatial setting (e.g. (Nowak and May 1992), group or multi-level selection (e.g. (Traulsen and Nowak 2006; Wilson 1975a), population structure (e.g. (Killingback et al. 2006) and network reciprocity (games on graphs; e.g. (Lieberman et al. 2005). Examples from the economics literature include some of the models of strong reciprocity, which we shall discuss in further detail below (e.g. (Bowles and Gintis 2004; Gintis 2000). In all of these cases the key role of relatedness and indirect fitness benefits can be shown by formal analysis (Grafen 2007b, c; Lehmann and Keller 2006; Lehmann et al. 2007a; Lehmann et al. 2007b; Lehmann et al. 2007c; Taylor et al. 2007a; Wild 2008).

Misconception 7: Relatedness is only high between close family members (Bowles and Gintis 2004; Boyd and Richerson 2005; Gintis 2000).

It is a 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, it is less well appreciated that this assumes a large panmictic population, and that if there is population structuring with limited migration (viscous populations or limited dispersal), then relatedness between group members can be relatively high. The reason for this is that limited dispersal will tend to increase the genetic similarity between interacting individuals, and relatedness is a measure of genetic similarity (Hamilton 1964, 1970; Hamilton 1971; Hamilton 1972; Hamilton 1975).

A famous encapsulation of this point is Wright's (1931) F_{ST} , which measures the degree of genetic homogeneity within a group relative to the whole population. For haploids, $F_{ST} = (1 - m)^2 / (N - (N - 1)(1 - m)^2)$, where m is the individual rate of dispersal (proportion of individuals that disperse from their natal patch before breeding) and N is the group size, and this is exactly the kin selection coefficient of relatedness for an individual to its group in this case. For example, if $m = 0.01$ and $N = 100$, then the average relatedness of group mates is approximately $1/3$; hence, the relatedness between first cousins will be $>1/3$, and not the commonly assumed $1/8$. For diploids, $F_{ST} = (1 - m)^2 / (2N - (2N - 1)(1 - m)^2)$, which for small m and large N is well approximated by $F_{ST} = 1 / (1 + 4Nm)$, representing one of the most famous equations of population genetics theory, and relatedness is given by $2F_{ST} / (1 + F_{IT})$, where F_{IT} is a measure of genetic homogeneity within an individual (inbredness; (Hamilton 1970; Rousset 2004) p. 141). These equations also show that it is incorrect to assume that kin selection cannot be important in large populations. Clear quantitative support for these predictions has been provided by experimental evolution studies on how population structure influences selection for cooperation in bacteria (Brockhurst et al. 2007; Griffin et al. 2004; Kümmerli et al. 2009).

A clear example of the confusion that can arise here is the various strong reciprocity theoretical models where it is argued that kin selection is not important (e.g. (Bowles and 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. 2007b; West et al. 2007a). We shall return to this example in misconception 16, when discussing strong reciprocity in further detail. Another example, is provided by the claim that group selection is an alternative mechanism to cooperation between 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 (Bowles 2006). We shall return to this example in misconceptions 10-14, when discussing group selection in further detail.

The above discussion for misconceptions 7 and 8 rest upon the understanding that relatedness is a statistical measure of genetic similarity (a regression coefficient). 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 1980's or later. 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; Hamilton 1975; Michod and Hamilton 1980), providing the regression definition of relatedness that is at the centre of modern social evolution theory (Frank 1998; Grafen 1985, 2006a; Taylor and 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 Dawkins' (1976, 1982) popularisations of inclusive fitness theory. The key point here is that 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 co-ancestry 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 and Goodnight 1989). The extent to which the statistical measure of relatedness is used by empirical biologists is clear from the fact that the Queller & Goodnight (1989) methods paper has been cited > 1100 times. Second, it is true that animal behaviour textbooks such as Krebs & Davies (1993) and Alcock (2005) define relatedness through co-ancestry and not statistically. However, the reason for this is that it is a useful approximation for teaching certain age groups of undergraduates. Compare with the teaching of Physics, where it is common for a new year of study to involve learning why what was taught in the previous year was incorrect, and just a useful approximation. A general point here is that the primary literature needs to build upon and relate to the primary literature, not to textbooks – and especially not those textbooks aimed at early stage undergraduates!

Misconception 8: Inclusive fitness only applies to interactions between relatives, and greenbeard genes can explain cooperation in humans (Bergstrom 1995, 1996, 2002; Bowles and Gintis 2004, 2008; Boyd and Richerson 2005; Frank 1987; Gintis 2000).

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 non-social characters. Second, as discussed in misconception 7, 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 (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; Hamilton 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 and West In press; Lehmann and 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 and Roze 2007).

Greenbeard genes are likely to be extremely rare in the real world (Gardner and West In press). The idea of greenbeards was initially developed as a thought experiment, to illustrate that what matters for inclusive (indirect) fitness is genetic similarity at the locus (or loci) being considered, rather than genealogical relationship *per se* (Hamilton 1964, 1970; Hamilton 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). Furthermore, there is strong selection for modifiers arising at other loci to suppress the cooperative behaviour (Gardner and West In press). 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 and West In press). 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 pleiotropic effects.

Models for the evolution of cooperation that rely upon greenbeards are unlikely to be important in humans (Gardner and West In press; 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, (Frank 1987) assumed that individuals who cooperate differ from individuals who cheat in “some observable characteristic” (p. 596) other than the cooperation phenotype itself, which represents the original green beard scenario. (Owren and 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 and 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. 2007b). 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 and Gintis 2004; Gintis 2000), they are actually examining the evolution of spiteful greenbeards (Lehmann et al. 2007b)! Consequently, 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.

Misconception 9: Greenbeards are a type of costly signaling (Henrich 2004; Owren and 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 and 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, 2008).

(Howard 1971)'s metagames with 'transparent disposition' and (Gauthier 1986)'s 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.

The four misconceptions given above emphasise how the generality of inclusive fitness theory and Hamilton's rule are often underappreciated. Hamilton's rule provides an encapsulation of inclusive fitness theory that can be applied to all forms of social behaviour, and not just altruism: r , b and c can each be positive or negative. Furthermore, it clarifies that the coefficient of relatedness is a measure of the correlation between two individuals in their genetic predisposition for a trait of interest, rather than a measure of their genome-wide similarity or genealogical closeness per se. However, in many scenarios the coefficient of relatedness will (on average) be the same over most of the genome, and will tend to coincide with the genealogical relationship of the two individuals – justifying J.B.S. Haldane's famous quip that he would give his life “for two brothers, or eight cousins”.

A possible source of confusion here is the term ‘kin selection’. Maynard Smith (1964) coined the term ‘kin selection’ to describe how indirect fitness benefits arise from helping relatives reproduce. Since then, the phrase kin selection has been used in two different ways (West et al. 2007a). The narrower use of kin selection works upon interactions between individuals who are genetically related due to common ancestry – i.e. indirect benefits due to limited dispersal or kin discrimination. The broader use of kin selection works upon interactions between individuals who are genetically correlated at the loci of interest, regardless of whether this is due to coancestry or some other mechanism – i.e. this also includes greenbeard effects. The difference between these usages is therefore whether kinship and relatedness are defined on the basis of average genetic similarity over most of the genome (narrow definition), or at the particular locus of the behaviour being examined (broad definition). However, the possibility for confusion is mainly theoretical, because kinship is by far the most common reason for indirect fitness benefits, with greenbeards being incredibly rare. In addition, the use of the phrase “inclusive fitness” should avoid confusion, because it has been defined broadly since its inception (Hamilton 1964, 1970; Hamilton 1971, 1975).

6.6. 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 and Grafen 2009; West et al. 2007a, 2008).

Misconception 10. 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 (Okasha 2004, 2006; West et al. 2007a, 2008). 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 not go extinct. Hence, by a process of differential survival of groups, behaviour evolved that was for the good of the group – i.e. selection would favour traits that maximize group success. During the 1960's and 1970's 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). Specifically, selection will only produce behaviours that maximize group success (group adaptations) if all the individuals within a group are genetically identical clones, or there is complete repression of competition within groups, such that the reproductive success of members of the group cannot differ (Gardner and Grafen 2009). This corresponds to the extreme cases where maximizing group success is the same as maximizing inclusive fitness.

Empirical work showed that individuals were reproducing at the rate that maximized their inclusive fitness, and were not adapted to maximize group fitness (Krebs and Davies 1987; Lack 1966; West et al. 2008). It is this 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; Soltis et al. 1995; 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 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).

In the 1970s and 1980s a 'new' form of group selection was championed by DS Wilson and others, which examined the consequences of interactions in small structured populations (Colwell 1981; Hamilton 1975; Wilson 1975a, 1977). These models 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 outweighed the benefits at the individual level. 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, whilst the earliest group selection models (e.g. (Colwell 1981; Traulsen and Nowak 2006; Wilson 1975a, 1977) were reinventing how indirect fitness benefits (kin selection) can work via limited dispersal, later models (e.g. (Wilson and Dugatkin 1997; Wilson and 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 and Keller 2006; Lehmann et al. 2007c; Maynard Smith 1976).

The key point here is that new group selection (multi-level selection) is just a different way of looking at the dynamics by which inclusive fitness is maximized. Put another way, they are mathematically identical, and are simply different ways of looking at the

same thing (Frank 1986, 1995a; Gardner 2008; Gardner and Grafen 2009; Gardner et al. 2007a; Grafen 1984; Hamilton 1975; Lehmann et al. 2007c; 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; (Bourke and Franks 1995; Frank 1986; Gardner et al. 2007a; Lehmann et al. 2007c; Wenseleers et al. 2004). This is not surprising given how they can both be formalized with the Price equation (Gardner 2008; Gardner et al. 2007a). As we shall discuss in further detail in misconception 14, 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).

[Table 5 here]

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 et al. 2003; Boyd and 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 in pairs, 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 6,7 & 16). 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.

