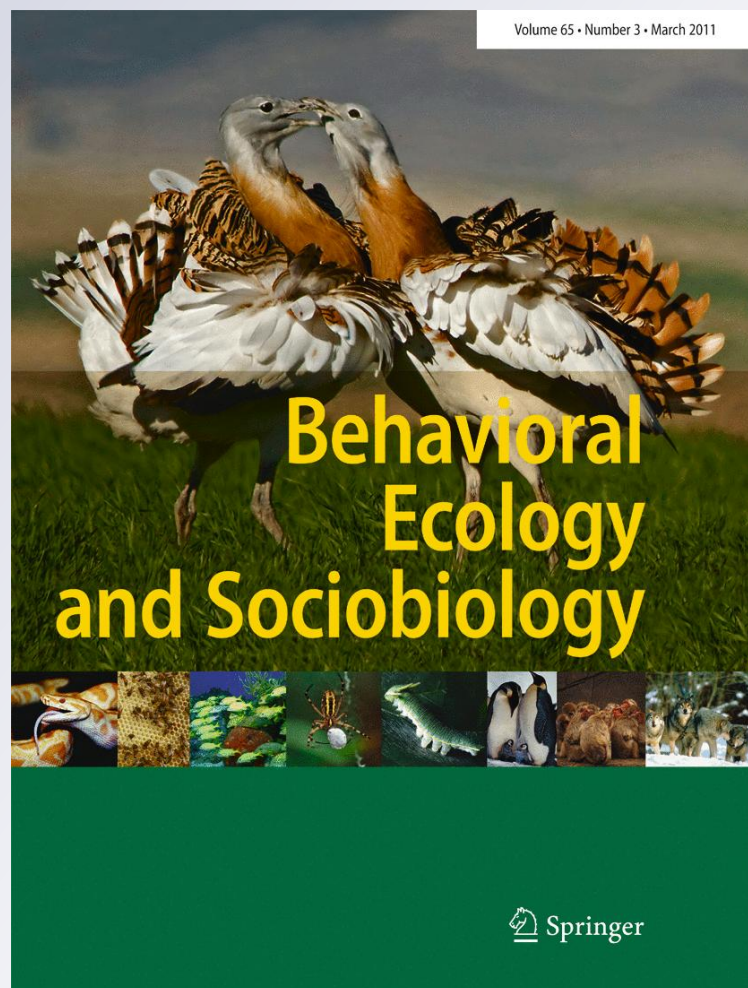


# *Rapid cultural adaptation can facilitate the evolution of large-scale cooperation*

**Behavioral Ecology and Sociobiology**

ISSN 0340-5443  
Volume 65  
Number 3

Behav Ecol Sociobiol (2010)  
65:431-444  
DOI 10.1007/  
s00265-010-1100-3



**Your article is protected by copyright and all rights are held exclusively by The Author(s). This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

# Rapid cultural adaptation can facilitate the evolution of large-scale cooperation

Robert Boyd · Peter J. Richerson · Joseph Henrich

Received: 14 May 2010 / Revised: 28 October 2010 / Accepted: 29 October 2010 / Published online: 1 December 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Over the past several decades, we have argued that cultural evolution can facilitate the evolution of large-scale cooperation because it often leads to more rapid adaptation than genetic evolution, and, when multiple stable equilibria exist, rapid adaptation leads to variation among groups. Recently, Lehmann, Feldman, and colleagues have published several papers questioning this argument. They analyze models showing that cultural evolution can actually reduce the range of conditions under which cooperation can evolve and interpret these models as indicating that we were wrong to conclude that culture facilitated the evolution of human cooperation. In the main, their models assume that rates of cultural adaptation are not

strong enough compared to migration to maintain persistent variation among groups when payoffs create multiple stable equilibria. We show that Lehmann et al. reach different conclusions because they have made different assumptions. We argue that the assumptions that underlie our models are more consistent with the empirical data on large-scale cultural variation in humans than those of Lehmann et al., and thus, our models provide a more plausible account of the cultural evolution of human cooperation in large groups.

**Keywords** Cooperation · Altruism · Culture

---

Communicated by Guest Editor J. Marshall

This contribution is part of the Special Issue “Mathematical Models in Ecology and Evolution: Darwin 200” (see Marshall et al. 2010).

**Electronic supplementary material** The online version of this article (doi:10.1007/s00265-010-1100-3) contains supplementary material, which is available to authorized users.

---

R. Boyd (✉)  
Department of Anthropology, UCLA,  
Los Angeles, CA 90095, USA  
e-mail: rboyd@anthro.ucla.edu

P. J. Richerson  
Department Environmental Science and Policy  
and Center for Population Biology, UC Davis,  
Davis, CA 95616, USA  
e-mail: pjricherson@ucdavis.edu

J. Henrich  
Departments of Psychology and Economics,  
University of British Columbia,  
Vancouver, BC, Canada V6T 1Z4  
e-mail: joseph.henrich@gmail.com

...although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection.

Charles Darwin (1874), *Descent of Man*, p. 132

Most contemporary human societies are larger and more cooperative than those of other mammals. In most mammal species, cooperation is limited to small groups; there is little division of labor, no trade, no large-scale conflict, no social support for sick or disabled, and no moral systems enforced by third parties (Clutton-Brock 2009). In stark contrast,

even in simple foraging societies people cooperate in large groups (e.g., Hill 2002). Division of labor, trade, and intergroup conflict are important nearly everywhere (Ridley 2010; Bowles 2009). The sick and disabled are often cared for, and social life is regulated by moral systems usually shared by thousands of individuals enforced, albeit imperfectly, by third party sanctions (Boehm 1999).

In chapter 5 of the *Descent of Man*, Darwin argued that the distinctive features of human sociality resulted from selection among groups with different standards of morality—conforming to the local moral system does not lead to individual disadvantage, but tribes with more effective moral systems would replace those with less effective systems. What Darwin did not seem to realize, but theoretical developments over the last 40 years or so have made clear, is that this kind of process can be effective only if something maintains sufficient heritable variation among groups (or equivalently, maintains sufficient relatedness within groups). In most animal species, heritable genetic variation among groups results from limited migration and small group size, and it seems unlikely that these processes can generate sufficient variation among large groups in mammal species with small families and substantial migration between groups.

In a series of publications, we have argued that cultural evolution leads to much variation among very large human groups and that this fact may explain our distinctive sociality. This hypothesis rests on three assumptions:

1. Mechanisms of reciprocity, reputation, signaling, and punishment can stabilize a vast range of heritable behaviors ranging from ruthless spite to prosocial cooperation (Henrich 2006). Adaptive learning biases in human cognition, including conformist learning, can both re-enforce these effects, and in some circumstances, stabilize many equilibria in the absence of reputational mechanisms (Boyd and Richerson 1985; Henrich and Boyd 2001; Henrich 2009).
2. Cultural adaptation is much more rapid than genetic adaptation. Adaptive cultural processes are strong relative to migration and other mixing processes, and thus, the cultural system can sustain large, persistent differences in behavioral patterns between neighboring social groups. By contrast, because natural selection on genes is typically weak relative to migration among neighboring groups of mammals, it does not typically maintain substantial genetically transmitted differences among such groups.
3. As hypothesized by Darwin, competition between groups favors the spread of culturally transmitted behaviors that enhance the competitive ability of groups.

We have developed a number of theoretical models that demonstrate the cogency of this argument, presented empirical data that suggests that the assumptions of the models are realistic, and reviewed empirical examples of cultural changes that result from competition between groups. Summaries and further references can be found in Richerson and Boyd (2005), Boyd and Richerson (2009), Henrich (2004, 2008), and Henrich and Henrich (2007).

In four recent papers, Lehmann and a series of collaborators have challenged the logical foundations of our hypothesis about the evolution of large-scale cooperation in humans. In two papers, Lehmann et al. (2008a, b) present models showing that a particular form of cultural transmission can actually reduce the range of parameters that allow the evolution of altruistic traits. Based on these results, they argue that we have been premature in concluding that culture facilitates the evolution of cooperation. Two other papers argue against specific components of our hypothesis: Lehmann and Feldman (2008) analyze models suggesting that a conformist social learning psychology does not enhance the evolution of altruism. Lehmann et al. (2007) argue that the punishment of noncooperators cannot evolve unless the transmitted variants giving rise to punishment and to cooperation are linked.

