REVIEW

# Questioning the cultural evolution of altruism

## J.-B. ANDRÉ*‡ & O. MORIN†

*CNRS UMR7206 Eco-anthropologie et ethnobiologie, MNHN, Université Paris Diderot, F75005 Paris, France
†Department of Cognitive Science, Central European University, Budapest, Hungary
‡Laboratoire Ecologie et Evolution, UMR 7625, CNRS, Ecole Normale Supérieure, Paris, France

## Abstract

The evolutionary foundations of helping among nonkin in humans have been the object of intense debates in the past decades. One thesis has had a prominent influence in this debate: the suggestion that genuine altruism, strictly defined as a form of help that comes at a net fitness cost for the benefactor, might have evolved owing to cultural transmission. The gene–culture coevolution literature is wont to claim that cultural evolution changes the selective pressures that normally act to limit the emergence of altruistic behaviours. This paper aims to recall, however, that cultural transmission yields altruism only to the extent that it relies on maladaptive mechanisms, such as conformist imitation and (in some cases) payoff-biased transmission. This point is sometimes obscured in the literature by a confusion between genuine altruism, maladaptive by definition, and mutualistic forms of cooperation, that benefit all parties in the long run. Theories of cultural altruism do not lift the selective pressures weighing on strictly altruistic actions; they merely shift the burden of maladaptation from social cognition to cultural transmission.

## Introduction

In models or verbal arguments, many scholars have suggested that cooperation in human societies cannot emerge solely because of its benefits to individuals. Mechanisms such as reciprocity or reputation, they argue, do not suffice (e.g. Boyd & Richerson, 1985, ch. 7; Boyd & Richerson, 1988; Fehr & Fischbacher, 2003). They have thus proposed that genuinely altruistic tendencies must constitute, in a way or another, the cement of human societies. These tendencies are often referred to as 'strong reciprocity' (Gintis, 2000; Gintis et al., 2003). Human beings, so they argue, are willing to sacrifice time and resources to the exclusive benefit of others, and this is why they are able to sustain cooperation on a large scale (Fehr & Fischbacher, 2003; Gintis et al., 2003).

Not only does this suggestion entail that human beings have altruistic motives in the psychological sense, it entails that they express altruistic behaviours in the *evolutionary* sense of that term: behaviours that increase

*Correspondence:* Jean-Baptiste André, 46 Rue d'Ulm, 75005 Paris, France.
Tel.: +33 14 432 2341; fax: +33 14 427 2694;
e-mail: jeanbaptisteandre@gmail.com

the biological fitness of other(s) at a net fitness cost for the benefactor (Hamilton, 1964; see also West et al., 2011, section 6.1.1). However, evolutionary theory predicts that altruism can evolve only if it is preferentially expressed towards genetically related partners (Hamilton, 1964; and see also Rousset, 2004; Lehmann & Keller, 2006; West et al., 2007a,b), and the genetic relatedness within most human societies is generally considered insufficient for this to occur. Therefore, the claim that human societies, but not other animal societies, rely upon genuine altruism must be backed by some specific evolutionary arguments. Two quite independent lines of research have proposed that the solution to this puzzle could be found in the fact that human social behaviours are culturally rather than genetically transmitted. In the general attempt to ground social sciences into naturalistic foundations, they play an important role by constituting arguments in favour of an altruistic view of human cooperation.

In this paper, our aim is to discuss, clarify and question the common hypotheses at the core of all such cultural explanations of altruism. We will first describe the mechanisms that may allow cultural transmission to favour the evolution of genuine altruism, even towards

nongenetic kin (section 'The Genetic and Cultural Evolution of Altruism').

Next, we will lay down one key assumption that any model of cultural altruism needs: social learning places strong constraints on our capacity to behave adaptively. In other words, altruism is a maladaptive side effect of social learning (section 'When Biological Altruism Evolves Culturally, it is Always a Maladaptive Byproduct of Imitation').

As a consequence, biological altruism cannot be conceived as arising from a process of gene–culture coevolution. If anything, genetic adaptation should prevent the spread of cultural altruism. This fact is not straightforwardly denied in the literature we review, but we shall give examples of ambiguous claims that seem to suggest the contrary (section 'Ambiguities in the Literature'). One important source of misunderstanding stems from confusions between mutualistic cooperation that benefit all individuals, and genuine biological altruism (section 'Cultural Group Selection: Disentangling Altruism and Cooperation').

We will conclude that any explanation for the existence of altruistic behaviours expressed towards nonkin must involve a lack of genetic adaptation. The maladaptation may come either from an inability to weigh perfectly the cost and benefits of social actions (maladaptations of social cognition) or from a tendency to imitate maladaptive behaviours indiscriminately (maladaptations of cultural cognition). However, cultural transmission in itself does not change, let alone reverse, the selective pressures acting on altruism.

## The genetic and cultural evolution of altruism

### Two theories of cultural altruism

The idea that culture could yield the evolution of altruism among nongenetic kin has been primarily put forward in the theory of cultural group selection (CGS) originally developed by Boyd & Richerson (1982, 1985, 2009a; Richerson & Boyd, 2005; Boyd et al., 2011; see also Allison, 1992a,b for similar ideas, Henrich & Boyd, 2001; Fehr & Fischbacher, 2003; Gintis, 2003; Gintis et al., 2003; Henrich, 2004; Henrich & Henrich, 2007, for further applications of CGS; and Lehmann et al., 2007, 2008a for a population genetics formalization). Boyd and Richerson's idea is that social behaviours are culturally transmitted in a way that tends to homogenize the phenotype of individuals within each group. They consider, in particular, one type of 'imitation rule' that could have this consequence: conformist-biased imitation, whereby individuals imitate whatever behaviour is the most frequent in their local group. Conformist imitation generates a strong phenotypic homogeneity within social groups. Some form of group selection can then act to favour the groups in which the shared cultural trait

happens to be group-beneficial (i.e. altruistic) rather than individual-beneficial (i.e. selfish). More recently, Lehmann et al. (2008a) has shown that the same process occurs also under another family of imitation rules called 'one to many' transmission, in which one, or a few, individual(s) are chosen randomly in each group to serve as cultural model(s).