The term group selection is also used when discussing “cultural group selection” or “gene-culture coevolutionary multi-level selection”. Cultural group selection is used to label situations when differential group success results from the expression of different cultural traits (Boyd and Richerson 2005; Fehr et al. 2002; Gintis 2003; Gintis et al. 2003; Henrich 2004; Henrich and Boyd 2001; Lehmann et al. 2008; McElreath and Henrich 2006; Richerson and Boyd 2005). This is analogous to the third use described above, and not a new one, because it is used to mean that competition occurs between groups. However, it should be noted that it refers to a different situation, because it is referring to selection on a cultural trait, rather than a genetically determined trait. In addition, just because competition is occurring between groups, this does not mean that group level adaptations are expected to evolve (Gardner and Grafen 2009). Consequently,

whilst it is often argued that the group is the fundamental unit of cultural evolution, or that cultural evolution is a group-level process (Boyd and Richerson 1985), there is no theoretical basis for this. The theory used to study evolution via genetic or cultural selection is the same – what matters is the inclusive fitness of the cultural trait (Cavalli-Sforza and Feldman 1981; Lehmann et al. 2008) – and so cultural evolution is no more a group level process than genetic evolution. Consequently, the cultural processes modelled by “cultural group selection” can be rewritten and entirely understood from an inclusive fitness or individual perspective. We return to the important process of cultural selection in misconception 18, and focus on genetic selection for the rest of this section. Finally, we also note that it has been suggested that there are even three different types of cultural group selection (Henrich 2004)!

The above discussion shows how the term group selection has been used to mean at least three very different things (Figure 4). 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 ‘newest’) 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 and Grafen 2009), which leads to authors confusingly switching between different meanings (Palmer et al. 1997; Trivers 1998a, b; West et al. 2007a, 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 and Wilson 1998; Wilson et al. 2008), whilst Bergstrom (2002) discusses all three types as if they are the same thing (old: p. 85-86; new: p. 71-72, 76-77, 80; newer: p. 81, 85-86). The lack of a formal theory of group selection contrasts clearly with the large literature formalizing inclusive fitness theory, examining the different ways in which it can be modeled mathematically, how this links to population and quantitative genetic theory, and showing that natural selection will produce individuals that maximize their inclusive fitness (see section 2). The idea that individuals strive to maximise their inclusive fitness holds irrespective of the intensity of selection between-groups; in contrast, group fitness is only maximised in the trivial cases of clonal groups or complete repression of competition within groups (Gardner and Grafen 2009). More generally, any debate about whether selection is at the individual or group level (e.g. (Bergstrom 2002; Burnham and Johnson 2005) is not very useful. Selection always operates at the level of inclusive fitness, and the extent to which this is dependent upon within-group versus between-group selection will vary hugely, depending upon model details and parameter values (Gardner and Grafen 2009; Grafen 2006a).

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 and Wilson 1998; Wilson 2008; Wilson and 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), p.136-140; (Boyd

and Richerson 1990), p.340; (Henrich 2004), p.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 and Richerson 2002; Boyd et al. 2005; Traulsen and Nowak 2006) relate to the inclusive fitness literature on the same issues (reviewed by (Lehmann et al. 2007c; Queller 1992b; West et al. 2002a; West et al. 2008); (ii) that some models (e.g. (Gintis 2000; Wilson and Dugatkin 1997) rely on greenbeard effects, and so are unlikely to be of general importance, especially in humans (see misconception 8) 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).

Misconception 11: 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; Gintis et al. 2003; Henrich 2004; Richerson and Boyd Manuscript (1999)).

This is incorrect. As discussed above, the old group selection ideas have only been shown to work in the extreme scenarios of clonal groups or complete repression of competition within groups, whereas individuals are expected to maximise their inclusive fitness irrespective of the relative strengths of within-group versus between-group selection, and the newer group selection is simply 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. (Boyd et al. 2005), p.215), where 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 6 & 7): “Cooperation among nonkin is commonly explained by one of two mechanisms: repeated interactions (Axelrod and Hamilton 1981; Trivers 1971; Clutton-Brock and Parker 1995) or group selection (Sober and 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 and Richerson 2002).

Misconception 12: Inclusive fitness or kin selection is a subset of group selection.

This is incorrect. No group selection model has ever been constructed where the same result cannot be found with kin selection theory (Table 4). Indeed, whilst 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 testable predictions (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 and Frank 1996; Taylor et al. 2007b).

Misconception 13: Group selection leads to group adaptations (Reeve and Hölldobler 2007; Sober and Wilson 1998; Wilson and Wilson 2007; Wilson and 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 organisms that are adapted to, or look as they were designed to, maximize their fitness. Hamilton's (Hamilton 1964) inclusive fitness theory built upon this showing that the general definition of fitness is inclusive fitness (Grafen 2006a, 2007a). In contrast, 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 and Hölldobler 2007; Sober and Wilson 1998; Wilson 2008; Wilson and Wilson 2007; Wilson and O'Brien 2009; Wilson and Hölldobler 2005; Wynne-Edwards 1962). However, a formal analysis of this problem has shown that 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; (Gardner and Grafen 2009). In contrast, the idea that individuals strive to maximise their inclusive fitness holds irrespective of the intensity of selection operating within and between groups.

Misconception 14: 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 and Wilson 1998; Traulsen and Nowak 2006; Wilson and 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 and Nowak 2006). We cannot stress enough that this is incorrect. Whilst the old group selection idea does not hold (selection does not maximise fitness at the group level except under the very special circumstances described in misconception 13), the new or newer ideas are correct. The reason that most evolutionary biologists, both theoretical and empirical, do not use the group selection approach, is just that it is less useful, and if they express negative views, it is because it has generated more confusion than insight (reviewed in detail by (West et al. 2007a, 2008). 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; (b) all the major developments in social evolution theory have been pioneered and led by the inclusive fitness approach, and not group selection - e.g. worker policing in social insects, kin discrimination in organisms ranging from slime moulds to cooperative breeding vertebrates, alarm calls, parent offspring conflict, sibling conflict, split sex ratios in social insects, selfish genetic

elements, parasite virulence, sex ratios in structured populations (local mate competition), queuing for reproduction in societies, and genomic imprinting (reviewed by (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). Within the social sciences, proponents of both individualist and collectivist schools of thought have used evolutionary theory to support their political world views (Trigger 1998). This debate frequently misunderstands the issues discussed above in misconceptions 10-14, leading to incorrect statements such as that biological evolution works at the individual level and cultural evolution at the group level (see also misconception 18).

6.7 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 and Gintis 2004, 2008; Boyd et al. 2003; Fehr and Gächter 2002; Fehr and Fischbacher 2003; Fehr and Rockenbach 2004; Fehr et al. 2002; Gintis 2000, 2003; Gintis et al. 2003, 2005a). This literature has contributed to 10 misconceptions, numbers 1, 2, 5-9 and the three given below. 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.

A number of elegant economic experiments have suggested that people have a propensity to cooperate, and punish individuals who do not cooperate (reviewed by (Burnham and Johnson 2005; Fehr and Fischbacher 2003; Gächter and 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 according to self-interested assumptions as their behaviour does not maximise their economic payoffs, even if they are given perfect knowledge.

It has been argued that strong reciprocity provides an explanation for cooperation in humans (see misconception 5 for quotations). However, as discussed in misconception 5, this uses a proximate mechanism to solve an ultimate problem. It is circular to argue that a propensity to cooperate and punish those that do not cooperate provides a solution to the problem of cooperation, because it immediately raises the question of why should individuals have a propensity to cooperate and punish? The theoretical models in which this question has been addressed are discussed below in misconception 16.

Misconception 15: Human cooperation in economic games requires the novel evolutionary force of strong reciprocity.

It is also argued that strong reciprocity presents a novel evolutionary force, contrasting with previous inclusive fitness explanations, that strong reciprocity can explain cooperation in one shot encounters between non-relatives, where individuals would do better by cooperating (Fehr and Gächter 2002; Fehr and Fischbacher 2003; Fehr and Rockenbach 2004; Fehr et al. 2002; Gintis et al. 2003, 2005b). For example, it has been claimed: “Strong Reciprocity cannot be rationalized as an adaptive trait by the leading evolutionary theories” (Fehr et al. 2002). There are three issues here. First, the discussion of possible explanations for cooperation focuses on kin selection and reciprocity, and hence ignores the majority of the direct fitness explanations for cooperation that have been developed (misconception 1). Second, as we shall discuss in misconception 16 (see also misconceptions 6-9), the strong reciprocity theoretical models rely on standard direct and indirect benefits (Lehmann et al. 2007b). It is simply that this has not been acknowledged. Third, the models of strong reciprocity do not explain cooperation between non-relatives in one-shot encounters.

The simplest explanation for cooperating and punishing in one-shot encounters is that it is a byproduct of selection for cooperation in other conditions (see also misconception 5). This is analogous to our examples discussed in section 5, such as why a female gorilla protects human children that fall into their pen, dolphins help exhausted swimmers to safety, and enslaved ants rear the broods of their captors. Specifically, humans have evolved a rule-of-thumb behaviour for cooperation and punishment, that improved inclusive fitness in the environment usually encountered, but which may lead to suboptimal performance in artificial laboratory games (Bateson et al. 2006; Binmore 2006; Burnham and Johnson 2005; Hagen and Hammerstein 2006; Haley and Fessler 2005; Nowak et al. 2000; Trivers 2004; West et al. 2007a).