In this paper, we show that Lehmann and colleagues reach different conclusions than we do because they make different assumptions, both about the processes that maintain variation among groups and about the selective processes that lead to the spread of group beneficial variants. In both studies of Lehmann et al. (2008a, b), they assume that rates of cultural adaption are so weak that cultural adaptation does not affect variation among groups. In Lehmann and Feldman (2008), they assume complete mixing every generation so that cultural adaptation cannot maintain persistent variation among groups no matter how strong it might be. In Lehmann et al. (2007), the conclusions about the necessity of linkage are based on a model that does not include different rates of extinction for groups held at different stable frequencies of cooperation by variation in the frequency of punishment. These differences in assumptions change both the nature of the forces that generate relatedness among interacting individuals and the way that groups compete. As a result, the nature of the evolutionary processes shaping social behavior is very different than in models we have made. Their work provides a competing explanation for the cultural evolution of human cooperation. The key questions are thus empirical not logical: which assumptions better fit with what we know empirically about human learning, cultural diffusion, cultural variation, and human cooperation?

Lehmann et al. do not clearly delineate how their models differ from ours, and in some places, seem to imply that



they are structurally similar, or that the differences in assumptions are unimportant. In what follows, we delineate the differences in assumptions between our work and that of Lehmann and colleagues and explain why these different assumptions yield different conclusions. We begin with the central issues discussed in Lehmann et al. (2008a, b) and then turn to more specific questions discussed in Lehmann and Feldman (2008) and Lehmann et al. (2007). We conclude by presenting data indicating that the Lehmann et al. (2008a, b) models seem inconsistent with the observed scale of cultural variation, and thus, do not provide a plausible alternative explanation for the evolution of large scale of human cooperation. However, these models may be relevant to the cultural evolution of cooperation on smaller demographic scales.

### Does culture facilitate the evolution of cooperation?

Lehmann et al. (2008b) suggest that we have argued that all forms of cultural inheritance make it easier for cooperation to evolve. They then refute this argument by analyzing models in which altruism evolves under a narrower range of conditions with cultural transmission than with vertical, gene-like transmission. However, we have never argued that culture *always* makes it easier for cooperation to evolve. Given the vast diversity of possible cultural transmission mechanisms, and the poor state of our empirical knowledge about such mechanisms, it would be foolish to claim that *any* imaginable form of cultural transmission facilitates the evolution of cooperation. Instead, we have made a much more specific argument. Human cooperation depends on systems of norms maintained by punishment, reputation, conformist cultural transmission, and other learning biases. Because cultural adaptation can be much more rapid than genetic adaptation, cultural evolution generates more stable behavioral variation among large groups, even very large groups. Like most models of this kind, our work reveals conditions both favorable and unfavorable to the spread of larger-scale cooperation via cultural evolution.

The models in Lehmann et al. (2008b) are different from ours in two important respects. First, they assume that adaptive forces are very weak, and as a result, their work closely resembles many recent inclusive fitness models used in population genetics (e.g., Lehmann and Keller 2006). Second, they assume island model rather than stepping stone population structure, as in Boyd and Richerson (2002), the model that they use as a comparison.

To see why these alternative assumptions yield different evolutionary dynamics, let us consider these models in detail. Lehmann et al. (2008a, b) seek to compare the evolution of altruism under horizontal and vertical cultural

transmission. The vertical model is similar to genetic models of the evolution of altruism in viscous populations (e.g., Taylor 1992) and thus serves as a baseline for comparison. The population is structured into a large number of groups. Local population regulation maintains groups at a fixed, finite size and during each generation groups exchange migrants with all other groups. There are two variants, an altruistic variant that produces a benefit to all group members and a selfish variant that does not. Benefits and costs affect individual survival so that altruists have lower survival rates than selfish individuals within their own group, but groups with more altruists have higher average fitness and produce more emigrants.

The horizontal cultural evolutionary model is meant to be a generalization of the model in Boyd and Richerson (2002). Again the population is structured into a large number of fixed, finite-sized groups. Lehmann et al. consider two different payoff structures: In the body of the paper, they present a model in which there is an altruistic variant and a selfish variant just as in the vertical model. This is a crucial modification of Boyd and Richerson (2002) where the core assumption is that there is are two norms, both stable when common. Altruistic variants cannot spread in that model because they are not favored in any group, and therefore, adaptive processes cannot maintain variation among groups. However, in the online supplementary materials of their paper, Lehmann et al. analyze a second model in which the payoff structure is given by a Stag Hunt game, as in Boyd and Richerson (2002). There are two variants. Each has higher payoff than the other when it is common, but one type increases the average payoff of all in the group while the other type does not. Cultural transmission is payoff biased: individuals meet another individual, the “model,” and adopt the model’s behavioral variant with a probability proportional to the difference in payoffs. With some probability, the model is drawn from another randomly chosen group; otherwise, the model is drawn from the individual’s own group. This leads to the exchange of cultural variants among groups. Models drawn from groups with higher frequencies of group beneficial behavior are more likely to be copied.

To analyze these models, Lehmann et al. restrict parameter values so that the adaptive forces, natural selection in the vertical model and payoff-biased imitation in the horizontal model, are much weaker than changes caused by the flow of cultural variants among groups. As a result, the relatedness within groups (or, equivalently, the variation among groups) rapidly comes to a “quasiequilibrium” determined only by interplay of random sampling variation and the flow of heritable variants among groups—adaptive forces (natural selection and payoff-biased transmission) are ignored. This assumption greatly simplifies the analysis for two reasons: First, because relatedness equi-

brates much more rapidly than changes in cultural variant frequencies due to adaptive forces, the equilibrium relatedness can be taken as a fixed parameter in calculating the inclusive fitness of each variant. Second, adaptive forces are slow enough that changes in frequency result from the average relatedness over all groups—once its effect on inclusive fitness is taken into account, group structure can be ignored. Lehmann and colleagues use this approach to analytically derive the conditions under which an altruistic variant can increase and show that the conditions are less restrictive under vertical inheritance than horizontal transmission because emigrants leave the group and therefore, there is less competition among descendants in the vertical case.

In contrast, we (Boyd and Richerson 2002, 1990) assume that adaptive processes in cultural evolution are strong compared to migration. This assumption is supported by empirical evidence from many sources. Recent studies of cultural transmission in humans suggest that learning mechanisms generate potent effects (see reviews in Henrich and Gil-White 2001; Henrich and Henrich 2007: chapter 2; McElreath et al. 2008; Mesoudi 2009), and the literatures on the diffusion of innovations, public health, business, history, and anthropology provide much evidence of rapid cultural change (reviewed in Richerson and Boyd 2005; Henrich 2001). Often, novel cultural traits, including new norms and practices, spread to fixation in less than one generation. Since trait frequencies evolve rapidly, local conditions are important. We also focus only on payoff structures typical for interactions involving reputation, repeated interaction, multi-stage games, and contingent behavior that readily generate multiple stable equilibria. This means that adaptive cultural forces like payoff-biased transmission can lead to substantial differences in variant frequencies in different groups.

Such systems cannot be analyzed using the weak selection, quasiequilibrium approach used by Lehmann et al., and instead, must be represented as high dimension dynamic systems that include a state variables representing the frequencies of the variants in each group. It is not easy to solve such systems analytically, particularly if groups are finite and therefore the system is stochastic. However, simulating the behavior of such systems is straightforward.

To illustrate the impact of the different assumptions about the strength of adaptive forces, we simulated a model very similar to the horizontal model in Lehmann et al. (2008b) with weak payoff bias and strong payoff bias. A population of size  $N$  is divided into groups with  $n$  individuals. There are two cultural variants, labeled 0 and 1. Let  $x_i$  be the frequency of variant 1 in group  $i$ . The life cycle has three steps: First, there is a mutation-like process. With probability  $\mu$ , each variant spontaneously transforms into the alternative variant. In all simulations,  $\mu=10^{-4}$ .