Cultural group selection, however, could be said to come in two distinct versions. Some models predict the cultural evolution of *altruistic* behaviours in a strict sense (e.g. Boyd & Richerson, 1985, ch. 7; Henrich & Boyd, 2001; Gintis, 2003; Lehmann et al., 2008a). Yet, other models, also pertaining to CGS, aim at explaining the cultural evolution of *cooperative* behaviours that are not altruistic, that is, behaviours that increase both the recipient's and the actor's fitness (e.g. Boyd & Richerson, 1990, 2002, 2009b). The points made in this paper are meant to apply to models of the first category, not to the second. The distinction between these two kinds of CGS (the one based on biological altruism and the one based on mutually beneficial cooperation) is easily overlooked, and some frequent confusions between the two will have to be lifted. This shall be done at the end of this paper.

Even though CGS is the only theory explicitly developed to explain the existence of cultural altruism towards nongenetic kin, its core mechanism operates in another important family of models: the study of social evolution in networks, also sometimes called 'evolution on graphs' (Skyrms, 2004; Nowak, 2006a,b; Ohtsuki et al., 2006; Taylor et al., 2007; Ohtsuki & Nowak, 2008; Santos et al., 2008; Tarnita et al., 2009, see also Allison, 1992a for a verbal description of the same process). In these models, individuals interact within networks, and neighbours on the network mutually imitate one another. In network models, individuals typically follow a family of rules called payoff-biased imitation. They observe the action and resulting payoff of their neighbours and preferentially imitate the action played by high-payoff individuals (payoff-biased imitation is also sometimes used in CGS – Henrich & Boyd, 2001; Boyd et al., 2003). Payoff-biased transmission is a cultural equivalent of biological reproduction, in which payoff affects cultural influence rather than biological fitness (see Hofbauer & Sigmund, 1998). Network models are hence explicitly built to represent indifferently either genetic or cultural evolution. In either case, they show that altruism can evolve among neighbours because neighbours do share common genes, or common cultural items. In their cultural version, therefore, these models open the possibility that altruism evolves among genetically unrelated individuals.

### Genetic and cultural evolution

The mechanisms leading to the cultural evolution of altruism in all models may be simply captured in the following way. Let us consider a social trait (helping,

punishment, or anything else), transmitted genetically. From the most basic principles of social evolution theory (detailed in Supporting Information), the effect of selection on that trait can be measured by a simple and general equation, Hamilton's rule (Hamilton, 1963, 1964), which states that the trait is favoured by selection when $-C_g + R_g B_g > 0$. In this equation, $-C_g$ is a partial regression coefficient measuring the statistical relationship between the trait value of a focal 'actor' individual (e.g. altruism or selfishness) and the *biological* fitness of this individual and is called the 'direct' effect of the trait on fitness. $B_g$ is a regression measuring the statistical relationship between the trait value of the actor and the biological fitness of another individual (called the 'recipient' in the case of a helping behaviour) and is called the 'indirect' effect of the trait on fitness. Finally, the third coefficient, $R_g$, is the *genetic* relatedness between the actor and the recipient, measured as the regression, on the focal trait, of the heritable component of the recipient's phenotype on the heritable component of the actor's phenotype.

The same principles apply also for a social trait transmitted culturally, that is, encoded by a cultural item (see Supporting Information). The trait is favoured by selection when $-C_c + R_c B_c > 0$, except that, now, $-C_c$ measures the statistical relationship between the actor's trait value and the cultural reproductive success of his cultural item (the specific item encoding the trait), $B_c$ measures the statistical relationship between the actor's trait value and the reproductive success of the homologous cultural item carried by a recipient individual and $R_c$ is the cultural relatedness between the actor and the recipient, measured as the regression of the culturally transmissible component of the recipient's phenotype on the culturally transmissible component of the actor's phenotype.

As we see, being a fully general partitioning of selective forces, the fundamental structure of Hamilton's rule is unchanged by the transmission mode of a trait (genetic or cultural). However, the *parameters* of Hamilton's rule are affected. First, the relatedness, on a given trait, between an actor and a recipient, shall depend on the way in which that trait is transmitted. As long as cultural transmission is not purely vertical, the relatedness it yields probably differs from that yielded by genetic transmission (i.e. $R_g \neq R_c$). Second, when cultural transmission is not purely vertical, the statistical relationship between the phenotypic expression of a cultural item (e.g. altruism or selfishness) and its transmission success is likely to differ from the strict biological effect of this item. When a given phenotypic trait is transmitted culturally, the parameters that matter to determine the effect of natural selection on that trait are not its effects on *biological fitness*, but its (direct and indirect) effects on the *cultural fitness* of the underlying cultural item, and these are likely to be different (i.e. $(B_g, C_g) \neq (B_c, C_c)$).

Simply stated, all this merely expresses the fact that, when two levels of replicators (genes and culture) coexist, there is no *a priori* reason why their evolutionary interests should always be aligned. Hence, the direction of selection, measured in Hamilton's rules, is not the same on both levels. A trait that would be maladaptive on the genetic level can be favoured by selection when it is transmitted on the cultural level, and vice versa. This general statement is almost trivial, but its implications are not, in particular in the case of altruism. Being an autonomous evolving system, culture could, at least in principle, yield us to sacrifice ourselves, or at least help perfect strangers, at a net cost to biological inclusive fitness.

## Biological altruism

Here, we are interested in *biologically* altruistic traits, defined as traits with a negative effect on the biological fitness of their carrier ($-C_g < 0$) and a positive effect on the biological fitness of other(s) ($B_g > 0$). These traits are interesting because, under regular genetic evolution, they should only evolve when they benefit genetically related individuals.

As we just explained, however, when a trait is transmitted culturally, the relevant $B$ and $C$ parameters are not its effects on biological fitness. Hence, a biologically altruistic trait is not necessarily culturally altruistic. For instance, a costly helping behaviour that would be easily imitated (e.g. because it is psychologically attractive) would be biologically but not culturally altruistic. In this paper, however, we are not interested in cultural selfishness/altruism defined in this sense. The aim of cultural models, and therefore the aim of the present review, is to understand how actual behaviours that have the paradoxical property to reduce the biological inclusive fitness of their carriers can evolve, that is, to understand how different transmission modes may or may not change the selective pressure upon *biological* altruism. Accordingly, in the following, when we speak of 'altruism', we always mean 'biological altruism'.