This rule-of-thumb suggestion is supported by a number of observations. When individuals play repeated games, they quickly adjust their level of cooperation, to cooperate less (e.g. approximately 90% of subjects giving nothing in a public goods game), unless there is some mechanism to enforce cooperation such as punishment or reciprocity (Binmore 2005a, 2006; Binmore and Shaked 2009). When players take part in one shot games, in which there are no future interactions, they still adjust their level of cooperation in response to artificial cues, such as the presence of eye-spot pictures on computer desktops, which trigger responses evolved for situations where they are being observed (Bateson et al. 2006; Burnham and Johnson 2005; Haley and Fessler 2005), and physiological factors such as testosterone levels (Burnham 2007). Cultural differences in experimental games (Gächter and Herrmann Manuscript; Henrich et al. 2006; Henrich et al. 2005) 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, punishment can provide a direct benefit if longer periods of interactions are allowed for (Gächter et al. 2008).

More generally, it is well accepted in the evolutionary and behavioural literature that evolution produces rule-of-thumb behaviours that work in the environments in which they evolve, not behaviours that will be perfect in any artificial situation - individuals perform less well in situations that they encounter less frequently (Davies 1992; Herre 1987; Herre et al. 2001; Krebs and McCleery 1984; Pompilio et al. 2006; Wehner 1987); see also the discussion of rationality versus saneness in (Ursprung 1988)). This does not represent maladaptation or a mistake (see below), in the sense that it can be the optimal state, just that the benefits of improving a behaviour have to be balanced or traded-off against the costs (Partridge and 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 regime needs to be considered. Few animal behaviourists or evolutionary biologists would assume that humans have evolved to behave as perfect maximising agents in every situation in which they can be placed.

This rule-of-thumb idea has been argued to be incorrect in several papers, where it is labelled the 'big mistake' or maladaptation hypothesis (Boyd and Richerson 2002; Fehr and 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, such as 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 that can associated with better living conditions (Hagen and 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 (Bateson et al. 2006; Burnham and Hare 2007; Haley and Fessler 2005).

The above discussion has assumed that the economic experiments prove humans cooperate at a level greater than that expected from selfish interests. However, we note that care should be made with the logic upon which this assumption is based. For example, 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 be spiteful punish at a level greater than that expected from selfish interests. Note that our purpose here is not to argue that humans are particularly spiteful (or cooperative), just that it is easy to overinterpret and put very different spins on the data.

Misconception 16: 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 and Gintis 2004; Fehr and Rockenbach 2003, 2004; 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 and 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 and 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 even though as a result they receive lower payoffs than other group members” (Bowles and 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 self regarding are not necessarily self regarding, in that they do not maximize the individual’s direct fitness, as discussed in misconception 2. The other source of confusion is that whilst 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 and West 2004; Lehmann et al. 2007b).

The strong reciprocity models allow cooperation and/or punishment to potentially provide both direct and indirect fitness benefits. As discussed in misconception 2, 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. Indeed, note the similarity here to the ‘group augmentation’ models from the evolutionary literature (see section 5.2), which were developed to show how cooperation could provide a direct fitness benefit. 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. Even with groups of relatively large sizes, relatedness can still take substantial values if the migration rate is low (misconception 7). For example, $r \approx 0.1$ in groups of size 50 if the migration rate is 0.1 (Lehmann et al. 2007b). 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 smaller group sizes of 20, and so relatedness will be higher, but claim that “there are many unrelated individuals, so altruism cannot be explained by inclusive fitness” (Bowles and 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. 2007b). Similarly, punishment can provide a direct or indirect benefit by reducing competition for the actor and 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 and West 2004; Lehmann et al. 2007b). 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, as well as the relative frequencies of cooperators and punishers in the population (Lehmann et al. 2007b). 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, advances in multilocus population genetic methodology (reviewed in (Gardner et al. 2007a) have been exploited 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. 2007b). Considering figure 3, the strong reciprocity models have involved selective forces that occur on multiple branches (e.g. non-enforced direct benefits (group augmentation effects); enforced direct benefits (punishment); indirect benefits by limited dispersal), as well as a branch outside the tree (spiteful green beards).

Misconception 17: 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 have a propensity to cooperate, 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. 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 actually show how competition between groups and limited dispersal can lead to direct and/or indirect benefits to cooperation, analogous to models of group augmentation. None of the theoretical models developed specifically predict cooperation in one-shot encounters, although as discussed, it might be explained as a byproduct from an evolved rule-of-thumb.

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.”

6.8 Cultural Evolution

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 and 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 (notable exceptions occur in bacteria; (Smith 2001; West et al. 2006a). Cultural evolution can be encapsulated with the same mathematical methods as genetical evolution, with the general result that a cultural trait will be selected for if it leads to an increase in cultural fitness of that trait (or meme (Dawkins 1976); for debate around this issue, see (Mesoudi et al. 2006) and related discussion). This leads to an equivalent Hamilton's rule, but where r is the cultural coefficient of relatedness that predicts the tendency of interacting individuals to express the same cultural trait (Lehmann et al. 2008; Werren and Pulliam 1981). The inheritance system (whether particulate or blending) does not affect this (Henrich and Boyd 2002). Cultural selection is frequently assumed to help cooperation spread on the basis that cultural traits can be transmitted horizontally within groups, which could lead to cultural r being higher than genetic r . As we shall see, this is not necessarily the case, depending upon the exact mechanism of cultural transmission.

Misconception 18: Cultural evolution solves the problem of cooperation (Bergstrom 1995; Boyd and Richerson 1985, 2002, 2005, 2006; Boyd et al. 2003; Fehr et al. 2002; Gintis 2003; Henrich 2004; Henrich and Boyd 2001; McElreath and Henrich 2006; Richerson and Boyd 2005).

It is sometimes assumed that cultural evolution is able to explain cooperation in cases where genetic selection cannot. One suggestion is that cultural evolution can make it easier for cooperation to evolve if 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 and Richerson 1985, 2002; Boyd et al. 2003; Henrich 2004). However, in this case, cultural selection actually reduces selection for cooperation, and can even lead to selection for harming behaviours (Feldman et al. 1985; Lehmann et al. 2008; Lehmann et al. 2007b). 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). A different conclusion was reached in earlier papers 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 (Lehmann et al. 2008; Lehmann et al. 2007b); (ii) Boyd & 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 (Lehmann et al. 2008).

Another mechanism for cultural transmission that has been suggested to favour cooperation is conformist transmission, where individuals tend to copy the most common behaviour (Boyd and Richerson 1982, 2005; Henrich 2004; Henrich and Boyd 2001; Richerson and Boyd 2005). This might be thought to favour cooperation because it will lead to a high cultural relatedness within groups. However, because this mechanism

causes individuals to copy the most frequent variant in the group, it strongly selects against cooperation until it has become the dominant phenotype in the population (Lehmann and Feldman 2008a)). This means that cooperation will only be selected for if it was there anyway, and so cultural evolution does not help solve the problem of cooperation.

A final mechanism for cultural transmission is if individuals tend to copy one or a few individuals in their group, such as ‘teachers’, ‘leaders’ or higher status individuals (Cavalli-Sforza and Feldman 1981; Lehmann and Feldman 2008a). In this case, cooperators would tend to occur together in a group, and so would non-cooperators. This would create a high cultural relatedness within groups that would reduce the ability of non-cooperators to exploit cooperators, and hence favour cooperation (Lehmann and Feldman 2008a). For example, if everybody in a group copies a single individual in the group with a probability τ and someone else with a probability $1 - \tau$, then the cultural relatedness between individuals within a group is approximately τ^2 , and is independent of group size (Cavalli-Sforza and Feldman 1981; Lehmann and Feldman 2008a). Consequently, if τ is large, then this can favour altruistic cooperation under conditions where it would not be favoured by genetic selection, such as large group size and high migration rate (Lehmann and Feldman 2008a).

The above makes clear that whether or not cultural evolution promotes cooperation depends upon the exact mechanism of cultural transmission. This raises the question of when will genetic selection favour the evolution of different sorts of mechanisms for cultural transmission? Put another way, when and what form of social learning does genetic selection favour, and how will social learning and cooperation coevolve (Boyd and Richerson 1985; Lehmann and Feldman 2008a; Nettle 2009b)? In particular, when would genetic selection favour the evolution and co-evolution of cultural transmission mechanisms that favour the cultural evolution of cooperation? In a panmictic population, forms of social learning that favour cooperation are not themselves favoured due to their impact upon the evolution of helping (Lehmann and Feldman 2008a). This suggests that a cultural learning mechanism that did favour cooperation would itself have to be favoured for some other reason, such as social learning allowing adaptation to temporally varying environments (Boyd and Richerson 1976, 1985), or that co-evolutionary selection pressures are different in some form of structured population (Lehmann and Feldman 2008a).

More generally, we stress that we are not arguing that cultural evolution cannot help explain cooperation in humans. Instead we are merely pointing out that whether it can help is not as clear as is usually assumed. Whilst there are examples of cultural transmission in the animal world, they pale in comparison to the situation with humans, where there is no doubt that cultural evolution has a huge impact on our behaviour (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981). Humans appear to be unique in the extent to which social learning allows them to adapt to local conditions (rapid social learning). However, with notable exceptions, cultural evolution has been relatively neglected in the biological literature. Determining situations in which genetic selection

favours forms of cultural transmission that select for higher levels of cooperation remains one of the major outstanding problems in this area.

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 (positive science), rather than what they ought to do (normative or regulative science; (Friedman 1953).

7.1 Why do humans cooperate?

The discussion surrounding misconceptions 2 and 7 make it clear that cooperation in humans could have originally evolved via genetic selection due to 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 suggest appreciable relatedness between interacting individuals (Lehmann et al. 2007b). Indeed, a synergy between direct and indirect benefits is also likely - as discussed in section 5.5, direct benefits are often more likely to become important when cooperation is already favoured due to indirect benefits. Higher levels of cooperation may then have been selected for by cultural evolution, although it is still not clear how the necessary form of cultural transmission would have been selected for (see misconception 18).