Second, individuals interact socially. The payoffs in group  $i$  are:

$$w_0 = 1 + gx_i \quad (1)$$

$$w_1 = 1 + s(x_i - \tilde{x}) + gx_i \quad (2)$$

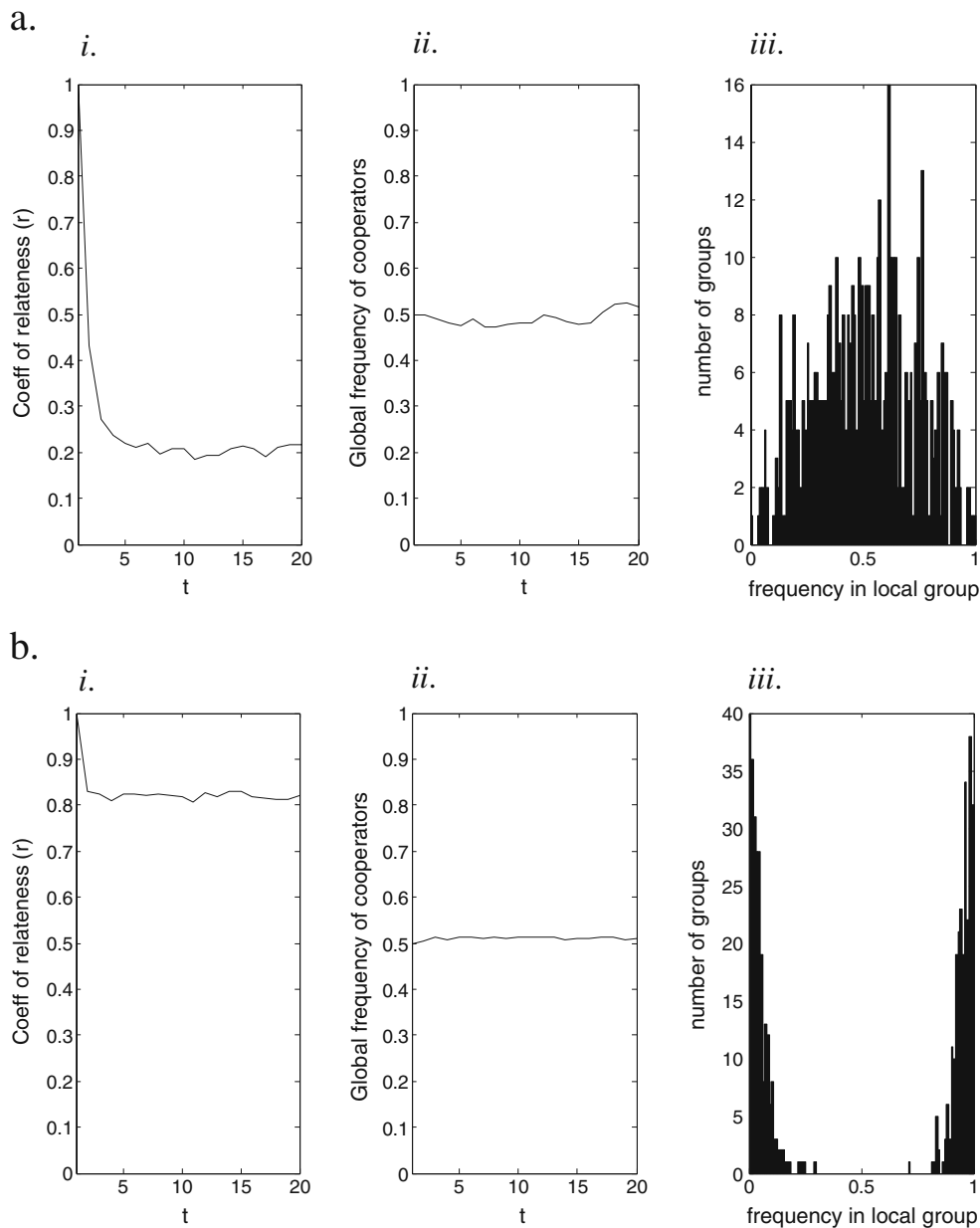
Thus, variant 1 produces a benefit to every member of the group proportional to  $g$  and has higher payoff than variant 0 if  $x_i > \tilde{x}$  where  $s>0$  and  $0 < \tilde{x} < 1$ . Thus, both variant 0 and variant 1 are favored when common. The parameter  $s$  controls the magnitude of this effect. Third, after social interaction, individuals meet a model and observe its payoff. With probability  $m$  the model is drawn from another, randomly chosen, group, and with probability  $1-m$  from the individual's own group. The learner adopts the model's variant with probability

$$\frac{1}{2}(1 + \beta(w_m - w_f)) \quad (3)$$

Where  $w_m$  and  $w_f$  are the payoffs of the model and focal, respectively, which captures individuals' tendency to switch to the model's behavior if the model has a higher payoff. The parameter  $\beta$  controls the strength of biased transmission, with larger values of  $\beta$  creating more rapid adaptive change. The MatLab code used in the simulations is given in the [supplementary materials](#) and is also available from the first author on request.

Increasing the strength of biased transmission changes the nature of the forces that shape relatedness within groups. To see why, consider the special symmetric case in which  $\tilde{x} = 0.5$  and  $g=0$ . There are 500 groups each with 100 individuals. Initially the frequency of variants 0 and 1 is one half, and groups are either all one type or all the other type. This means the relatedness within groups (approximately the fraction of variance among groups for groups of this size) in the population is initially equal to one. We adopt these artificial symmetrical initial conditions for clarity. The qualitative conclusions listed below will hold as long as the initial conditions lead to a steady state in which each variant is at high frequency in at least one group in the strong bias case.

Figure 1(a) shows the results for a weak transmission bias ( $s=0.1$ ,  $\beta=0.01$ ). Relatedness declines rapidly to steady state value of around 0.2, about the value predicted by the weak bias approximation given in Lehmann et al. (2008b), and as a result, adaptive processes like selection and payoff-biased transmission can lead to the spread of low levels of individually costly group beneficial behaviors. Compare these results to those shown in Fig. 1(b) where bias is strong ( $s=0.1$ ,  $\beta=0.5$ ). Now, the relatedness within groups stabilizes at a much higher value, around 0.8. Note that we are conforming to the contemporary definition of relatedness as a measure of the extent to which an



**Fig. 1** **a** Model behavior when biased cultural transmission is weak and both variants have a higher payoff when common. The payoffs are symmetric so that the basins of attraction of both equilibria are the same, and there is no group benefit. There are 500 groups each with 100 individuals. The probability of choosing a model from outside the group is 0.02. Initially, half of the groups are all one variant and other half are the other variant. *i* The relatedness within groups converges rapidly to a value predicted by the analytical treatment given in Lehmann et al. (2008b). *ii* The overall frequency does not change due to the symmetry of the model. *iii* The distribution of frequencies across groups in the final time period. The distribution is unimodal, but because relatedness is approximately equal to 0.2, the variance of this distribution is much greater than that would be predicted if groups were sampled at random

with probability 0.5. **b** Model behavior when biased cultural transmission is strong. Parameter values as in **a**, except that the strength of payoff-biased transmission is increased by a factor of 20 ( $\beta=0.5$ ). *i* The relatedness within groups converges rapidly to a value that is much higher than in the weak bias case. Relatedness here is a measure of the extent to which one individual's type predicts the types of others in its group, but is mostly not due to common descent. *ii* The overall frequency does not change. *iii* The distribution of frequencies across groups in the final time period. This shows why relatedness is so high—the cultural analog of disruptive selection creates a bimodal distribution of frequencies across groups. Because most of the groups are either mostly one variant or mostly the alternative variant, an individual's own variant is a good predictor of the variants of others in its group

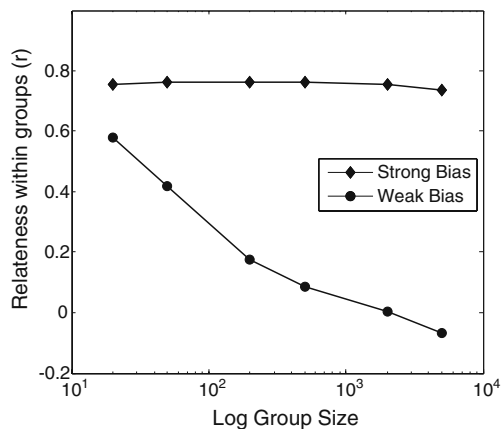
individual's variant predicts the variants of others in the group extended to cultural inheritance (Alison 1992). When selection is strong, this does not, necessarily, measure the extent to which individuals are similar by common descent.

When groups are large, relatedness is approximately the proportion of variation between groups.

To see why relatedness within groups is greater in the strong bias case, consider the distribution of frequencies of

the group beneficial trait across groups in the two cases. When bias is weak, the distribution of frequencies across groups is unimodal. The variance across groups exceeds the level that would result from random group formation, so an individual's own type predicts the types of others in its group. When bias is strong, the cultural analog of disruptive selection creates a bimodal distribution of frequencies across groups. Because most of the groups are composed of mostly one variant or mostly the alternative variant, an individual's own variant is a much better predictor of the variants of others in its group than in the weak bias case.