Models found in the literature typically consider the evolution of two types of biologically altruistic traits. Some consider exclusively the evolution of *helping* behaviours, whereby a helper provides a good or a service to a recipient (Boyd & Richerson, 1985, ch. 7; Gintis, 2003; Lehmann & Feldman, 2008; Lehmann *et al.*, 2008a). Others consider the joint evolution of helping and *punishing* behaviours, whereby a punisher pays a cost to reduce the fitness of other individuals, because they have refused to provide help in a first stage of the game (Henrich & Boyd, 2001; Boyd *et al.*, 2003; Guzman *et al.*, 2007; Lehmann *et al.*, 2007). The only important thing to understand, here, is that the characterization of a trait as altruistic (or not) is independent of whether the trait consists in helping or punishing (Gardner & West, 2004; Lehmann *et al.*, 2007). A trait is altruistic if it increases

the direct biological fitness of other(s) at a net cost to one's own. Helping is altruistic when it comes at a net personal cost for the helper. Punishment can be altruistic also, if it comes at a net personal cost for the punisher, and if it benefits nonpunished individuals by reducing the intensity of local competition, or by increasing the average amount of help expressed by others.

## The hypothetical cultural facilitation of biological altruism

The effect of selection on any trait depends on its transmission mode (see above, section 'Genetic and Cultural Evolution'). In itself, this is sufficient to entail that altruism may, at least in principle, evolve culturally when its genetic evolution is impossible, and vice versa. However, models of the cultural evolution of altruism make a more specific statement (at least in the case of CGS). Not only do they claim that cultural and genetic evolution differ, they also claim that, under realistic conditions, cultural evolution is generally *more* likely than genetic evolution to yield altruism. In the following we review what we take to be the two major mechanisms, at work separately or in conjunction, that lend credibility to this claim: (i) culture can 'increase' relatedness (i.e. $R_c > R_g$) and (ii) culture can 'reduce' the direct cost of altruism (i.e. $C_c < C_g$).

### *Cultural relatedness may be larger than genetic relatedness*
The (genetic or cultural) relatedness in a given group, or neighbourhood, is the outcome of a balance between two antagonistic forces: (i) sampling effects that occur each time reproduction takes place and tend to increase local homogeneity and (ii) migration and mutation that reintroduce outside polymorphism and tend to decrease homogeneity. One of the effect of cultural transmission, in models, is to boost the importance of sampling effects relative to migration. This occurs in two distinct ways.

*Cultural reproduction is faster.* It is empirically reasonable to assume that more reproduction events take place for culture than for genes in one given amount of time. 'Cultural generations' are shorter. If the cultural migration rate is not proportionately larger, then the migration rate *per generation* is mechanically lower in culture, and hence cultural relatedness is larger. Interestingly, this occurs even if cultural transmission is subject to the same (or similar) selective forces than genetic transmission (i.e. under 'payoff-biased' transmission; e.g. Boyd *et al.*, 2003; Ohtsuki *et al.*, 2006).

For instance, Boyd *et al.* (2003) consider a migration rate between groups where <1% of individuals 'migrate' every generation, which yields a strong relatedness. What allows this is, of course, the fact that the generations considered in the model are cultural generation. Thus, generations last only a year (whereas biological genera-

tions might last approximately 25 years). The same holds in network models. Ohtsuki *et al.* (2006) show that helping evolves if the relative benefit of help ($b/c$) is larger than the number ($k$) of social partners per individual. This can reasonably be achieved if 'individuals' are actually cultural items that reproduce every year (or more) and can interact with as few as approximately 5 other items every cultural generation, not if they are biological organisms reproducing every 25 years and interacting with several hundreds of partners in their lives. The migration rates that happen to favour altruism in group-structured populations, the social graphs that happen to favour altruism in network models, are empirically more reasonable under cultural than genetic evolution.

*Cultural transmission is more homogenizing.* The balance between homogenizing and diversifying forces is also tipped by culture for a second reason, related to the specific forms that cultural transmission takes in many models: conformist-biased transmission (Boyd & Richerson, 1985, ch. 7; Henrich & Boyd, 2001; Guzman *et al.*, 2007; Lehmann & Feldman, 2008; Lehmann *et al.*, 2008a) and one to many transmission (Lehmann *et al.*, 2007, 2008a). With these modes of transmission, not only is cultural reproduction faster than biological reproduction, it is also more homogenizing (a single or a few cultural items are imitated by every individual). This reinforces even further the amount of cultural relatedness relative to genetic relatedness.

### *The cultural costs and benefits of biological altruism may differ from their biological counterparts*
When we think of the effect of culture on biological altruism, we usually exclusively think of the increased relatedness. However, when comparing the effect of selection under genetic vs. cultural transmission, the other parameters of Hamilton's rule must also be compared, namely the cost, $C$, and benefit, $B$, of altruism. Let us insist, however, to avoid any confusion. Here, we consider culturally transmitted traits that happen to have the property to be *biologically* altruistic (i.e. $-C_g < 0$ and $B_g > 0$), and we say that their *cultural* costs and benefits ($C_c$ and $B_c$) are likely to differ from their biological costs and benefits ($C_g$ and $B_g$). By no means do we intend to say that cultural transmission has an effect on the biological effects of traits.

To our knowledge, in fact, the only paper in which this second outcome of culture has been explicitly highlighted is a model in which cultural transmission impedes rather than facilitates the spread of altruism. Under a form of payoff-biased transmission, Lehmann *et al.* (2008a; see also Lehmann *et al.*, 2007) show that the direct cost of altruism is larger under cultural than genetic transmission (i.e. $C_c > C_g$, compare their equations 16 and 22) because the effect of local competition is stronger. In this case, *ceteris paribus*, altruism is less likely under cultural than genetic transmission.

In other models, the cultural benefit-to-cost ratio of altruistic traits ($B_c/C_c$) seems more favourable to their success than their biological counterpart (e.g. Boyd & Richerson, 1985, ch. 7; Henrich & Boyd, 2001; Gintis, 2003). However, this is not explicitly put forward in the papers, and it would require more formalization to be better understood. Essentially, we think this change in benefit-to-cost ratio arises because (i) altruistic behaviours are assumed to be transmitted just as well (or almost as well) as selfish behaviours within local groups, that is, the individual biological cost of altruism, does not translate into an equivalent cultural cost in local transmission, whereas (ii) the collective benefit of altruism does translate into a cultural advantage that accrues exactly equally to all the individuals in each group (e.g. via a larger group survival, see e.g. Boyd & Richerson, 1985, ch. 7; Gintis, 2003). In other words, and to put it bluntly, in these models, the individual cost of altruism is reduced but its collective benefit is maintained. This probably accounts for a significant part of these models' outcome.