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. 2007b). We suggest that researchers are unlikely to be able to obtain sufficiently good parameter estimates about ancestral humans to address this problem with sufficient confidence. Indeed, 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 and West 2002; Jennions and Macdonald 1994). In particular, while it appears that the indirect benefits of cooperation vary across cooperative breeding vertebrates, as predicted by Hamilton's rule (Griffin and West 2003), we have a relatively poor understanding of how the direct benefits vary across species, or their importance relative to indirect benefits. Consequently, we suggest that the main point is that both direct and indirect benefits are likely to have been important.

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 6-8, 16); (2) an altruistic group-beneficial trait is being modelled, when actually the trait can be mutually beneficial (misconceptions 1, 2 & 16), or even spiteful and costly at the group level (misconception 8); (3) cultural evolution aids the evolution of helping in situations where it actually hinders it (misconception 18); (4) proximate data provides an answer to an ultimate question (misconception 5). Similar examples can be found elsewhere, such as discussions on how and when selection favours hostility between groups (compare (Choi and Bowles 2007) with (Lehmann and Feldman 2008b)).

We believe that these examples emphasise a potentially huge benefit of greater interactions between the evolutionary and social sciences. Empirical issues such as the proximate versus ultimate debate (example 4) have long been resolved in the behavioural sciences (Mayr 1961; Tinbergen 1963). Theoretical issues, such as those discussed in examples 1-3 of the previous paragraph, stress the advantages of approaches that allow analytical solutions, and the importance of relating new research to that which has gone before. Whilst there are many examples of evolutionary biology benefiting from methods of analysis developed by economists, this appears to be a potential benefit in the other direction. There have been considerable advances over the last 30 years in the methodologies for modelling social evolution (reviewed by (Frank 1997, 1998; Gardner et al. 2007a; Rousset 2004; Taylor 1996; Taylor and Frank 1996; Taylor et al. 2007b), and these methods can provide extremely useful tools for addressing theoretical problems raised by the social sciences, in terms of both tractability and making the evolutionary forces transparent. For example, as recently pointed out by (Lehmann et al. 2008): “where economists strive to evaluate the conditions of invasion and stabilities of mutant strategies by computing the complete distributions of the number of copies of the strategies within and among groups (e.g., Ellison 1993; Kandori et al. 1993), adopting an inclusive fitness approach reduces the problem to the much simpler task of computing the probabilities that pairs of strategies sampled within and among groups are identical”.

7.2 Are humans special?

It is frequently assumed that the form of cooperation in humans is special (Boyd and Richerson 2002; Boyd et al. 2003; Fehr and Fischbacher 2003; Fehr and 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 and Rockenbach 2004) or “Human altruism goes far beyond that which has been observed in the animal world” (Fehr and 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 and Fischbacher 2003). In this section we critically assess the different ways in which human cooperation may be special.

Do humans have especially high levels of altruism (Fehr and Fischbacher 2003; Warneken et al. 2007)? No, a number of organisms have higher levels of altruism than humans, ranging from slime moulds and bacteria to ants and cooperative breeding vertebrates. In both slime moulds 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 slime moulds and bacteria these are the stalk cells, that raise 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 and Franks 1995; Hamilton 1972). In cooperative vertebrates, helping is sometimes mutually beneficial, and sometimes altruistic, depending upon the species (Griffin and 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 and Hatchwell 2004; Russell and 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. Confusion here comes for the mistaken use of altruism, as discussed in misconceptions 1 & 2.

Are humans special because cooperation occurs between non-relatives (Boyd and Richerson 2002; Boyd et al. 2003; Fehr and Fischbacher 2003; Fehr and Rockenbach 2004; Henrich 2004)? No, cooperation between non-relatives occurs in a range of organisms. Many forms of cooperation occur between non-relatives in 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 and West 2002; Jennions and Macdonald 1994; Krebs and Davies 1997). Even in social insects such as ants and wasps, there are some examples where non-relatives come together for mutually beneficial cooperation (Bernasconi and Strassman 1999; Queller et al. 2000). However, perhaps the most extreme examples of cooperation between non-relatives are the various examples of cooperation between species, termed mutualisms (Herre et al. 1999; Sachs et al. 2004; West et al. 2007b). 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 (Table 2), such as the incorporation of symbiotic bacteria that became mitochondria, in the transition to eukaryotes ((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. For example, clients chase and attack cleaner fish that do not cooperate (Bshary and Grutter 2002, 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), and honey bees destroy (police) eggs laid by workers (Ratnieks and Visscher 1989).

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 and Gächter 2002; Fehr and Rockenbach 2003; Fehr and Fischbacher 2003, 2004; Henrich et al. 2005; Semmann et al. 2004; West et al. 2006b). For example, human cognitive abilities allow individuals to alter their level of cooperation in response to whether there is the possibility for punishment (Fehr and 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 (Puurtilinen and Mappes 2009). Importantly, this fine-tuning of behaviour can be done in response to both previous experience (learning) and observations of others (social learning).

Whilst 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 and Hauser 2004; Stevens et al. 2005). More complex and unique mechanisms to enforce cooperation have arisen in humans, such as contracts, laws, justice, trade and norms (Binmore 1994, 1998, 2005b; Boyd and 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. Cultural evolution allows a potential mechanism for testing different strategies, although, as discussed in section 6.8, the extent to which this favours cooperation is still not clear.

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 we have the best cognitive abilities for all behaviours related to cooperation. For example, considering indirect fitness benefits, whilst slime moulds 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), p. 104-106).

Finally, are humans special because of cultural evolution (social learning)? Probably, but maybe not for the reasons that are usually assumed. There is no doubt that cultural evolution is far more important for humans than other organisms, and that it has been relatively ignored by evolutionary biologists (Binmore 2005b; Boyd and Richerson 1985). However, it remains unclear whether this has enabled higher levels of cooperation than favoured by genetic selections. As discussed in misconception 18, whilst cultural evolution can favour higher levels of cooperation, it can also select against cooperation, depending upon the underlying mechanism of cultural transmission (Lehmann and Feldman 2008a; Lehmann et al. 2008). Determining whether and how cultural evolution influences levels of cooperation remains a major outstanding task.

8. Acknowledgements

We thank Jean-Baptiste André, Ken Binmore, John Brenner, Terry Burnham, Dominic Johnson, Laurent Lehmann, Daniel Rankin, Mark Schaffer, Thom Scott-Phillips, Francis Ratnieks and Peyton Young for comments and/or supplying reprints; the Leverhulme Trust, Royal Society, European Research Council and Biotechnology and Biological Sciences Research Council for funding.

References

- Akerlof, G. A. & Kranton, R. A. (2000). Economics and Identity. *Quarterly Journal of Economics*, 115, 715-753.
- Alcock, J. (2005). *Animal Behavior Eighth Edition*. Sunderland, Massachusetts: Sinauer Associates.
- Alexander, R. D. (1974). The evolution of social behavior. *Annu. Rev. Ecol. Syst.*, 5, 325-383.
- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Alvergne, A., Faurie, C. & Raymond, M. (2007). Differential resemblance of young children to their parents: who do children look like more. *Evolution and Human behavior*, 18, 135-144.
- Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal*, 100, 464-477.
- Apicella, C. L. & Marlowe, F. W. (2004). Perceived mate fidelity and paternal resemblance predict men's investment in children. *Evolution and Human behavior*, 15, 371-378.
- 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. (1959). Acceptable points in general cooperative n-person games. In: *Contributions to the Theory of Games IV* (Ed. by Tucker, A. W. & Luce, R. D.), pp. 287-324. Princeton: Princeton University Press.
- Aumann, R. (1981). Survey of repeated games. In: *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern* (Ed. by Bohm, V.), 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. *Am. Nat.*, 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.
- Benabou, R. & Tirole, J. (2004). *Incentives and prosocial behaviour*. Princeton: Princeton University Press.
- Benoit, J.-P. & Krishna, V. (1985). Finitely repeated games. *Econometrica*, 53, 905-922.
- Bergmüller, R., Bshary, R., Johnstone, R. A. & Russell, A. F. (2007). Integrating cooperative 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. & Strassman, J. E. (1999). Cooperation among unrelated individuals: the ant foundress case. *Trends Ecol. Evol.*, 14, 477-482.
- Binmore, K. (1994). *Game Theory And The Social Contract Volume 1: Playing Fair*. Cambridge, Massachusetts: MIT Press.
- Binmore, K. (1998). *Game Theory And The Social Contract Volume 2: Just Playing*. Cambridge, Massachusetts: 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.
- Binmore, K. & Shaked, A. (2009). Experimental economics: where next? *Journal of Economic Behavior & Organization*, in press.
- Bonder, R. & Prelec, D. (2003). Self-signalling and diagnostic utility in everyday decision making. In: *The Psychology of Economic Decisions. Vol. 1. Rationality and Well-being* (Ed. by Brocas, I. & Carrillo, J.), pp. 105-126. 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. & Morgan, E. D. (2003). Informational constraints on optimal sex allocation in ants. *Proc. Natl. Acad. Sci. U.S.A.*, *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. & 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: *The New Palgrave Dictionary of Economics* (Ed. by Blume, L. & Durlauf, S.): 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. (1976). A simple dual inheritance model of the conflict between social and biological evolution. *Zygon*, *11*, 254-262.
- Boyd, R. & Richerson, P. J. (1982). Cultural transmission and the evolution of cooperative behavior. *Human Ecology*, *10*, 325-351.
- Boyd, R. & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Boyd, R. & Lorberaum, J. P. (1987). No pure strategy is evolutionary stable in the repeated Prisoner's Dilemma game. *Nature*, *327*, 58-59.
- Boyd, R. & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, *132*, 337-356.
- Boyd, R. & Richerson, P. J. (1990). Group selection among alternative evolutionarily stable strategies. *J. Theor. Biol.*, *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 structured population. *J. Theor. Biol.*, *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. (Ed. by Levinson, S. & Enfield, N.). Oxford: Berg.
- Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.*, *100*, 3531-3535.
- Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2005). The evolution of altruistic punishment. In: *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (Ed. by Gintis, H., Bowles, S., Boyd, R. & Fehr, E.), pp. 215-227. Cambridge, Massachusetts: MIT Press.
- 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.
- Brown, S. P. & Johnstone, R. A. (2001). Cooperation in the dark: signalling and collective action in quorum-sensing bacteria. *Proc. Roy. Soc. Lond. B*, *268*, 961-965.