As is shown in Fig. 2, the effect of group size on relatedness is very different in weak bias and strong bias cases. When payoff-biased transmission is weak ( $\beta=0.01$ ), relatedness gets smaller as group size increases because relatedness derives from common descent and the probability that two individuals have the same cultural parent declines as groups get bigger. In contrast, when the payoff-biased transmission is strong ( $\beta=0.5$ ), variation among groups and (therefore, relatedness within groups) is mainly created and maintained by biased transmission. Because the strength of bias does not depend on group size, the within group relatedness remains high even when groups are very large, and as a consequence, very few individuals are similar due to shared descent. In fact, in the limit of infinite groups (as assumed in Boyd and Richerson 2002), the probability of common descent is zero, but relatedness has the approximately same value shown in Fig. 2.



**Fig. 2** Mean relatedness within groups during the last period of the simulation as a function of group size. When payoff-biased transmission is weak ( $\beta=0.01$ ), relatedness among group members declines as group size increases because relatedness derives from common descent, and the probability that two individuals have the same cultural parent declines as groups get bigger. In contrast, when the payoff-biased transmission is strong ( $\beta=0.5$ ), variation among groups (and therefore, relatedness within groups) is mainly created and maintained by bias and common descent plays a minor role. Since the strength of bias does not depend on group size, the within group relatedness does not depend on group size

Both strong and weak biases can generate group beneficial behavior but in very different ways. Lehmann et al. (2008b) derive a condition for the group beneficial variant to increase when rare. In the notation of the present model, this condition can be rearranged to become:

$$s\tilde{x} < \frac{R(s + mg)}{1 - R(1 - m)} \tag{4}$$

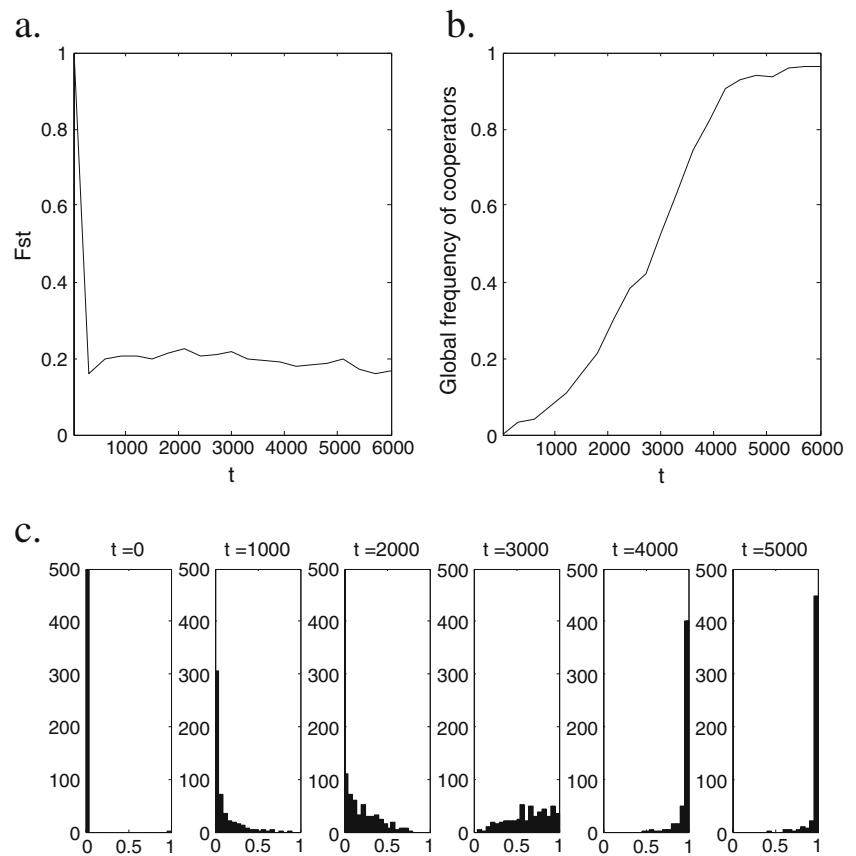
Where  $R = 1/(nm(2 - m))$  is the equilibrium relatedness if bias is weak and groups are large. When individuals with the group beneficial variant are rare, they suffer a payoff disadvantage relative to the common variant. The magnitude of this disadvantage is proportional to  $s\tilde{x}$ , so the left hand side is the cost associated with the group beneficial variant when it is rare. The right hand side is proportional to the relatedness and gives the inclusive fitness benefit associated with the group beneficial variant. Because  $R$  is proportional to  $1/n$ , this condition becomes hard to satisfy when groups are large, and thus, this mechanism cannot lead to the spread of the group beneficial variant in the infinite groups assumed in Boyd and Richerson (2002). The dynamics in the weak bias case are given in Fig. 3. Relatedness quickly converges to the predicted equilibrium value, and the group beneficial trait increases because the inclusive fitness benefits exceed the cost. Relatedness here is due to common descent.

In contrast, now suppose that payoff bias is strong enough relative to mixing that once either trait is common within a group, it will remain common, and that the group beneficial trait is initially common in a single group. In this case, nothing happens even though (4) is satisfied. The group beneficial trait still raises payoffs, and individuals in the group in which it is common are still disproportionately imitated by individuals in other groups, and relatedness is high. However, unlike weak selection models, producing more emigrants is not enough. The group beneficial trait does not spread to groups in which it is not common because payoff bias acts strongly against the trait in such groups. Thus, the group beneficial trait remains common in the initial population, but cannot spread.

The group beneficial trait can spread, even in very large groups, if the model is modified in one of two ways. First suppose that groups with a higher frequency of the group beneficial trait are less likely to suffer extinctions, and that empty habitats are recolonized by individuals drawn from a single randomly selected group (Boyd and Richerson 1990). This assumption is consistent with ethnographic, historical, and archeological research (Soltis et al. 1995; Keeley 1997; Bowles 2009). Figure 4 shows that the group beneficial trait increases and relatedness remains high even though the groups are an order of magnitude larger than in the weak bias case. When  $n=1,000$  and payoff bias is weak,



**Fig. 3** The dynamics of the group beneficial trait when pay-off bias is weak ( $n=100$ ,  $m=0.02$ ,  $s=0.2$ ,  $g=1$ ,  $\bar{x}=0.25$ ,  $b=0.05$ ). There are 500 groups. Initially, there is one group in which the frequency of the group beneficial variant is one in one group and zero in all other groups. **a** Relatedness quickly attains the predicted equilibrium value ( $\approx 0.2$  for these parameter values) and **b** shows that because the inclusive fitness benefits exceed the costs, the group beneficial trait increases in frequency. **c** Distribution of frequencies across groups as the group beneficial trait increases. Individuals with the group beneficial variant rapidly diffuse throughout the population, and then the distribution of frequencies results from the interplay of migration and common descent. Adaptation can be understood as responding to the average over the entire distribution