## When biological altruism evolves culturally, it is always a maladaptive byproduct of imitation

One might capture the basic argument of the models we review as stating that culture, being an autonomous evolving system, may in principle yield altruism when genes cannot. They suggest in particular that, under reasonable assumptions, altruism might evolve culturally more easily than genetically. However, the apparent simplicity with which they achieve this effect, by changing the parameters of Hamilton's rule, is misleading. It might hide the fact that these models rely on a crucial empirical assumption.

Genuine altruism towards nonkin (or, more generally, genetically maladaptive altruism) can evolve and/or stabilize culturally if individuals use imitation rules, such as conformist, one to many or payoff-biased imitation, that result in its cultural diffusion. Yet, a maladaptive behaviour, even imitated from someone else, is still a maladaptive behaviour. When an individual behaves in an altruistic manner towards nonkin, whether because she is 'directly' mistaken, or because she has mistakenly imitated another altruist, she is simply expressing a genetically maladaptive trait. Thus, cultural evolution solves the problem of altruism only to convert it into another problem: the fact that individuals learn from others in a way that eventually leads them to acquire maladaptive behaviours. Let us consider three examples before reaching a general conclusion.

### Conformist imitation

Conformist imitation consists in imitating the behaviour expressed by the majority of individuals in one's group.

According to CGS theory, it is an adaptive strategy because behaviours expressed by the majority tend to be more adaptive than average (Boyd & Richerson, 1985, ch. 7). Yet, CGS scholars also recognize that frequency is a highly indirect, limited and easily mistaken proxy. Many things can cause a majority of individuals to express maladaptive behaviours (e.g. founder effects and cultural drift), and these behaviours are then imitated all the same, which is maladaptive. In other words, with respect to biological fitness, conformist imitation is mistake-prone, and this mistake-proneness is the very reason why it can lead to the acquisition of maladaptive behaviours such as altruism.

Two theoretical results seem surprisingly at odds with the above reasoning, claiming to show that conformist imitation can be favoured by genetic evolution, even when its *only* effect is to trigger the imitation of altruism (Henrich & Boyd, 2001; Guzman *et al.*, 2007). Both models consider punishment rather than primary helping as the focal altruistic trait. In a model in which conformist imitation allows the cultural spread of altruistic punishment, Henrich & Boyd (2001) show that natural selection favours genes that predispose individuals to imitate more faithfully the social behaviour of others, and Guzman *et al.* (2007) show that natural selection favours genes that directly code for conformist imitation itself. In a situation where genetic evolution should not lead to altruistic punishment, evolution would favour a genetic predisposition (Henrich & Boyd, 2001) or a genetic learning rule (Guzman *et al.*, 2007) leading individuals to acquire altruistic punishment. How is that possible?

The explanation is that both results are permitted by an implicit assumption. In Guzman *et al.* (2007), individuals are assumed to be either conformist with respect to *both* helping and punishment or not conformist at all. Yet, what is adaptive in this model is to faithfully imitate the *helping* part of others' behaviour (because when others cooperate they generally also punish, and therefore one should cooperate as well). But imitating the punishment part of their behaviour is not adaptive (because punishment is simply *always* costly). However, being maladaptive does not keep this norm from being adopted, because no flexible imitation strategy exists, that would allow agents to copy the beneficial norm but not the costly one.

In Henrich & Boyd (2001), punishing is not always detrimental to punishers, because several layers of punishment are authorized: agents who do not punish may be punished for this, and so on up to the $n$th level. In this case, all intermediate levels of punishment are in the direct interest of individuals', and only the $n$th level is genuinely costly because it faces no upper-level sanction. Why is punishment at this last degree nevertheless adopted by evolved imitation rules? The answer is that in Henrich & Boyd (2001), just as in Guzman *et al.* (2007), the same genetic predisposition is assumed to affect *equally* the imitation of all social traits – whereas

the more adaptive strategy would consist in imitating only those traits that benefit the individual. The imitation of the last layer of punishment is hence a costly side effect of the (adaptive) imitation of other layers.

Overall, even though these two models are presented as exceptional instances in which conformist imitation evolves even when its only effect is to trigger the imitation of altruism, they have nothing exceptional in this respect. As in all other models, altruism is a costly side effect of a relative maladaptation of imitation rules.

Note that Lehmann & Feldman (2008) also surprisingly found that unbiased conformist imitation (in which each individual imitates a random member of her group), as well as one to many transmission (in which every group member imitates the same peer), can be favoured by genetic evolution in a model in which their only effect is to favour the cultural spread of helping (towards nonkin). But a closer look offers an interpretation of this result. When random imitation occurs (unbiased conformism, or one to many transmission), individuals can *directly* benefit from being helpers, because this may influence the phenotype of others in their group and prompt them to help in return through imitation. This effect obtains when each individual has a large expected influence on others, which is exactly what Lehmann & Feldman (2008) find (this occurs when social groups are small; see also Boyd & Richerson, 1988). Under such circumstances, owing to a type of selective pressure also at work in the evolution of reciprocity, where the expression of helping by one individual increases the tendency of others to express it, imitation can be favoured because of its effect on helping. However, by definition, helping is not altruistic in this case, being directly reciprocated.

### Payoff-biased imitation

At first, payoff-biased imitation is even more perplexing than conformist transmission. Payoff-biased imitation looks like a perfect manner to take behavioural decisions. As long as the observed 'payoff' is a faithful measure of biological fitness, payoff-biased imitation should be immune to the mistaken acquisition of genetically maladaptive behaviours. This intuition explains why, in contrast with conformism, scholars who use payoff-biased imitation do not (to our knowledge) raise the question of its evolutionary rationale. However, it turns out that this intuition can be erroneous. With payoff-biased imitation, one does risk to acquire maladaptive behaviours, but only a specific category of maladaptive behaviours: altruistic ones. Altruism is the flaw, so to speak, of payoff-biased imitation.

This flaw manifests itself when interacting individuals are phenotypically correlated because an individual's payoff then partly depends on her neighbours' strategy. Consequently, their payoff is a poor indicator of the

adaptiveness of their behaviour. In a social network or in a group-structured population, altruists have a larger payoff than 'selfish' individuals, but that is because their *neighbours* are more altruistic than average, not because *they* are altruistic (see Allison, 1992a). This is precisely the difference between a phenotypic correlation emerging from reciprocity and a correlation emerging from co-ancestry. In the latter, the correlation is not the reflection of an underlying causal relationship between the actor's phenotype and the recipient's. Helpers do have a larger payoff because they happen to interact more often with helpers, but helping does not *cause* an increase in one's payoff. Imitating high-payoff individuals thus leads to the mistaken acquisition of an altruistic behaviour that did not cause their large payoffs. Altruism, there again, is a consequence of the learning rule's proneness to mistakes.