- Bshary, R. (2002). Biting cleaner fish use altruism to deceive image-scoring client reef fish *Proc. Roy. Soc. Lond. B*, 269, 2087-2093.
- 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. (2002). Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, 63, 547-555.
- Bshary, R. & Grutter, A. S. (2005). Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters*, 1, 396-399.
- Bshary, R. & Bergmüller, R. (2008). Distinguishing four fundamental approaches to the evolution of helping. *Journal of Evolutionary Biology*, 21, 405-420.
- Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society of London Series B*, 274, 2327-2330.
- Burnham, T. C. & Johnson, D. P. (2005). The biological and evolutionary logic of human cooperation. *Analyse & Kritik*, 27, 113-135.
- Burnham, T. C. & Hare, B. (2007). Engineering human cooperation: does involuntary neural activation increase public goods contributions? *Human Nature*, 18, 88-108.
- Burt, A. & Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish genetic Elements*. Cambridge, Massachusetts: Harvard University Press.
- Cavalli-Sforza, L. L. & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton: Princeton University Press.
- Cernoch, J. M. & Porter, R. H. (1985). Recognition of maternal axillary odors by infants. *Child Development*, 56, 1593-1598.
- 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. (In press). Reciprocity and social investment in animal societies
- 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. *A. Rev. Ecol. Syst.*, 29, 141-177.
- Colwell, R. K. (1981). Group selection is implicated in the evolution of female-biased sex ratios. *Nature*, 290, 401-404.
- Crespi, B. J. (2001). The evolution of social behavior in microorganisms. *Trends Ecol. Evol.*, 16, 178-183.
- Crozier, W. R. (2001). Blushing and the exposed self: Darwin revisited. *Journal for the Theory of Social Behaviour*, 31, 61-72.
- Daly, M. & Wilson, M. (1982a). Whom are newborn babies said to resemble? *Ethology and Sociobiology*, 3, 69-78.
- Daly, M. & Wilson, M. (1982b). Homicide and kinship. *American Anthropologist*, 84, 372-378.
- 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. *Z. Tierpsychol.*, *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. *Q. Rev. Biol.*, *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.
- Edwards, A. W. F. (1994). The fundamental theorem of natural selection. *Biological Reviews*, *69*, 443-474.
- Eisenberg, A. R. (1988). Grandchildren's perspectives on relationships with grandparents: the influence of gender across generations. *Sex Roles*, *19*, 205-217.
- Emlen, S. T. (1997). Predicting family dynamics in social vertebrates. In: *Behavioural Ecology* (Ed. by Krebs, J. R. & Davies, N. B.), pp. 228-253. Oxford, UK: Blackwell.
- Ewens, W. J. (1989). An interpretation and proof of the Fundamental Theorem of Natural Selection. *Theoretical Population Biology*, *36*, 167-180.
- Falk, A. & Fischbacher, U. (2001 (MS)). A theory of reciprocity. *CEsifo Working Paper Np. 457*.
- 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. & Henrich, J. (2003). Is strong reciprocity a maladaptation? In: *Genetic and Cultural Evolution of Cooperation* (Ed. by Hammerstein, P.), pp. 55-82. Cambridge, MA: MIT Press.
- Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785-791.
- Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63-87.
- Fehr, E. & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology*, *14*, 784-790.
- 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. (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. (2006). A general model for the evolution of mutualisms. *Journal of Evolutionary Biology*, 19, 1283-1293.
- Foster, K. R., Wenseleers, T. & Ratnieks, F. L. W. (2006). Kin selection is the key to altruism. *Trends Ecol. Evol.*, 21, 57-60.
- Frank, R. H. (1987). If *Homo Economicus* could choose his 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. *Theoret. Popul. Biol.*, 29, 312-342.
- Frank, S. A. (1995a). George Price's contributions to evolutionary genetics. *J. Theor. Biol.*, 175, 373-388.
- Frank, S. A. (1995b). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, 377, 520-522.
- Frank, S. A. (1996a). Policing and group cohesion when resources vary. *Anim. Behav.*, 52, 1163-1169.
- Frank, S. A. (1996b). Models of parasite virulence. *Q. Rev. Biol.*, 71, 37-78.
- Frank, S. A. (1997). Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *J. Theor. Biol.*, 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. & Slatkin, M. (1992). Fisher's fundamental theorem of natural selection. *Trends in Ecology & Evolution*, 7, 92-95.
- Friedman, M. (1953). The methodology of positive economics In: *Essays in Positive Economics* (Ed. by Friedman, M.), 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.
- Fudenberg, D. & Maskin, E. (1990). Evolution and cooperation in noisy repeated games. *New Developments in Economic Theory*, 80, 274-279.
- Fudenberg, D., Levine, D. & Maskin, E. (1994). The folk theorem with imperfect public information. *Econometrica*, 62, 997-1039.
- Gächter, S. & Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Phil. Trans. R. Soc. Lond. B*, 364, 791-806.
- Gächter, S. & Herrmann, B. (Manuscript). The limits of self-governance when cooperators get punished: experimental evidence from urban and rural Russia.
- 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. & 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. (In press). Greenbeards. *Evolution*.

- Gardner, A., West, S. A. & Buckling, A. (2004). Bacteriocins, spite and virulence. *Proc. Roy. Soc. Lond. 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. & Alpedrinha, J. (2009). Budding dispersal and the sex ratio. *Journal of Evolutionary Biology*, in press.
- 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., Strassman, J. E. & Queller, D. C. (2007). High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 8913-8917.
- Gintis, H. (2000). Strong reciprocity and human sociality. *J. Theor. Biol.*, 206, 169-179.
- Gintis, H. (2003). The hitchhiker's guide to altruism: gene-culture coevolution, and the internalization of norms. *J. Theor. Biol.*, 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: *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (Ed. by Gintis, H., Bowles, S., Boyd, R. & Fehr, E.), pp. 3-39. Cambridge, Massachusetts: MIT Press.
- Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, 298, 425-426.
- Grafen, A. (1984). Natural selection, kin selection and group selection. In: *Behavioural Ecology: An Evolutionary Approach* (Ed. by Krebs, J. R. & Davies, N. B.), pp. 62-84. Oxford, UK: Blackwell Scientific Publications.
- Grafen, A. (1985). A geometric view of relatedness. *Oxford Surv. Evol. Biol.*, 2, 28-89.
- Grafen, A. (1990a). Biological signals as handicaps. *J. Theor. Biol.*, 144, 517-546.
- Grafen, A. (1990b). Do animals really recognise kin? *Anim. Behav.*, 39, 42-54.
- Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proc. Roy. Soc. Lond. B*, 266, 799-803.
- Grafen, A. (2002). A first formal link between the Price equation and an optimization program. *J. Theor. Biol.*, 217, 75-91.
- Grafen, A. (2003). Fisher the evolutionary biologist. *The Statistician*, 52, 319-329.
- Grafen, A. (2006a). Optimisation of inclusive fitness. *Journal of Theoretical Biology*, 238, 541-563.
- Grafen, A. (2006b). A theory of Fisher's reproductive value. *J. Math. Biol.*, 53, 15-60.
- Grafen, A. (2007a). The formal Darwinism project: a mid-term report. *Journal of Evolutionary Biology*, 20, 1243-1254.
- Grafen, A. (2007b). An inclusive fitness analysis of altruism on a cyclical network. *Journal of Evolutionary Biology*, 20, 2278-2283.