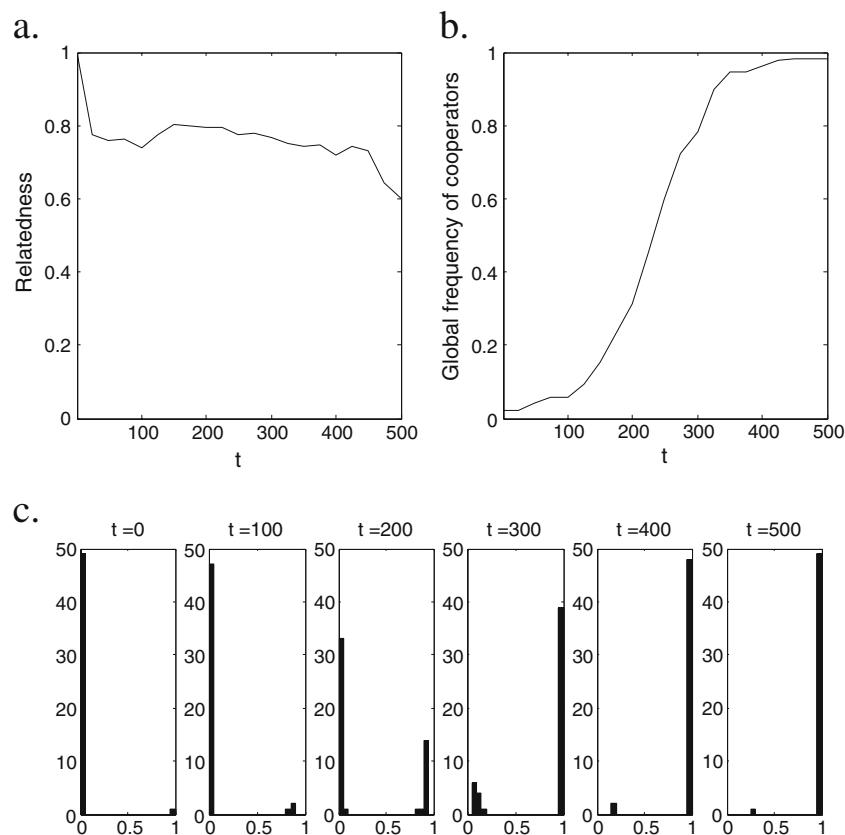
$R$  is very low, and the group beneficial trait is unlikely to spread. Notice that the dynamics of the distribution of frequencies across groups is very different than in the weak bias case—throughout the process, strong bias maintains quite different frequencies of the group beneficial trait among groups, and adaptation occurs because groups with a low frequency of the group beneficial variant are more likely to go extinct than groups with a high frequency of the variant. Very similar dynamics result if extinctions are random, but groups with high frequencies of the group beneficial variant are more likely to grow and colonize empty patches, or if groups engage in conflicts with other groups and groups with a higher frequency of the group beneficial variant are likely to be victorious.

The second way to modify the strong bias model is to use a stepping stone population structure so that individuals only imitate models in a small number of neighboring groups (Boyd and Richerson 2002). Lehmann et al. assume island model migration, and this is sensible given their assumption of weak bias—there is only a modest difference between island and stepping stone models in the weak bias case because traits diffuse rapidly throughout the population. However, when bias is strong, the difference is crucial. If the group beneficial trait becomes common in one group, the high payoff causes individuals in neighboring groups to adopt the group beneficial variant, which can tip the

neighbors into the basin of attraction of the group beneficial trait. This results in a cascade that spreads the group beneficial trait throughout the population. This process is formally similar to genetic models of the third phase of Wright's shifting balance process (e.g., Gavrillets 1996), the dynamics of hybrid zones (Barton 1979), and early models of reciprocal altruism (Boorman and Levitt 1973, 1980). The difference is that cultural adaptation can maintain sharp heritable behavioral differences among neighboring human social groups—neighboring ethnolinguistic groups numbering a few thousand people living a few kilometers apart can have mutually unintelligible languages and strikingly different moral systems. Step clines on this scale do not seem to occur with genetically transmitted influences on social behavior within other large mobile mammal species because, we believe, selection is usually not strong enough relative to migration among local groups.

### Conformism can facilitate the evolution of cooperation

In a series of studies, we (e.g., Boyd and Richerson 1985: chapter 7; Henrich and Boyd 2001) have argued that a “conformist” bias in social learning may facilitate the evolution of altruism. Conformist bias occurs when individuals are disproportionately likely to acquire the



**Fig. 4** The dynamics of the group beneficial trait when bias is strong and the group beneficial trait lowers extinction rates ( $n=1,000$ ,  $m=0.02$ ,  $s=0.2$ ,  $g=0$ ,  $\bar{x}=0.5$ ,  $\beta=1.0$ ). There are 50 groups. The probability of extinction in group  $i$  each time period is  $\varepsilon(1-x_i)$  where  $x_i$  is the frequency of the group beneficial trait in group  $i$  and  $\varepsilon=0.015$ , a value that when combined with the distribution of frequencies yields extinction rates roughly consistent with those observed in tribal societies (Soltis et al. 1995) assuming simulation time periods of 1 year. Empty habitats are recolonized by immigrants from a single surviving group. Initially, there is one group in which the frequency of the group beneficial variant is one and zero in all other groups. **a**

Relatedness quickly reaches an equilibrium value of about 0.8 even though groups are quite large because strong bias maintains the group beneficial norm at either a high or low frequency in every group. Here, relatedness is mainly not the result of common descent. **b** The group beneficial trait spreads because groups with a high frequency of the group beneficial trait are much less likely to become extinct. **c** Distribution of frequencies across groups as the group beneficial trait increases. Throughout the process, strong bias maintains groups at strongly different frequencies, and adaptation occurs because groups with a low frequency of the group beneficial variant are more likely to go extinct than groups with a high frequency of the variant

more common variant from among the variants that they observe. So, for example, if three of four models have one variant and the fourth model has a second variant, the probability of acquiring the common variant is greater than three quarters. Such a bias creates an evolutionary force that increases the frequency of the more common variant in the population. Theoretical work predicts that natural selection will favor conformist-biased social learning in some kinds of spatially and temporally varying environments (Henrich and Boyd 1998; Nakahashi 2007; Wakano and Aoki 2007; Kendal et al. 2009) and when social learning has a high error rate (Henrich and Boyd 2002). In both cases, more than one cultural variant will coexist, but the most adaptive variant will tend to be more common. Thus, all other things being equal, a predisposition to adopt the more common variant increases the chance of acquiring the most adaptive variant. Recent experimental work indicates that human

social learning is subject to conformist bias (Efferson et al. 2008; McElreath et al. 2008).

Conformist bias can facilitate the evolution of altruism in large groups because it creates multiple stable equilibria, which in turn can create and sustain variation among groups. To see how this works, consider what happens when a potentially altruistic trait evolves under the influence of payoff-biased transmission alone. Altruists produce a benefit to the group at a cost to themselves. This means that payoff bias decreases the frequency of altruists in every group unless groups are small enough or migration rates are low enough that there is enough relatedness within groups to create a compensating inclusive fitness benefit (or equivalently, maintain sufficient variation among groups so that there is sufficient between group selection in favor of altruists). Now, suppose that in addition to payoff bias, there is also a conformist bias. Remember that conformist

bias tends to increase the common type. So if conformist bias is strong enough, it can maintain altruists at high frequency in some groups even though they achieve a lower payoff. Thus, conformist bias creates multiple equilibria in situations with an altruistic payoff structure, and if the bias is strong compared to migration in a structured population, this allows altruism to evolve in very large groups for the same reasons described above.

This mechanism may be of particular importance for the maintenance of punishment (Henrich and Boyd 2001). Because punishment suffers a smaller disadvantage when cooperation is common, even weak conformism (combined with payoff-biased transmission) can stabilize punishment against the invasion of second-order free riders who cooperate but do not punish. This in turn can stabilize cooperation, and thus, create the conditions that allow selection among groups to spread punishment and cooperation in much larger groups than is possible without conformism (Guzmán et al. 2007). Interestingly, these authors also show that conformism is favored in this model without any other form of variation in payoffs.

Lehmann and Feldman (2008) study the effect of conformism on the evolution of altruism. They analyze a model in which individuals interact in finite groups that are formed each generation by random sampling from the population. Once in groups, individuals then undergo an episode of horizontal conformist cultural transmission that can amplify initial differences in frequencies among groups caused by sampling. Lehmann and Feldman show that when this horizontal transmission episode is subject to a conformist bias, it is harder for rare altruistic cultural variants to increase. They conclude, “Our results illustrate that a frequency-dependent assimilation rule such as biased conformist transmission...is unlikely to promote the evolution of altruistic helping in situations where it is otherwise difficult to explain, that is, in populations of large size when the trait is initially rare. (p. 514)”

Lehman and Feldman reach different conclusions than we do about the effect of conformist bias because they have made different assumptions about the effects of adaptation and mixing. In their model, conformist-biased cultural transmission cannot maintain persistent variation among groups because there is complete mixing every generation. Suppose that the altruistic variant is rare in the population. If chance assortment leads to a group with a high frequency of the altruistic variant, it will have higher average fitness. However, offspring produced by the group are randomly mixed with offspring from other groups. As a result, the altruistic variant will remain rare, and conformist bias will act to decrease its frequency. However, if biased transmission is strong compared to mixing, this group will persist in maintaining the altruistic variant at high frequency, and altruism can spread through the population as a whole

either due to the differential extinction or stepping stone processes discussed above. Thus, the altruistic variant can spread when it is rare in the population as a whole as long as some random nonadaptive process causes it to become common in a *single* local group. The same is true for the case of alternative norms discussed above.