### Social learning constrains behavioural adaptation

Being a genetically maladaptive trait, altruism towards nonkin entails a lack of adaptation on the side of genes, whether because individuals mistakenly act altruistically, or because they mistakenly imitate other altruists. Hence, cultural transmission does not change the nature of the selective pressures, acting upon genes, with respect to altruism. What cultural transmission does is only to introduce an intermediate control variable (the rules of social learning) in the evolution of social traits, which changes the constraints applying to behaviours, and limit their potential for adaptation in different ways (see Lehmann *et al.*, 2008b). This is interesting and important, but does not by itself cancel the paradox of altruism.

### Strong constraints generate strong maladaptations

What is more, in their models, students of cultural altruism assume particularly strong constraints over social learning. They assume that individuals learn from others using *stereotyped* rules. Individuals copy the most frequent behaviour of their group, or the behaviour of their highest payoff-peer, but they are unable to evaluate the actual properties of the behaviour (e.g. 'Does it involve to spend some resources?' and 'Does it entail a physical risk?'). The ability of these simple rules to subtly control cultural acquisitions is hence limited and opens the possibility of many maladaptive decisions, including the imitation of altruism. Assuming only simple imitation rules is like imposing a very strong limit over the genetic adaptation of behaviour.

The assumption that evolution has not been able to provide human beings with better mechanisms than stereotyped rules of imitation, in important domains such as social behaviour, is quite a surprising one. For instance, the rule of payoff-biased imitation used in models of evolution on graphs rests on the assumption that humans are able to track others' fitness gains in complex circum-

stances, but unable to understand the simple fact that unreciprocated help is costly, while being helped is beneficial (see the section 'Payoff-Biased Imitation').

This raises the question: Why do scholars follow assumptions of this sort? The answer, we think, is that it is part of the implicit way of looking at things in cultural evolution. Under the influence of dual-inheritance theory (Cavalli-Sforza & Feldman, 1981), scholars see culture as a supplementary system of inheritance, parallel to genes. Hence, like genes, cultural items are assumed to be transmitted according to simple stereotyped rules (e.g. biparental inheritance for nuclear genes). This leaves a lot of autonomy to culture and thus yields a lot of maladaptations.

### Weaker constrains would generate weaker maladaptations

Yet, the ability to modulate one's social behaviour in function of others' could also be seen as a complex form of adaptive plasticity, rather than a supplementary system of inheritance. Instead of a general drive to imitate others, we could be endowed with a domain-specific ability to construct our social behaviour, in line with the contingent strategies found in reciprocity (Trivers, 1971). Even though they could never be perfect and adaptive in all circumstances, such abilities would leave much less space for maladaptive altruism.

This debate is an old matter in the dual-inheritance literature and dates back to the foundation of the field (Cavalli-Sforza & Feldman, 1981; Lumsden & Wilson, 1981). Defenders of simple imitation heuristics argue that we are bound to use stereotyped transmission strategies for all purposes, because cultural behaviours and innovations can take an infinite variety of aspects. No genetically encoded mechanism could have prepared us to deal with all of them (Richerson & Boyd, 2005, ch. 5; Boyd & Richerson, 2006). We do not aim to enter, or even less to settle, this debate here. We simply note that Richerson and Boyd's general argument is certainly well applicable to decisions that we, as a species, have seldom (or never) been confronted with (e.g. the making of a kayak or a bow). But it is unclear to us how this same argument applies in a domain, such as social life, to which our ancestors have been confronted for long. The possible ways of helping others, although they are infinitely numerous, do share some general properties, and evolution is likely to have endowed us with innate abilities that help us deal specifically with these situations (as some recent experiments in preschool children seem to suggest, see e.g. Hamlin *et al.*, 2007; Warneken *et al.*, 2010; Hamann *et al.*, 2011).

### Weaker maladaptations are only weakly altruistic

Boyd, Richerson and Henrich do often argue in favour of the existence of stereotyped imitation in cultural evolu-tion (e.g. Richerson & Boyd, 2005, ch. 5; Boyd & Richerson, 2006). Yet, they also often seem well aware of the fact that, as products of evolution, humans are unlikely to be so crudely adapted that they are ready to imitate anything. Individuals shall use simple imitation rules, they often agree, only when they have no better sources of information. For instance, Richerson and Boyd state that conformist transmission can operate only if 'individuals have difficulty evaluating the costs and benefits of alternative cultural variants' (Richerson & Boyd, 2005, p. 206; see also Henrich & Henrich, 2007, p. 66).

For this reason, they have put forward the idea of 'altruistic punishment' (Boyd & Richerson, 1992; Henrich & Boyd, 2001; Fehr & Gachter, 2002), according to which people primarily cooperate because they fear punishment (i.e. the helping actions per se are not altruistic but directly beneficial to individuals), but the norms of punishment are themselves grounded in the fact that people are ready to punish even at one's personal cost (i.e. the *second-order* retributive actions are genuinely altruistic). This is interesting because, punishment being less frequently expressed (in equilibrium, most individuals cooperate and are never punished), it is biologically 'cheaper' in average than helping. Therefore, so the argument goes, individuals shall have more difficulty detecting the cost of punishment than the cost of helping and are thus more likely to use content-blind imitation in the former (Henrich & Boyd, 2001; Henrich & Henrich, 2007, pp. 66–67).

Leaving aside the fact that it is not clear (at least to us) how the mere rarity of punishment makes it *more* likely to be imitated, it is important to understand that the reservations of these scholars on the amount of mistaken imitations in humans, however interesting and reasonable they may be, are also, de facto, reservations on the amount of genuine altruism present in societies. If Boyd, Richerson, Henrich and others are ready to consider that humans imitate altruism only in rare instances in which they have difficulty evaluating its costs (e.g. only for punishment), they must also agree that genuine altruism is rare.

## Ambiguities in the literature

### Gene–culture coevolution

Many statements found in the literature do honestly present altruism as a costly side effect of social learning, which entails that genes that make the cultural acquisition of altruism more likely cannot be favoured by selection *because* of their effect on altruism, but at best *in spite* of this effect (see e.g. Boyd & Richerson, 1985, p. 227; Gintis, 2003).