- Grafen, A. (2007c). Detecting kin selection at work using inclusive fitness. *Proc. Roy. Soc. Lond. B*, 274, 713-719.
- Griffin, A. S. & West, S. A. (2002). Kin selection: fact and fiction. *Trends Ecol. Evol.*, 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.
- 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. *Am. Nat.*, 97, 354-356.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour, I & II. *J. Theor. Biol.*, 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: *Man and Beast: Comparative Social Behavior* (Ed. by Eisenberg, J. F. & Dillon, W. S.), pp. 57-91. Washington, DC: Smithsonian Press.
- Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annu. Rev. Ecol. Syst.*, 3, 193-232.
- Hamilton, W. D. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In: *Biosocial Anthropology* (Ed. by Fox, R.), 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.
- Hauber, M. E. & Sherman, P. W. (2001). Self-referent phenotype matching: theoretical considerations and empirical evidence. *Trends in Neurosciences*, 24, 609-616.
- Hauk, E. (2001). Leaving the prison: permitting partner choice and refusal in prisoner's dilemma games. *Computational Economics*, 18, 65-87.
- Helanterä, H. & Sundström, L. (2007). Worker policing and nest mate recognition in the ant *Formica fusca*. *Behav. Ecol. Sociobiol.*, 61, 1143-1149.
- Helanterä, H. & Bargum, K. (2007). Pedigree relatedness, not greenbeard genes, explains eusociality. *Oikos*, 116, 217-220.
- 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. & Boyd, R. (2002). Culture and cognition: why cultural evolution does not require replication of representations. *Culture and Cognition*, 2, 87-112.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767-1770.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q. & Tracer, D. (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: *Adaptionism and Optimality* (Ed. by Orzack, S. & Sober, E.), pp. 191-218. Cambridge: 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 Ecol. Evol.*, 14, 49-53.
- 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 selection is the key to the evolution of eusociality. *Science*, 320, 1213-1216.
- Jennions, M. D. & Macdonald, D. W. (1994). Cooperative breeding in mammals. *Trends Ecol. Evol.*, 9, 89-93.
- Kahana, B. & Kahana, E. (1970). Grandparenthood from the perspective of the developing grandchild. *Developmental Psychology*, 3, 98-105.
- 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 commons. *Proc. Roy. Soc. Lond. B*, 273, 1477-1481.
- Knight, C. (1998). Ritual/speech coevolution: a solution to the problem of deception. In: *Approaches to the Evolution of Language* (Ed. by Hurford, J. R., Studdert-Kennedy, M. & Knight, C.). 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. *Proc. R. Soc. Lond. B*, 268, 187-196.
- Krebs, J. R. & McCleery, R. H. (1984). Optimization in behavioural ecology. In: *Behavioural Ecology, 2nd edn.* (Ed. by Krebs, J. R. & Davies, N. B.). Oxford: Blackwell.
- Krebs, J. R. & Davies, N. B. (1987). *An Introduction to Behavioural Ecology, Second Edition*, 2nd edn. Oxford, England: Blackwell Scientific Publications.
- Krebs, J. R. & Davies, N. B. (1993). *An Introduction to Behavioural Ecology, Third Edition*, Third edn. 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.
- Lack, D. (1966). *Population Studies of Birds*. Oxford: Clarendon 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). The co-evolution of culturally inherited altruistic helping and cultural transmission under random group formation. *Theoretical Population Biology*, 73, 506-516.
- Lehmann, L. & Feldman, M. W. (2008b). War and the evolution of belligerence and bravery. *Proceedings of the Royal Society of London, Series B*, in press.
- Lehmann, L., Bargum, K. & Reuter, M. (2006). An evolutionary analysis of the relationship between spite and altruism. *Journal of Evolutionary Biology*, 19, 1507-1516.
- 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., Rousset, F., Roze, D. & Keller, L. (2007b). Strong-reciprocity or strong-ferocity? A population genetic view of the evolution of altruistic punishment. *American Naturalist*, 170, 21-36.
- Lehmann, L., Keller, L., West, S. A. & Roze, D. (2007c). Group selection and kin selection. Two concepts but one process. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 6736-6739.
- 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? *Proc. Natl. Acad. Sci. (USA)*, 80, 2985-2989.
- Leigh, E. G. (1991). Genes, bees and ecosystems: the evolution of a common interest among individuals. *Trends in Ecol. & Evol.*, 6, 257-262.
- Leimar, O. & Hammerstein, P. (2006). Facing the facts. *Journal of Evolutionary Biology*, 19, 1403-1405.

- Lessard, S. (1997). Fisher's fundamental theorem of natural selection revisited. *Theoretical Population Biology*, 52, 119-136.
- 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. *Proc. Roy. Soc. Lond. B*, 270, 819-826.
- Lieberman, D., Tooby, J. & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445, 727-731.
- 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.
- 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. *Proc. Roy. Soc. Lond. B*, 269, 721-727.
- Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, 201, 1145-1147.
- Maynard Smith, J. (1976). Group selection. *Q. Rev. Biol.*, 51, 277-283.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- Maynard Smith, J. (1983). Models of evolution. *Proc. Roy. Soc. Lond. B*, 219, 315-325.
- 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.
- McBurney, D. H., Simon, J., Gaulin, S. J. C. & Geliebter, A. (2002). Matrilateral biases in the investment of aunts and uncles replication in a population presumed to have high paternity certainty. *Human Nature*, 13, 391-402.
- McElreath, R. & Henrich, J. (2006). Modeling cultural evolution. In: *Oxford Handbook of Evolutionary Biology* (Ed. by Dunbar, R. I. M. & Barrett).
- McLain, D. K., Setters, D., Moulton, M. P. & Pratt, A. E. (2000). Ascription of resemblance of newborns by parents and nonrelatives. *Evolution and Human Behavior*, 21, 11-23.
- Mehdiabadi, N. J., Jack, C. N., Farnham, T. T., Platt, T. G., Kalla, S. E., Shaulsky, G., Queller, D. C. & Strassman, J. S. (2006). Kin preference in a social microbe. *Nature*, 442, 881-882.
- Mesoudi, A., Whiten, A. & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, 29, 329-383.
- 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 defection 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. (2009a). *Evolution and Genetics for Psychology*. Oxford: Oxford University Press.
- Nettle, D. (2009b). Beyond nature versus culture: cultural variation as an evolved characteristic. *Journal of the Royal Anthropological Institute*, in press.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560-1563.
- Nowak, M. A. & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*, 826-829.
- Nowak, M. A. & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*, 56-58.
- 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., Bonhoeffer, S. & May, R. M. (1994). Spatial games and the maintenance of cooperation. *Proc. Natl. Acad. Sci. U.S.A.*, *91*, 4877-4881.
- Nowak, M. A., Page, K. M. & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, *289*, 1773-1775.
- Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, *428*, 646-650.
- 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.
- 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.
- Owren, M. J. & Bachorowski, J.-A. (2001). The evolution of emotional expression: a "selfish-gene" account of smiling and laughter in early hominids and humans In: *Emotions: Current Issues and Future Directions* (Ed. by Mayne, T. J. & Bonanno, G. A.). New York: The Guilford Press.
- Palameta, B. & Brown, W. M. (1999). Human cooperation is more than by-product mutualism. *Animal Behaviour*, *57*, F1-F3.

- 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. *Phil. Trans. Roy. Soc. Lon. B*, 332, 3-13.
- Pellmyr, O. & Huth, C. J. (1994). Evolutionary stability of mutualism between yuccas and yucca moths. *Nature*, 372, 257-260.
- Pfennig, D. W., Collins, J. P. & Ziemba, R. E. (1999). A test of alternative hypotheses for kin recognition in cannibalistic tiger salamanders. *Behav. Ecol.*, 10, 436-443.
- Platak, S. M., Keenan, J. P. & Mohamed, F. B. (2005). Sex differences in the neural correlates of child facial resemblance: an event-related fMRI study *NeuroImage*, 25, 1336-1344.
- Platak, S. M., Raines, D. M., Gallup Jr, G. G., Mohamed, F. B., Thomson, J. W., Myers, T. E., Panyavin, I. S., Levin, S. L., Davis, J. A., Fonteyn, L. C. M. & Arigo, D. R. (2004). Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains. *Evolution and Human behavior*, 25, 394-405.
- Pollet, T. V., Nettle, D. & Nelissen, M. (2007). Maternal Grandmothers do go the Extra Mile: Factoring Distance and Lineage into Differential Contact with Grandchildren. *Evolutionary Psychology*, 5, 832-843.
- Pompilio, L., Kacelnik, A. & Behmer, S. T. (2006). State-dependent learned valuation drives choice in an invertebrate. *Science*, 311, 1613-1615.
- Porter, R. H. & Cernoch, J. M. (1983). Maternal recognition of neonates through olfactory cues. *Physiology & Behavior*, 30, 151-154.
- Porter, R. H., Balogh, R. D., Cernoch, J. M. & Franchi, C. (1986). Recognition of kin through characteristic body odors. *Chemical Senses*, 11, 389-395.
- Poundstone, W. (1988). *Labyrinths of Reason*. New York: Doubleday.
- Price, G. R. (1972). Fisher's 'fundamental theorem' made clear. *Ann. hum. Genet.*, 35, 485-490.
- Puurttinen, 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? *Trend. Ecol. Evol.*, 7, 322-324.
- Queller, D. C. (2000). Relatedness and the fraternal major transitions. *Phil. Trans. R. Soc. Lond. 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. & Strassmann, J. E. (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. & Fehr, E. (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. *Annu. Rev. Entomol.*, *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.
- Regalski, J. M. & Gaulin, S. J. C. (1993). Whom are Mexican infants said to resemble? Monitoring and fostering paternal confidence in the Yutacan. *Ethology and Sociobiology*, *14*, 97-113.
- 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. (Manuscript (1999)). Complex societies: the evolutionary origins of a crude superorganism. *Human Nature*.
- Robson, A. J. (2008). Group selection. In: *The New Palgrave Dictionary of Economics. Second Edition*. (Ed. by Durlauf, S. & Blume, L. E.): Palgrave Macmillan.
- Rose, M. R. & Lauder, G. V. (1996). *Adaptation*. San Diego: Academic Press.
- Rossi, A. S. & Rossi, P. H. (1990). *Of Human Bonding: Parent-Child Relations across the Life Course*. New York: Aldine.
- 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.
- Russel, D. E. H. (1976). Human olfactory communication. *Nature*, *260*, 520-522.
- Russel, M. J., Mendelson, T. & Peeke, H. V. S. (1983). Mothers' identification of their infant's odor. *Ethology and Sociobiology*, *4*, 29-31.
- Russell, A. F. & Hatchwell, B. J. (2001). Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proc. Roy. Soc. Lond. 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. *Q. Rev. Biol.*, *79*, 135-160.
- 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.
- Smith, J. (2001). The social evolution of bacterial pathogenesis. *Proc. Roy. Soc. Lond. 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, Mass.: 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. *Q. J. Econ.*, 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. *Annu. Rev. Ecol. Evol. Syst.*, 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. *Am. Nat.*, 135, 95-106.
- Taylor, P. D. (1996). Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.*, 34, 654-674.
- Taylor, P. D. & Frank, S. A. (1996). How to make a kin selection model. *J. Theor. Biol.*, 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. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 10952-10955.
- Trigger, B. (1998). *Sociocultural Evolution: Calculation and Contingency*. Oxford: Blackwell.
- Trivers, R. (2004). Mutual benefits at all levels of life. *Science*, 304, 964-965.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.*, 46, 35-57.
- Trivers, R. L. (1974). Parent-offspring conflict. *Am. Zool.*, 14, 249-264.
- Trivers, R. L. (1998a). As they would do to you. *Skeptic*, 6, 81-83.
- Trivers, R. L. (1998b). Think for yourself. *Skeptic*, 6, 86-87.
- Ursprung, H. W. (1988). Evolution and the economic approach to human behavior. *J. Social Biol. Struct.*, 11, 257-279.