There are at least three plausible processes that can do this. First, sampling variation leads to cultural drift, a process closely analogous to genetic drift (Cavalli-Sforza and Feldman 1981; Neiman 1995; Shennan 2001). Such cultural drift can lead to “peak shifts” for the same reasons as genetic drift. Moreover, if as some authors have argued (Claidiere and Sperber 2007; Griffiths et al. 2008; Henrich 2009), the cultural analog of mutation rates are much higher than genetic mutation rates, the equilibrium frequency of deleterious traits resulting from the balance of adaptive bias and cultural mutation will be higher, and therefore, waiting times for peak shifts should be much shorter. Note that this mechanism depends on sampling variation and should be less effective in large groups. It will also be less effective when adaptive forces are strong.

There are, however, two mechanisms in which waiting times for peak shifts are not necessarily reduced in large populations or when adaptive forces are strong. In random environments, linkage leading to “genetic draft” (Gillespie 2000, 2001) can also lead to peak shifts, and the rate at which this occurs does not depend strongly on group size and may actually increase with the strength of selection. Linkage in cultural transmission means that you acquire two traits from the same person, either because that person is a particularly salient model or because acquiring one trait increases receptivity to a second trait. This leads to correlations between traits analogous to linkage disequilibrium. Then payoff biases that increase the frequency of one trait also tend to increase the frequency of correlated traits, and in a fluctuating environment, this leads to random, nonadaptive temporal variation in frequencies that can cause the shift from one basin of attraction to another. For example, suppose weather patterns shifted in lakeside village and that fishermen, who previously formed a small cooperative of low status men who could not become hunters or warriors, suddenly became the primary providers of protein, and locally prestigious for their, now-valued, fishing skills. Selected as potent cultural models for their fishing skills, these men might also transmit their cooperativeness broadly across the village, tipping the community into a cooperative basin of attraction.

Finally, both individual learning and biased transmission depend on environmental cues. One cue will cause an individual to preferentially adopt one variant, while a different cue will cause her to adopt the alternative variant. The cues observed by members of a group may often be highly correlated. For example, the disastrous loss of WWII

seems to have shifted the Japanese from a militaristic moral system to a more pacifistic one (Dower 2000). However, such cues often have a strong random component, and as a result, will lead to random fluctuations in the frequency of different behaviors. For example, the Battles of the Coral Sea and Midway might easily have gone differently, and if they had and the US sued for peace, the Japanese might have “learned” that militarism pays. This process does not depend on group size and will lead to more rapid shifts when learning processes are strong.

### **Punishment can evolve even when not linked to cooperation**

In Boyd et al. (2003), we presented a simulation model showing how competition among groups could enhance the evolution of costly punishment when adaptive forces are strong. The model assumed that there is a population structured into groups of size  $n$ . Each generation individuals can contribute to a collective good at a cost  $c$ . Then individuals can punish any other individual reducing their fitness an amount  $p$  at a cost  $k$ . Next, with probability  $\varepsilon$ , each group enters into a conflict with another randomly chosen group. It wins the conflict with a probability proportional to the difference in the frequency of cooperators between the two groups. Losing groups go extinct and are replaced by a clone of the winning group. Traits “mutate” with probability  $\mu$ . Then, individuals choose another randomly chosen individual and acquire their trait from a model with a probability proportional to the difference in payoffs according to the rule given in (3) above. Finally, a fraction  $m$  of models are drawn at random from the population as a whole and  $1-m$  from the individual's own group. Boyd et al. considered competition between three strategies, cooperators who contribute to the collective good but do not punish, punishers who contribute and punish noncontributors, and defectors who neither contribute nor punish. The simulations in that paper indicate that plausible amounts of intergroup conflict can maintain cooperation and punishment at high levels as long as payoff biases are strong compared to migration.

Lehmann et al. (2007) analyze a model of the evolution of cooperation and punishment that is similar to that in Boyd et al. (2003). The main differences are that in Lehmann et al. (2007) greater collective action in a group did not reduce its chances of extinction, and instead, it increased its average fitness, and the two traits were transmitted genetically, not culturally. They conclude that selection can only lead to the evolution of punishment when the locus that controls punishment is tightly linked to the locus that controls cooperation. While acknowledging that their model is not directly comparable to Boyd et al.'s,

they note that punishment and cooperation are linked in that model, and conjecture that selection would not lead to the evolution of punishment if this were not the case.

This conjecture is incorrect. We have modified the simulation used in Boyd et al (2003) so that individuals acquire the variant of the cooperation trait (contribute or defect) and the variant of the punishment trait (punish or do not punish) from two different, randomly chosen models. (This simulation was written in Visual Basic 5, and the necessary form and module files are available from the first author). This corresponds to a recombination rate equal to one, and means that in each time period immediately after transmission, the correlation due to linkage between the cooperation trait and the punishment trait *within* groups is very low. It is not exactly zero because the flow of variants among groups with different frequencies of each trait generates a low level of correlation.

When transmission biases are weak, Lehmann et al. are correct—punishment does not evolve. However, when biases are strong compared to migration between groups, punishment does evolve, and the results are qualitatively similar to the original results which assumed that punishment and cooperation were linked. To see why, consider the results shown in Fig. 5. Both simulations assume that groups consist of 128 individuals and that migration rate is 1%, and both assume that initially one group has high frequencies of punishment and cooperation and the rest have no punishment and no cooperation. The figure shows the distribution of frequencies across groups after 1,000 time periods. In (a) bias is weak ( $\beta=0.05$ ). At steady state, the frequency of cooperation is about 0.3, the frequency of punishment is close to zero, and there is no correlation across groups. This makes sense. Finite group size and limited migration lead to an equilibrium relatedness of around 0.3, in this case, due to common descent. Thus, the lower extinction rates that are generated by cooperation lead to an inclusive fitness benefit to cooperators. Punishment does not substantially increase the frequency of cooperation within groups because the transmission bias is weak, so punishment cannot create an inclusive fitness benefit, and does not increase in frequency.

In contrast, in (b), the bias is strong ( $\beta=0.5$ ). Now at steady state, the population average frequencies of punishment and cooperation are about 0.4 and 0.9, respectively, and there is a substantial positive correlation between cooperation and punishment across groups—groups with more punishers have more cooperators—even though there is no correlation within groups. In groups in which punishers are common, defectors are heavily punished, have a lower payoff than cooperators who are not punished, and thus, are less likely to be imitated than cooperators. This decreases the within group frequency of defectors in groups in which punishers are common, and as a result, the