Other statements, however, revolving around the idea of gene–culture coevolution, are less clear. For instance, in a widely cited review, Gintis *et al.* (2003) define strong

reciprocity as 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', which is genetically maladaptive by definition, and yet claim that 'strong reciprocity is *adaptive* in the sense of emerging from a gene–culture coevolutionary process' (the emphasis is ours). Richerson and Boyd provide another example in a recent book, when they state that gene–culture coevolution 'set up an arms race that drove social evolution to ever greater extremes of in-group cooperation' such that eventually 'people were ready to punish [...] even when personal interests were not directly at stake' (Richerson & Boyd, 2005, p. 214). Even more surprisingly, Henrich & Boyd (2001) write in their abstract that 'prosocial genes favouring cooperation and punishment may invade', and Guzman *et al.* (2007) write in their conclusion that conformism coevolves genetically with cultural altruism with 'no need to assume that costly conformism is a spin-off from individually beneficial conformism'.

These are misleading, if not simply false, statements. They certainly cannot help readers to comprehend the fact that altruism can be nothing but a costly side effect of imitation and can even help spreading the false idea that, thanks to culture, altruism towards nonkin ends up being adaptive, even from a gene's point of view. There are at least three sources of ambiguity that play a role here. We rapidly mention two of them. We then devote a specific subsection to the third and most important one.

First, the idea that the two sides of an interaction coevolve can mean quite different things. Coevolution can be antagonistic when the two sides have conflicting interests. Or it can be 'reinforcing' when the two sides have aligned interests. Regarding biological altruism with nonkin, by definition, genes and culture have conflicting interests. Hence, altruism can be the outcome of a coevolution between genes and culture only in the former sense, not in the latter. The claim that gene–culture coevolution can lead to altruism is thus literally true, but it is misleading. It gives the false impression that altruism is even stronger when genes come into play, whereas it is the exact opposite that is true. The correct statement is that *cultural* evolution can lead to altruism, and that *genetic* counter-evolution is not always fully able to prevent it. This is not what readers are likely to understand when they read that 'strong reciprocity is adaptive in the sense of emerging from a gene–culture coevolutionary process'.

Second, as we already discussed, Henrich & Boyd (2001) and Guzman *et al.* (2007) claim to show that genes coding for imitation can be favoured even when their 'only' effect is to yield altruism, but they forget to mention that this results from a constraint over the adaptation of learning (see section 'Conformist Imitation'). The imitation of costly altruism is, in these models as in others, a costly side effect of the imitation of beneficial traits. Therefore, the statement that 'there is no need to assume that costly conformism is a spin-off from individually beneficial conformism' is simply wrong.

## Cultural group selection: disentangling altruism and cooperation

Third, as we already acknowledged, CGS comes in two significantly different versions, and this leads to some ambiguities. The first version is the target of this paper. Here, we call it the 'altruistic' version because it aims to explain how culture may yield the evolution of altruistic behaviours *stricto sensu*. In contrast, the second version, called here 'mutualistic', aims to explain the cultural evolution of cooperative behaviours that are not altruistic (see West *et al.*, 2007b), that is, behaviours that increase both the recipient's and the actor's fitness through incentives such as reciprocity, reputation or retribution, performing what is known in game theory as 'equilibrium selection' (e.g. Boyd & Richerson, 1990, 2002, 2009b). Although these two versions have initially been explicitly distinguished (e.g. in Boyd & Richerson, 1990), this is no more the case in recent writings that neglect to distinguish genuine altruism from mutualistic cooperation (e.g. Richerson & Boyd, 2005, ch. 6; Boyd & Richerson, 2009a).

For instance, Richerson & Boyd (2005) begin their chapter six by explaining why genetic group selection is limited to small groups of close kin, whereas CGS is possible in larger groups (pp. 201–204). Even though they never explicitly define it, this strongly suggests that they deal with genuine *altruism* (they even mention the case of social insects). But then they go on to explain that human 'tribal social instincts' evolved *genetically* because people lacking these instincts were punished and ostracized (p. 214), which corresponds to the *mutualistic* version of their theory where cooperation is individually beneficial. Finally, they implicitly move back to altruism when they ask why genes that prevent the acquisition of costly cooperative behaviours do not invade (p. 214). In a more recent review, Boyd & Richerson (2009a) seem to be mostly dealing with the mutualistic version of their theory, mentioning the problem of equilibrium selection (p. 3281), and stating that genetic instincts have 'coevolved with culturally transmitted social norms' (p. 3287). However, they still choose to define cooperation as a 'costly behaviour performed by one individual that increases the payoff of others' (p. 3283), which is really *altruism* and therefore falls under the scope of the other version of CGS.

We fear that such expositions of CGS, which fail to distinguish mutualism and altruism, could lead to important misunderstandings of the assumptions and explanatory power of the theory. The altruistic version of CGS can explain the existence of genuinely altruistic

behaviours (e.g. strong reciprocity), but it relies on the existence of a *mismatch* between genes and culture (the mistaken imitation of costly cultural traits), not on a reinforcing coevolution of the two. On the contrary, the mutualistic version of CGS can only explain the existence of cooperative, not altruistic, behaviours, but it can be the outcome of adaptive learning mechanisms and thus emerge from a reinforcing coevolution of genes and culture. Confusions between the two versions are hence dangerous, because one can take the mild restrictions of one version to be the sufficient conditions for the other. For instance, and most symptomatically, in their widely cited review on human altruism (Fehr & Fischbacher, 2003), have a specific section on gene–culture coevolution, where they state that it can 'provide a solution to the puzzle of strong reciprocity'. Nowhere in this section is the underlying assumption of genetic maladaptation mentioned.

### How to disambiguate cultural group selection theory

To avoid any risk of misinterpretation, it would be useful if, for each version of their theory (or mixture of versions), proponents of CGS could spell out *together* (i) its underlying assumptions (e.g. regarding the degree of mismatch between genes and culture) and (ii) its explanatory power (e.g. regarding the amount of genuine altruism in human behaviour). In view of most of their recent articles (in particular Henrich & Henrich, 2007, ch. 3; Boyd & Richerson, 2009a,b; Boyd et al., 2011), it seems to us (but we might be wrong) that Boyd, Richerson and Henrich now endorse a primarily mutualistic version of their theory, in which helping is mostly individually beneficial, and genetic instincts can coevolve with, and reinforce, cultural helping (note that this was not the case in earlier writings, e.g. Boyd & Richerson, 1985, p. 227). However, if they do currently favour such a view, CGS scholars must also explicitly state that their theory does not support many claims, found in the literature, about the alleged selflessness of humans. For instance, a moderate version of their theory cannot support the claim that human cooperation is a 'huge anomaly' (Fehr & Fischbacher, 2003), or that it is 'fundamentally incompatible with the biologists' model of the self-regarding reciprocal altruist' (Gintis et al., 2003; see West et al., 2011, for more examples). Initially, the originality of CGS lied in its ability to explain cooperation in anonymous contexts (in which it cannot have direct benefits), and it has been invoked in many claims on this issue (e.g. Fehr & Fischbacher, 2003; Gintis et al., 2003). Yet, a primarily mutualistic version of CGS loses this specific explanatory power. At best, with such a view, one can explain that humans sometimes make mistakes in their social decisions (e.g. by cooperating when it turns out to have no benefit), not that they generally and systematically differ from nonhumans by being ready to help and punish with no inclusive fitness benefits.