- van Rooy, R. (2003). Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. In: *Proc. of the 4th Conference on Theoretical Aspects of Rationality and Knowledge* (Ed. by Tennenholtz, M.), 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*, *5*, e184.
- 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. *J Comp Physiol A*, *161*, 511-531.
- Weibull, J. W. (1995). *Evolutionary Game Theory*. Cambridge, Massachusetts: MIT Press.
- Wenseleers, T. & Ratnieks, F. L. W. (2006). Enforced altruism in insect societies. *Nature*, *444*, 50.
- Wenseleers, T., Helanterä, 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.
- Werren, J. H. & Pulliam, H. R. (1981). An intergenerational transmission model for the cultural evolution of helping behavior. *Human Ecology*, *9*, 465-493.
- West, S. A. (2009). *Sex Allocation*. Princeton: Princeton University Press.
- 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). 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. (2007b). Evolutionary explanations for cooperation. *Current Biology*, *17*, R661-R672.
- 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? *J. Evol. Biol.*, *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., Diggle, S. P., Buckling, A., Gardner, A. & Griffin, A. S. (2007c). The social lives of microbes. *Annual Review of Ecology, Evolution and Systematics*, *38*, 53-77.
- West, S. A., Gardner, A., Shuker, D. M., Reynolds, T., Burton-Chellow, M., Sykes, E. M., Guinnee, M. A. & Griffin, A. S. (2006b). Cooperation and the scale of competition in humans. *Current Biology*, *16*, 1103-1106.
- Wild, G. (2008). Toward evolutionary graphs with two sexes: a kin selection analysis of a sex allocation problem. *Journal of Evolutionary Biology*, *21*, 1428-1437.
- 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. *Proc. Natl. Acad. Sci. U.S.A.*, 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. *Am. Nat.*, 149, 336-351.
- Wilson, D. S. & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.*, 327-348.
- Wilson, D. S. & O'Brien, D. T. (2009). Evolutionary theory and cooperation in everyday life. In: *Games, Groups, and the Global Good* (Ed. by Levin, S. A.). 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. *Proc. Natl. Acad. Sci. U.S.A.*, 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. *Am. Nat.*, 153, 254-266.
- Woolfenden, G. E. (1975). Florida scrub jay helpers at the nest. *The Auk*, 92, 1-15.
- Wynne-Edwards, V. C. (1962). *Animal Dispersion in Relation to Social Behaviour*. Edinburgh: Oliver and Boyd.
- 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. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 12005-12010.
- Young, H. P. (2003). The power of norms. In: *Genetic and Cultural Evolution of Cooperation* (Ed. by Hammerstein, P.). Cambridge: MIT Press.
- Young, H. P. & Foster, D. (1991). Cooperation in the short and long run. *Games and Economic Behavior*, 3, 145-156.
- Zeder, M. A. (2009). The Neolithic macro-(r)evolution: macroevolutionary theory and the study of culture change. *J Archaeol Res*, 17, 1-63.

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 and Lauder 1996)
Altruism	A behaviour that is costly to the actor and beneficial to the recipient. Cost and benefit 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. 2007b).
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.
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. 2007b).
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 (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).

Table 1. Glossary.

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

Table 2. Social Behaviours. A Hamiltonian classification scheme for social behaviours that have been selected for by natural selection (West et al. 2007a).

Replicating molecules	⇒	Populations of molecules in compartments
Independent replicators	⇒	Chromosomes
RNA as gene and enzyme	⇒	DNA & protein (genetic code)
Prokaryotes	⇒	Eukaryotes (cell nucleus & organelles)
Asexual clones	⇒	Sexual populations
Protists	⇒	Multicelled animals, plants & fungi (with cell differentiation, i.e. organs)
Solitary individuals	⇒	Colonies (with non-reproductive worker castes)
Primate societies	⇒	Human societies (language)

Table 3. The major evolutionary transitions. It is generally accepted that there have been eight major evolutionary transitions, each of which has led to a new level of organismal complexity (Maynard Smith and Szathmary 1995). Whilst major transitions can lead to a marked reduction in within-group competition, there can still be potential for conflicts, such as between genes in a genome or individuals in a society. The problem (of cooperation) in most of these transitions, is why did selection on entities to maximize their fitness at the lower level not prevent integration at the higher levels? The potential solutions are discussed in sections 5 and 6, and the care needed with defining adaptation at a group level is discussed in detail elsewhere (Gardner and Grafen 2009). Cultural evolution is not included in the table, because it is a separate evolutionary process not a new biological transition.

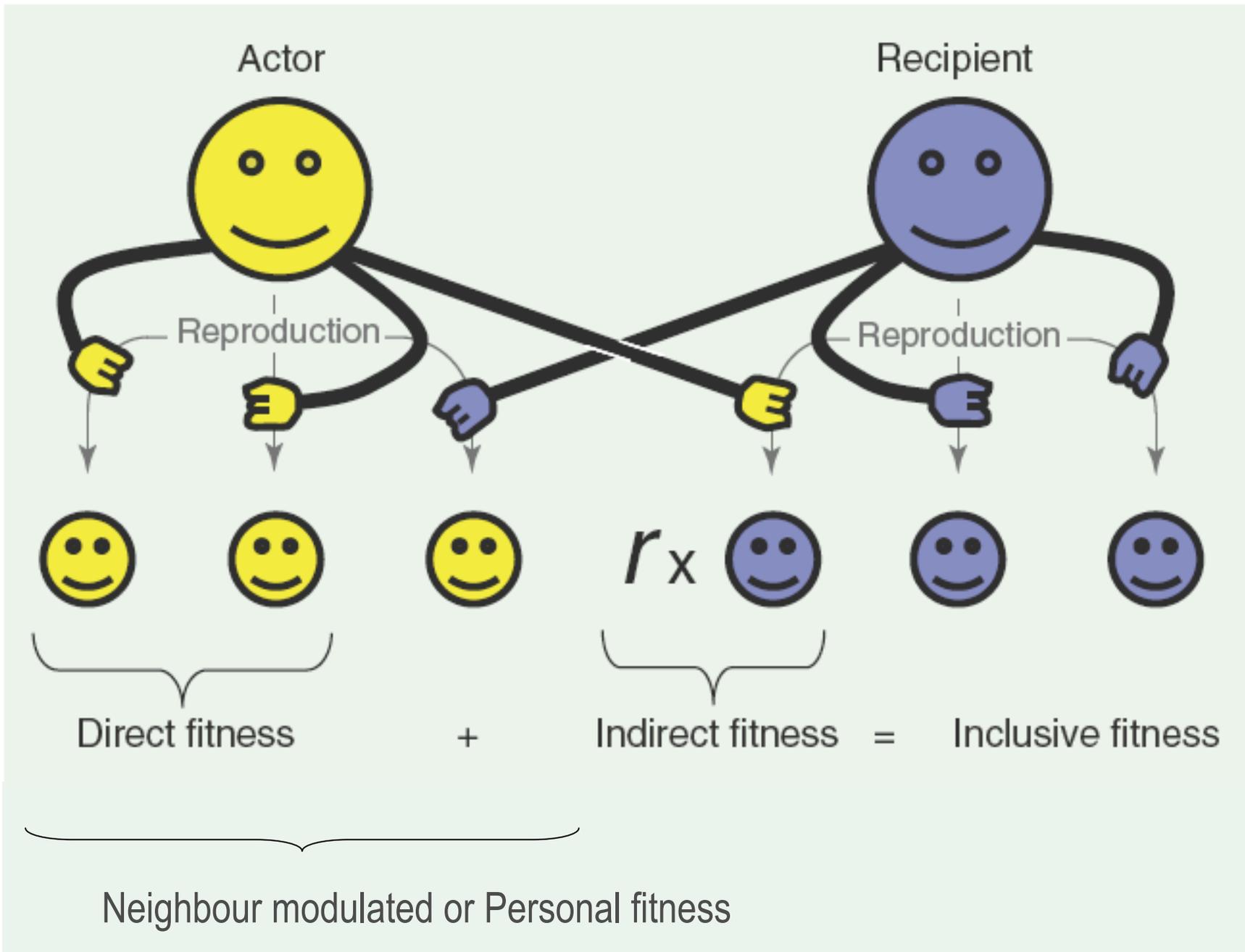
Misconception	Reality
1. Kin selection and reciprocity are the major competing explanations for altruism in biological theory.	In the context of reciprocity, cooperation is not altruistic, and there are many other mechanisms by which cooperation can be favoured due to direct fitness benefits.
2. The 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 hides the underlying selective forces.
3. The evolution of cooperation is encapsulated by the Prisoners' Dilemma (PD) and how it can be solved by the behavioural strategy 'tit-for-tat'.	The PD is only a special case; the disproportionate focus on this game has hindered progress. Tit-for-tat is not the evolutionary stable strategy in the iterated PD game
4. 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.
5. Proximate and ultimate explanations.	Proximate answers cannot provide a solution to ultimate problems.
6. Kin selection requires kin discrimination.	A sufficiently high relatedness can also arise through limited dispersal.
7. 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.
8. 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.
9. Greenbeards are a type of costly signaling	Greenbeards and costly signalling are two different things.
10. Group selection is a formal theory with one meaning.	Group selection is used to mean at least four different things.
11. 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.
12. Kin selection is a subset of group	No group selection model has ever been