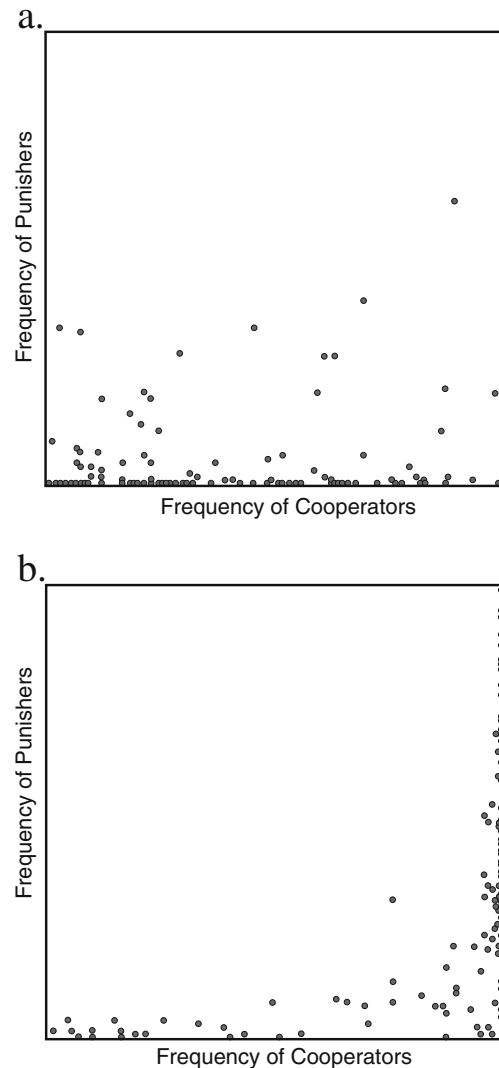


**Fig. 5** The distribution of group frequencies of cooperation and punishment when bias is weak and when it is strong. Each dot represents the frequencies of punishment and cooperation in one of 128 groups during the last period (1,000) of the simulation. In both simulations,  $c=0.2$ ,  $p=k=0.8$ ,  $m=0.01$ , and  $n=128$ . There is complete recombination each time period. Individuals acquire their punishment variant and their cooperation variant from different randomly selected models so that no linkage exists. **a** Weak forces ( $\beta=0.05$ ,  $\varepsilon=0.0015$ ,  $\mu=0.0001$ ). Relatedness (based on common descent) builds up to substantial levels, and since extinction rates are proportional to the frequency of cooperators in groups, selection increases the frequency of cooperation to a modest level (averaging about 0.3) at steady state. However, punishment and cooperation are uncorrelated both within and across groups, so punishment is selected against. It is maintained in the population by the cultural analog of mutation. **b** Strong forces ( $\beta=0.5$ ,  $\varepsilon=0.015$ ,  $\mu=0.001$ ). As in the weak bias case, relatedness builds up due to finite populations and limited migration. However, in groups with a high frequency of punishers, defectors are selected against thus maintaining a high frequency of cooperators. Complete recombination means that there is no correlation between cooperation and punishment within groups, but there is a strong correlation across groups generated by the fact that punishment lowers the payoff of defectors. Thus, the extinction of groups with few cooperators increases the frequency of punishers, and since punishment has low cost in groups in which cooperators are common, punishment is maintained at a substantial frequency (about 0.4) and cooperation at a higher frequency (about 0.9) than in the weak forces case

frequency of cooperation in such groups is higher than in groups in which punishers are rare. Groups with a low frequency of cooperators go extinct at a higher rate than those with a high frequency of cooperators, and because there is a positive correlation across groups, this means that groups with high frequency of punishers have a lower extinction rate than groups with a low frequency of punishers. Of course, punishers always have lower payoffs than nonpunishers in their group, and thus, the frequency of punishment within groups tends to decrease. However, in groups in which defectors are rare, there is little cost to being a punisher, and thus, frequency of punishment declines slowly. As long as the increase in frequency due to differential extinction is greater than the decrease within groups, punishment and cooperation are sustained at high frequencies.

### Evidence that cultural evolution is subject to strong adaptive forces

Our models and those of Lehmann et al. (2008b) are based on different assumptions about the processes that maintain cultural variation among groups, and the processes that select among groups. Their work does not refute ours. Rather, it explores an alternative hypothesis about the processes that govern the cultural evolution of large-scale cooperation in human populations in which variation among groups is maintained by common descent, not adaptation to local social conditions. Since both accounts



are cogent, we must turn to what is known about human learning, cultural variation, and cooperation.

One of the striking puzzles about human sociality is that people frequently cooperate in large groups. This is obviously true in the agricultural societies of the last 10,000 years in which thousands of individuals are mobilized for military activity and the construction of large capital facilities like roads, fortifications, and ceremonial centers. However, it is also true for small-scale human societies. For example, hunter–gatherers recruit war parties numbering in the hundreds of individuals. (See Richerson and Boyd 2005 and Henrich and Henrich 2007 for references). Thus, a successful account of human cooperation must explain how substantial cultural variation among large groups (or equivalently substantial cultural relatedness within groups) arises and is maintained.

In the Lehmann et al. models, variation among groups arises from common descent as in genetic models, thus naively, one might expect that there would be low cultural

relatedness in large groups. However, as Lehmann and colleagues (2008a, b, Lehman and Feldman 2008) point out, the disproportionately important role that prestigious individuals often play in cultural evolution (Henrich and Gil-White 2001) could lead to substantial cultural variation among larger groups and to a high degree of cultural relatedness. They illustrate this idea with the “teacher” model (Cavalli-Sforza and Feldman 1981). With probability  $t$ , each individual in the group acquires his or her cultural variant from a single focal individual, the teacher, and, with probability  $1-t$ , imitates a randomly chosen individual. This means that when groups are very large, the cultural relatedness within groups converges to  $t^2$ , a result they believe provides an explanation for observed cultural variation among large groups. Thus, the idea is that the cultural analog of reproductive skew reduces the effective population size and that as a result, common descent generates substantial variation among large groups.

This account is hard to reconcile with the observed scale of human cultural variation. In the modern world, there is substantial variation in beliefs and norms among ethnic groups and nation states that number millions of individuals. For example, Bell et al. (2009) show that a lower bound on cultural  $F_{ST}$  is more than ten times the genetic  $F_{ST}$  for neighboring nation states ( $F_{ST}$  is the fraction of total heritable variation that is among groups. For large groups, it is approximately equal to the relatedness within groups). They also show that lower bounds on the cultural  $F_{ST}$  values for four large East African ethnolinguistic units are quite high, even when ecological variation is controlled for. It is not plausible that four million Kamba (one of the East African groups) share a language and many beliefs because a substantial fraction of them acquired their beliefs by imitating a small number of people. Consistent with our model, the traits being measured by Bell et al. are norms that are plausibly subject to the analog of disruptive selection; other kinds of traits might have lower  $F_{ST}$  values.

Nor is this account plausible for smaller scale societies because the scale of cultural variation in such societies is typically much larger than the scale of everyday interaction. For example, among Australian aboriginal foragers, ethnolinguistic units that shared a common language and culture typically numbered between 500 and 5,000 (Keen 2004), and migration rates between ethnic groups were probably substantial. If we assume that bands numbered between 10 and 100 people, and that everybody in a band imitates a single individual, then the formula used by Lehman and colleagues predict that only a small fraction of cultural variation will be between ethnolinguistic units.

Moreover, empirical evidence indicates that while some individuals are more important in cultural transmission than others, probabilities of common descent are too small to lead to substantial relatedness in very large groups.

Assuming discrete traits, and accurate social learning, the probability that two individuals acquire the same variant by common descent is  $\sum a_i^2$  where  $a_i$  is the probability that the  $i$ th individual in the group is imitated by a learner in the next generation. In the limit of very large groups,  $\sum a_i^2$  is the relatedness among group members, a generalization of the teacher model. Henrich and Broesch (2010) have estimated these  $a_i$  values in a small Fijian village with 210 residents. In 2003, a sample of 146 subjects were asked to indicate which other individuals they would go to if they were seeking information in three different domains: knowledge about (1) fishing, (2) yam horticulture, and (3) medicinal plants. Assuming that this is a measure of the importance of individuals in cultural transmission, these data can be used to estimate  $\sum a_i^2$ . The values are 0.043 for fishing, 0.053 for yam horticulture, and 0.053 for medicinal plants. Five years later, the protocol was repeated, this time obtaining 0.040 for fishing, 0.043 for yam horticulture, and 0.034 for medicinal plants. These data indicate that some individuals have more influence than others, but the probability that two individuals acquire their beliefs by common descent is still fairly small. For comparison, the average genetic relatedness computed from a complete three generation genealogy taken in 2003 from this group is 0.018.

Nor can the teacher model explain the persistence of differences between large neighboring groups over hundreds of generations unless it is assumed that migration rates are unrealistically low. For example, the Romance/Germanic linguistic boundary is roughly where the Roman advance came to rest two millennia ago despite massive flows of people across the boundary. This boundary also separates peoples with different norms that lead to measurably different behavior in important economic contexts (Brügger et al. 2009). Lehmann et al.’s account would seem to require that Germans keep speaking German and keep adhering to German social norms and the French do the same because there is a significant probability that they acquire their linguistic and social norms from a small number of people. We believe that it is more plausible that strong biased transmission maintains cultural boundaries (McElreath et al. 2003; Boyd and Richerson 1987). When people move from one culture to another, they, and especially their children, modify their language and social behavior in response to local linguistic and social norms, so that they will be understood, and approved of. If this process is sufficiently rapid compared to the rate of migration, the boundary will be maintained.

## Conclusion

Our models and those of Lehmann and colleagues lead to very different pictures of the cultural evolution of human

cooperation. In the Lehmann et al. (2008b) world, groups have different norms and different beliefs and values affecting cooperation. These differences arise and are maintained by chance events determining which individuals happen to be imitated each generation. The incremental effects of alternative culturally transmitted ideas and practices on welfare or fitness are small, and, as a result, ideas and practices move easily from one group to another. Variation among groups is necessary for the evolution of norms that are group beneficial but costly when rare. However, if their effect averaged over all groups is positive, norms evolve throughout the population at approximately the same rate. In our picture, groups have different norms that affect cooperation. Individuals who adhere to the norms that are common in their group achieve higher payoffs than individuals who espouse different norms, and as a result, beliefs and values that make a person likely to be imitated in one group may have the opposite effect in other groups. This, in turn, means that beliefs and values do not move easily from group to group, and this acts to maintain variation among groups, even very large groups. Group beneficial ideas spread because groups in which those ideas are common replace groups in which they are not, or because they have a big enough effect on neighboring groups that these groups shift to the norms of their successful neighbors.