### Conclusion

We do sometimes behave in a strictly altruistic manner towards nonkin. For instance, economic experiments have shown that, even when perfect anonymity is guaranteed, some experimental subjects still give away monetary resources to help or punish others, with no clear individual benefit (although anonymity does strongly reduce subjects' generosity, see Hoffman et al., 1996). Our aim here has not been to discuss this literature. We discussed one specific kind of evolutionary explanation of altruism: cultural explanations.

Any explanation for a maladaptive trait, whether it involves culture or not, must rely on the fact that natural selection has been unable to optimize it. Cultural altruism is sometimes thought of as an alternative to maladaptation, but this is a misunderstanding. Like any evolutionary account of a maladaptation, cultural evolution models do rely on the inability of natural selection to remove the maladaptation. The only difference between cultural accounts of altruism, and others, has to do with the kind of maladaptation involved (Fig. 1): Do they come from the limitations of our social cognition, or from those of our capacities for cultural learning?

### Imperfections of social cognition

Genuine altruism could emerge as a consequence of imperfections of human social cognition. We are endowed with evolved predispositions to handle social behaviour. They help us manage our reciprocal exchanges and reputation. These predispositions should be adapted, in average, to maximize our genetic fitness, but we cannot expect them to be perfect (especially not in experimental settings).

Virtually every aspect of our phenotype is the product of compromises: compromises between the costs and benefits of complexity; compromises between one function and another; etc. For instance, a 'snake-detecting' ability cannot both detect all snakes with near-certainty and never generate false positives (this has been termed the 'smoke detector principle' by Nesse, 2001). Relative maladaptations are unavoidable in snake detection. Similarly, the set of cognitive modules that play a role in our social decisions (to help or not help, punish or not punish, etc.) must unavoidably be subject to constraints. Therefore, relative maladaptations are unavoidable in this domain as well (see West et al., 2011 section 6.6.1). As an illustration, in the vein of the smoke detector principle, one may imagine that it could generally be adaptive to overestimate the importance of being cooperative in social situations, for the sake of one's reputation. Yet, this would come at the cost of false positives: we would sometimes behave in a cooperative manner when it turns out that there was no benefit to gain (see e.g. Hagen & Hammerstein, 2006).
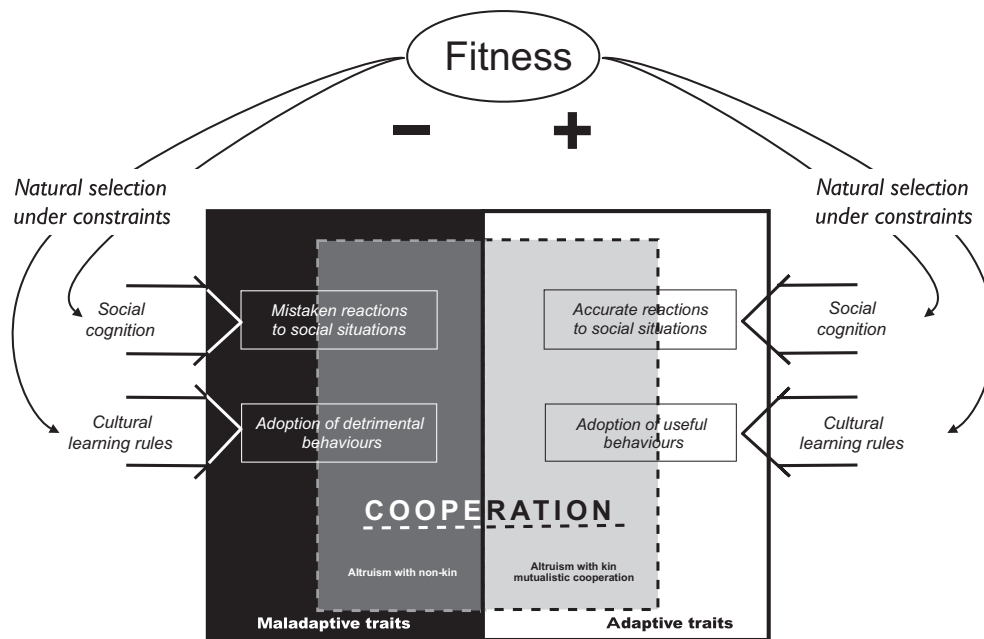
**Fig. 1** Culture does not change the selective pressure on genes, which always tends to maximize inclusive fitness. Relative maladaptations can nevertheless occur, because the action of natural selection is constrained. In the case of cooperative behaviours (that benefit at least one recipient), relative maladaptations may have at least two origins. Social cognition may be imperfect, leading (among other maladaptive behaviours) to altruism with no inclusive fitness benefits. Likewise, cultural learning rules may be imperfect, leading to the adoption of maladaptive behaviours, some of which may happen to be altruistic. Influential versions of cultural group selection consider that biological altruism is particularly important in humans because of this second mechanism.

## Imperfections of cultural cognition

The originality of the cultural accounts of altruism is that, rather than relying on the necessary imperfection of the set of evolved mechanisms involved specifically in social decisions, they rely on imperfections of another, more transversal, set of evolved mechanisms: the mechanisms we use to learn from others. Because we generally imitate others, so the argument goes, we often end up imitating also their maladaptive courses of actions. Scholars also explain this relative imperfection by the existence of a compromise. In the domain of social learning, they argue, we face a trade-off between the risk to imitate maladaptive behaviours, and the benefit of imitating interesting novelties. Cultural altruism is thus one of the costly consequences of this trade-off, like our fear of snake-shaped wood sticks is a consequence of the trade-offs faced by our snake-detecting ability.