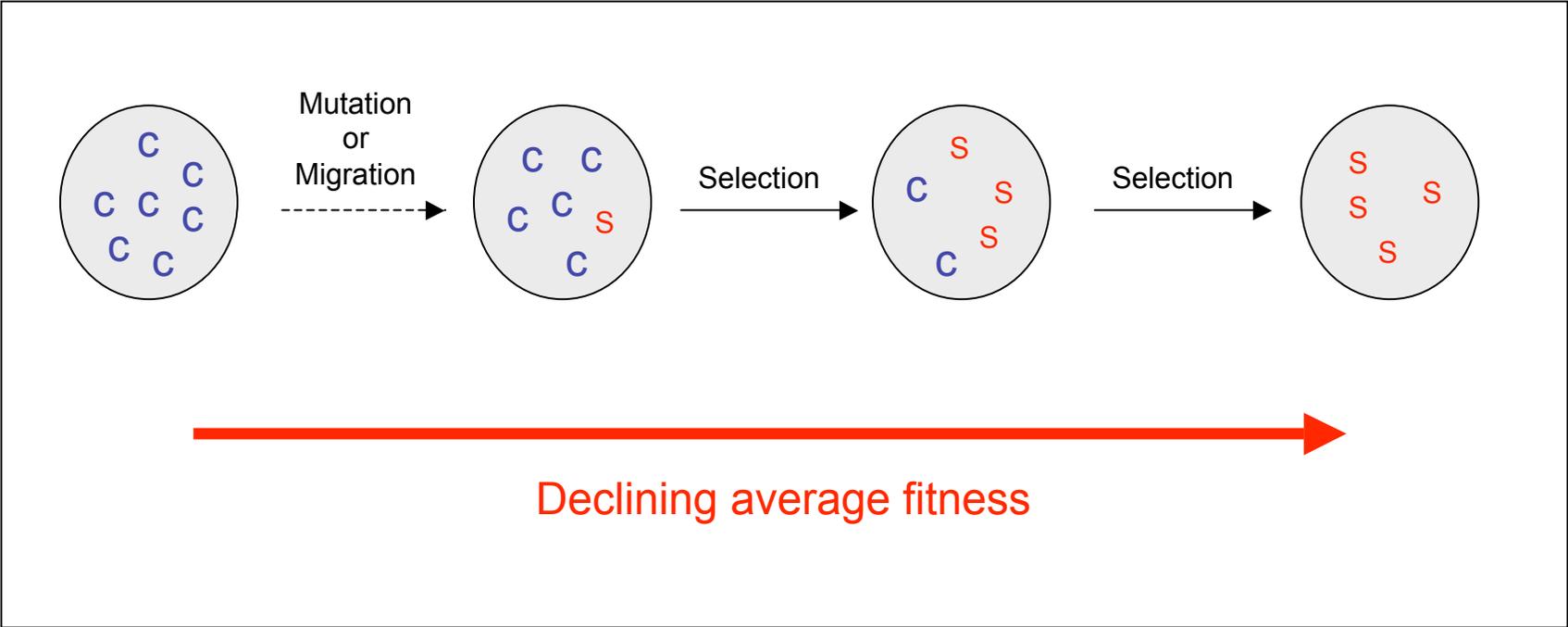
selection.	constructed where the same result cannot be found with kin selection theory. The reverse is not necessarily true.
13. Group selection leads to group adaptations	Group selection will only lead to group adaptations in the special circumstances where either: (a) the group is composed of genetically identical individuals (clonal groups, $r=1$), or (b) there is complete repression of competition between groups (i.e. no conflict within groups).
14. 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.
15. Human cooperation in economic games requires the novel evolutionary force of strong reciprocity.	The simplest explanation for cooperating and punishing in one-shot encounters is that it is a byproduct of selection for cooperation in other conditions.
16. 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.
17. 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.
18. Culture is a magic bullet	In some cases where it has been claimed that cultural evolution will ease the evolution of cooperation, it is actually more likely to hinder it, and may even favour harming behaviours.

Table 4. Eighteen common misconceptions about social evolution theory.

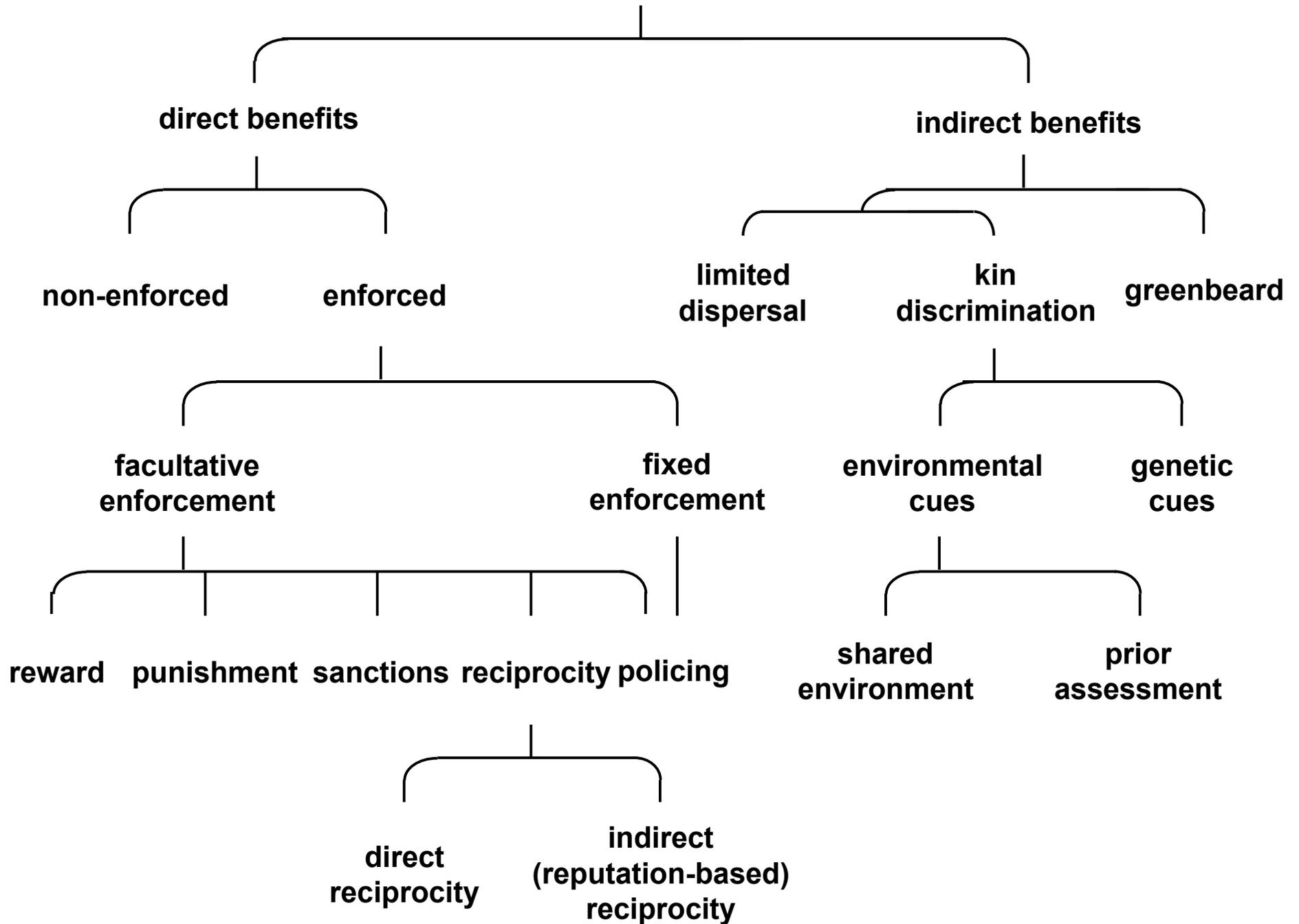
Area	Papers claiming that a result requires group selection	Papers showing equivalent result can be obtained with kin selection / inclusive fitness
Sex ratios with local mate competition	(Colwell 1981; Sober and Wilson 1998; Wilson and Colwell 1981)	(Frank 1986; Grafen 1984; Harvey et al. 1985)
Sex ratios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

Table 5. The equivalence of kin selection and new group section. 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).

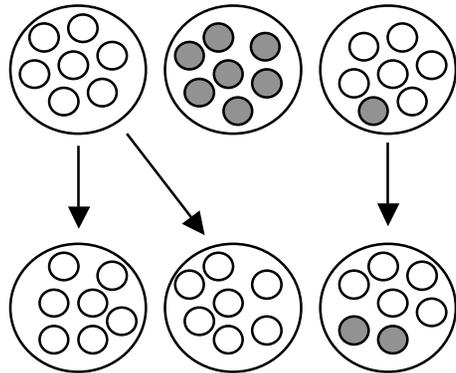




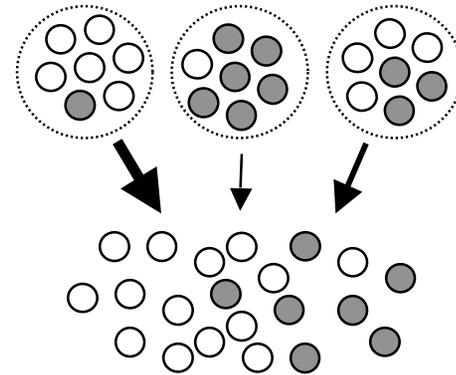
cooperation



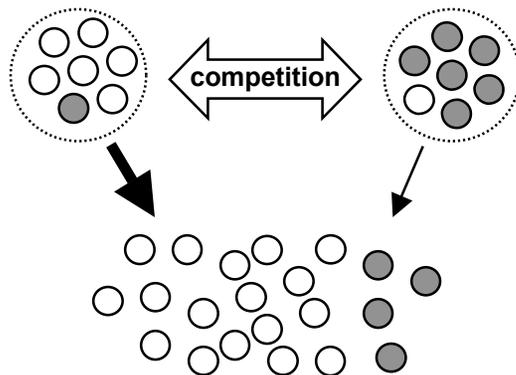
A) “old” group selection



B) “new” group selection



C) “newer” group selection



D) cultural group selection