Whether either of these models captures the processes that have led to the evolution of human cooperation is an empirical question, both models are cogent; the question is: does either fit the data? For cultural variation on the scale of ethnolinguistic groups numbering thousands of individuals, we think that the answer is that our model fits better than that of Lehmann et al. for the reasons discussed above. However, it is possible that the Lehmann et al. models will be useful for understanding the cultural evolution of cooperation on smaller scales or during earlier periods of human evolution prior to the emergence of self-enforcing norms and our current sophisticated forms of cultural learning. Variation among bands or villages within an ethnic group could be due to sampling.

Human cultural evolution is usefully conceptualized as a population process, and as a result, theoretical tools from population biology can be very helpful in understanding cultural evolution. However, it is important to resist the temptation to think that cultural transmission is just like genetic transmission. The theory of the genetic evolution of social behavior is highly developed, replete with subtle, powerful insights, and well-worked out mathematical tools. In genetic evolution, selection is often weak enough that relatedness through common descent is sufficient to predict patterns of social interaction, providing a powerful tool for understanding the evolution of social behavior. The evidence suggests that cultural variation is affected by

strong biased transmission, and limited diffusion. If so, different tools may be necessary to understand the cultural evolution of group beneficial norms in large groups.

**Conflict of interest** The authors declare that they have no conflict of interest regarding the materials presented in this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Alison P (1992) Cultural relatedness under oblique and horizontal transmission rules. *Ethol Sociobiol* 13:153–169
- Barton N (1979) The dynamics of hybrid zones. *Heredity* 43:341–359
- Bell AV, Richerson PJ, McElreath R (2009) Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc Natl Acad Sci* 106:17671–17674
- Boehm C (1999) Hierarchy in the forest: the evolution of egalitarian behavior. Harvard University Press, Cambridge
- Boorman S, Levitt P (1973) A frequency-dependent natural selection model for the evolution of social cooperation networks. *Proc Natl Acad Sci* 70:187–189
- Boorman S, Levitt P (1980) The genetics of altruism. Academic, NY
- Bowles S (2009) Did warfare among ancestral hunter–gatherer groups affect the evolution of human social behaviors. *Science* 324:1293–1298
- Boyd R, Gintis H, Bowles S, Richerson PJ (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100:3531–3535
- Boyd R, Richerson PJ (1985) Culture and the evolutionary process. University of Chicago Press, Chicago
- Boyd R, Richerson PJ (1987) The evolution of ethnic markers. *Cult Anthropol* 2:65–79
- Boyd R, Richerson PJ (1990) Group selection among alternative evolutionarily stable strategies. *J Theor Biol* 145:331–342
- Boyd R, Richerson PJ (2002) Group beneficial norms spread rapidly in a structured population. *J Theor Biol* 215:287–296
- Boyd R, Richerson PJ (2009) Culture and the evolution of cooperation. *Philos Trans R Soc B* 364:3281–3288
- Brügger B, Lalive R, Zweimüller J (2009) Does culture affect unemployment? Evidence from the *Röstigraben*, Institute for the Study of Labor, University of Zurich, Discussion Paper No. 4283
- Cavalli-Sforza LL, Feldman MW (1981) Cultural transmission and evolution. Princeton University Press, Princeton
- Claidiere N, Sperber D (2007) The role of attraction in cultural evolution. *J Cogn Cult* 7:89–111
- Clutton-Brock TH (2009) Cooperation between non-kin in animal societies. *Nature* 462:51–57
- Darwin C (1874) The descent of man and selection in relation to sex, 2nd edition, John Murray, London.
- Dower J (2000) Embracing defeat: Japan in the wake of World War II. W.W. Norton, Co, NY
- Efferson C, Lalive R, Richerson PJ, McElreath R, Lubell M (2008) Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evol Hum Behav* 29:56–64
- Gavrilets S (1996) On phase three of Wright's shifting balance theory. *Evolution* 50:1034–1041

- Gillespie JH (2000) Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155:909–919
- Gillespie JH (2001) Is the population size of a species relevant to its evolution? *Evolution* 55:2161–2169
- Griffiths TL, Kalish ML, Lewandowsky S (2008) Theoretical and experimental evidence for the impact of inductive biases on cultural evolution. *Philos Trans R Soc* 363:3503–3514
- Guzmán A, Rodríguez-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
- Henrich J (2001) Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change and much of sociocultural evolution. *Am Anthropol* 103:992–101
- Henrich J (2004) Cultural group selection, coevolutionary processes and large-scale cooperation. *J Econ Behav Organ* 53:3–35
- Henrich J (2006) Cooperation, punishment, and the evolution of human institutions. *Science* 312:60–61
- Henrich J (2009) The evolution of costly displays, cooperation, and religion: credibility enhancing displays and their implications for cultural evolution. *Evol Hum Behav* 30:244–260
- Henrich J, Boyd R (1998) The evolution of conformist transmission and the emergence of between-group differences. *Evol Hum Behav* 19:215–242
- 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 J, Boyd R (2002) Culture and cognition: why cultural evolution does not require replication of representations. *Cult Cogn* 2:87–112
- Henrich J, Broesch J (2010) On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philos Trans R Soc B* (in press)
- Henrich J, Gil-White F (2001) The evolution of prestige: freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evol Hum Behav* 22:1–32
- Henrich J, Henrich N (2007) Why humans cooperate. Oxford University Press, Oxford
- Hill K (2002) Altruistic cooperation during foraging by the Aché, and the evolved predisposition to cooperate. *Hum Nat* 13:1045–1067
- Keeley L (1997) War before civilization. Oxford University Press, Oxford
- Keen I (2004) Aboriginal economy and society: Australia at the threshold of colonisation. Oxford University Press, Oxford
- Kendal J, Giraldeau L, Laland K (2009) The evolution of social learning rules: payoff-biased and frequency-dependent biased transmission. *J Theor Biol* 260:210–219
- Lehmann L, Feldman MW (2008) The coevolution of culturally inherited altruistic helping and cultural transmission under random group formation. *Theor Popul Biol* 73:506–516
- Lehmann L, Feldman MW, Foster K (2008a) Cultural transmission can inhibit the evolution of altruistic helping. *Am Nat* 172:12–24
- Lehmann L, Foster K, Borenstein E, Feldman MW (2008b) Social and individual learning of helping in humans and other species. *Trends Ecol Evol* 23:664–671
- 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* 170:21–36
- Marshall J, McNamara J, Houston A (2010) The state of Darwinian theory. *Behav Ecol Sociobiol*. doi:10.1007/s00265-010-1100-3
- McElreath R, Bell AV, Efferson C, Lubell M, Richerson PJ, Waring T (2008) Beyond existence and aiming outside the laboratory: estimating frequency-dependent and payoff-biased social learning strategies. *Philos Trans R Soc B* 363:3515–3528
- McElreath R, Boyd R, Richerson PJ (2003) Shared norms can lead to the evolution of ethnic markers. *Curr Anthropol* 44:122–130
- Mesoudi A (2009) How cultural evolutionary theory can inform social psychology and vice versa. *Psychol Rev* 116:929–952
- Nakahashi W (2007) The evolution of conformist transmission in social learning when the environment changes periodically. *Theor Popul Biol* 72:52–66
- Neiman FD (1995) Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *Am Antiquity* 60:7–36
- Richerson PJ, Boyd R (2005) Not by genes alone: how culture transformed human evolution. University of Chicago Press, Chicago
- Ridley M (2010) The rational optimist: how prosperity evolves. Harper, NY
- Shennan S (2001) Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Camb Archaeol J* 11:5–16
- Soltis J, Boyd R, Richerson PJ (1995) Can group functional behaviors evolve by cultural group selection? An empirical test. *Curr Anthropol* 36:473–494
- Taylor PD (1992) Inclusive fitness in a homogeneous environment. *Proc R Soc B* 249:299–302
- Wakano JH, Aoki K (2007) Do social learning and conformist bias coevolve? Henrich and Boyd revisited. *Theor Popul Biol* 72:504–512