This paper is not the place to discuss in detail the relative merits and explanatory power of these two lines of research. We simply wish to highlight their profound similarity. In both cases, constraints limit the adaptive perfection of our decisions and introduce unavoidable mistakes in our behaviours. In one case, the mistakes are specifically related to social life. In the other, the mistakes concern equally all the behaviours we can learn from others. Usually, because it has to do with culture, the second explanation is seen as profoundly different. It is not, and we think that the respective merit of cultural and social explanations of genuine altruism are easier to compare when seen as similar in this respect.

## Acknowledgments

## References

Allison, P. 1992a. The cultural evolution of beneficent norms. *Soc. Forces* **71**: 279–301.

Allison, P. 1992b. Cultural relatedness under oblique and horizontal transmission. *Ethol. Sociobiol.* **13**: 153–169.

Boyd, R. & Richerson, P.J. 1982. Cultural transmission and the evolution of cooperative behavior. *Hum. Ecol.* **10**: 325–351.

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

Boyd, R. & Richerson, P.J. 1988. The evolution of reciprocity in sizable groups. *J. Theor. Biol.* **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. *Ethol. Sociobiol.* **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. 2006. Culture, adaptation, and innateness. In: *The Innate Mind: Culture and Cognition* (P. Carruthers, S. Stich & S. Laurence, eds), pp. 23–28. Oxford University Press, New York.

Boyd, R. & Richerson, P.J. 2009a. Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **364**: 3281–3288.

Boyd, R. & Richerson, P.J. 2009b. Voting with your feet: Payoff biased migration and the evolution of group beneficial behavior. *J. Theor. Biol.* **257**: 331–339.

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

Boyd, R., Richerson, P.J., & Henrich, J. 2011. Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behav. Ecol. Sociobiol.* **65**: 431–444.

Cavalli-Sforza, L. & Feldman, M. 1981. *Cultural Transmission and Evolution*. Princeton University Press, Princeton.

Fehr, E. & Fischbacher, U. 2003. The nature of human altruism. *Nature* **425**: 784–791.

Fehr, E. & Gachter, S. 2002. Altruistic punishment in humans. *Nature* **415**: 137–140.

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

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., Bowles, S., Boyd, R., & Fehr, E. 2003. Explaining altruistic behavior in humans. *Evol. Hum. Behav.* **24**: 153–172.

Guzman, R., Rodriguez-Sickert, C., & Rowthorn, R. 2007. When in rome, do as the romans do: the coevolution of altruistic punishment, conformist learning, and cooperation. *Evol. Hum. Behav.* **28**: 112–117.

Hagen, E.H. & Hammerstein, P. 2006. Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theor. Popul. Biol.* **69**: 339–348.

Hamann, K., Warneken, F., Greenberg, J.R., & Tomasello, M. 2011. Collaboration encourages equal sharing in children but not in chimpanzees. *Nature* **476**: 328–331.

Hamilton, W.D. 1963. The evolution of altruistic behavior. *Am. Nat.* **97**: 354–356.

Hamilton, W. 1964. The genetical evolution of social behaviour, i & ii. *J. Theor. Biol.* **7**: 1–52.

Hamlin, J.K., Wynn, K., & Bloom, P. 2007. Social evaluation by preverbal infants. *Nature* **450**: 557–559.

Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *J. Econ. Behav. Organ.* **53**: 3–35.

Henrich, J. & Boyd, R. 2001. Why people punish defectors. weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* **208**: 79–89.

Henrich, N. & Henrich, J. 2007. *Why Humans Cooperate*. Oxford University Press, Cambridge, MA.

Hofbauer, J. & Sigmund, K. 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

Hoffman, E., McCabe, K., & Smith, V. 1996. Social distance and other-regarding behavior in dictator games. *Am. Econ. Rev.* **86**: 653–660.

Lehmann, L. & Feldman, M.W. 2008. The co-evolution of culturally inherited altruistic helping and cultural transmission under random group formation. *Theor. Popul. Biol.* **73**: 506–516.

Lehmann, L. & Keller, L. 2006. The evolution of cooperation and altruism – a general framework and a classification of models. *J. Evol. Biol.* **19**: 1365–1376.

Lehmann, L., Rousset, F., Roze, D., & Keller, L. 2007. Strong-reciprocity or strong-ferocity? a population genetic view of the evolution of altruistic punishment. *Am. Nat.* **170**: 661.

Lehmann, L., Feldman, M.W., & Foster, K.R. 2008a. Cultural transmission can inhibit the evolution of altruistic helping. *Am. Nat.* **172**: 12–24.

Lehmann, L., Foster, K.R., Borenstein, E., & Feldman, M.W. 2008b. Social and individual learning of helping in humans and other species. *Trends Ecol. Evol.* **23**: 664–671.

Lumsden, C.J. & Wilson, E.O. 1981. *Genes, Mind, and Culture: The Coevolutionary Process*. Harvard University Press, Cambridge, MA.

Nesse, R. 2001. The smoke detector principle. natural selection and the regulation of defensive responses. *Ann. N. Y. Acad. Sci.* **935**: 75–85.

Nowak, M.A. 2006a. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press of Harvard University Press, Cambridge, MA.

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

Ohtsuki, H. & Nowak, M. 2008. Evolutionary stability on graphs. *J. Theor. Biol.* **251**: 698–707.

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M.A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**: 502–505.

Richerson, P.J. & Boyd, R. 2005. *Not by Genes Alone : How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.

Rousset, F. 2004. *Genetic Structure and Selection in Sudivided Populations*. Princeton University Press, Princeton.

Santos, F.C., Santos, M.D., & Pacheco, J.M. 2008. Social diversity promotes the emergence of cooperation in public goods games. *Nature* **454**: 213–216.

Skyrms, B. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge.

Tarnita, C.E., Antal, T., Ohtsuki, H., & Nowak, M.A. 2009. Evolutionary dynamics in set structured populations. *Proc. Natl Acad. Sci. USA* **106**: 8601–8604.

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

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

Warneken, F., Lohse, K., Melis, A., & Tomasello, M. 2010. Young children share the spoils after collaboration. *Psychol. Sci.* **22**: 267–273.

West, S., Griffin, A., & Gardner, A. 2007a. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: R661–R672.

West, S., Griffin, A., & Gardner, A. 2007b. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**: 415–432.

West, S., El-Mouden, C., & Gardner, A. 2011. 16 misconceptions about the evolution of cooperation in humans. *Evol. Hum. Behav.* **32**: 231–262.

## Supporting information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Hamilton's rule for genetic and cultural change.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.