Cultural group selection, coevolutionary processes and large-scale cooperation

Joseph Henrich*

Department of Anthropology, Emory University, Geosciences Building, Atlanta, GA 30322-1720, USA

Abstract

In constructing improved models of human behavior, both experimental and behavioral economists have increasingly turned to evolutionary theory for insights into human psychology and preferences. Unfortunately, the existing genetic evolutionary approaches can explain neither the degree of prosociality (altruism and altruistic punishment) observed in humans, nor the patterns of variation in these behaviors across different behavioral domains and social groups. Ongoing misunderstandings about why certain models work, what they predict, and what the place is of "group selection" in evolutionary theory have hampered the use of insights from biology and anthropology. This paper clarifies some of these issues and proposes an approach to the evolution of prosociality rooted in the interaction between cultural and genetic transmission. I explain how, in contrast to non-cultural species, the details of our evolved cultural learning capacities (e.g., imitative abilities) create the conditions for the cultural evolution of prosociality. By producing multiple behavioral equilibria, including group-beneficial equilibria, cultural evolution endogenously generates a mechanism of equilibrium selection that can favor prosociality. Finally, in the novel social environments left in the wake of these cultural evolutionary processes, natural selection is likely to favor prosocial genes that would not be expected in a purely genetic approach.
© 2003 Published by Elsevier B.V.

1. Introduction

As the experimental evidence for non-selfish preferences continues to mount in empirical work (Fehr and Gächter, 2000; Henrich et al., 2001; Bolton et al., 1998; Kagel and Roth,

* Tel.: +1-404-727-5248; fax: +1-404-727-2860.
E-mail address: jhenric@emory.edu (J. Henrich).

1995), an increasing number of economists are turning to evolutionary biology and biological anthropology in search of plausible theoretical foundations (e.g. McCabe et al., 2000; Fehr and Gächter, 2002; Fehr et al., 2002; Gintis et al., 2001; Gintis, 2000; Ben-Ner and Putterman, 2000). Of particular interest to economic theorists is the evolution of prosocial preferences (e.g., tastes for altruism and altruistic punishment) that are capable of explaining cooperation and fairness in large groups and among unrelated strangers in non-repeated contexts. Unfortunately, the interdisciplinary nature and complexity of the literature on cooperation and altruism has led to some confusion about: (1) what constitutes an evolutionary 'solution' to the puzzle of prosociality, and (2) the process and potential importance of 'group selection'. In this paper, I will first layout a general framework for understanding evolutionary approaches to prosociality. Then, using this framework, I will critique and clarify some of the most common approaches to show how they fail to explain the prosocial preferences that promote large-scale cooperation among non-relatives without repeated interaction. Building on this foundation, the second half of the paper presents an alternative route to explain prosociality via culture–gene coevolution that keys on three things: (1) the difference between cultural transmission and genetic transmission, (2) the differences between genetic group selection and cultural group selection, and (3) the evolutionary interaction between genes and culture.

2. The underlying structure of prosociality

All genetic evolutionary explanations to the altruism dilemma are successful to the degree that they allow natural selection to operate on statistically reliable patterns or regularities in the environment. All too often, the assumptions or constraints that maintain this regularity are submerged in the basic structure or setup of the model—and not explicitly analyzed. To expose the required regularity, I develop a very simple model of cooperation. We begin with a population of N individuals indexed by i and subdivided into a number of smaller groups labeled with j. Each individual possesses either an altruistic gene or an egoistic gene tracked by x_i. When $x_i = 1$ individual i possesses the altruistic allele (version of the gene), and $x_i = 0$ marks i as possessing the egoistic allele. To derive the conditions for the spread of an altruistic gene, we will use the famous Price equation (Price, 1970, 1972):

$$\bar{w}\Delta\bar{x} = \text{Cov}(w_i, x_i) + E(w_i\Delta x_i) \tag{1}$$

The Price equation is a simple statistical statement that relates the expected change in the frequency of a gene ($\Delta\bar{x}$) per generation, the absolute fitness (w_i: the number of offspring of i), the average fitness of the population ($\bar{w}$), and the current frequency of the gene (x_i). The $\text{Cov}(w_i, x_i)$ term gives the covariance between absolute fitness and gene frequency, and $E(w_i\Delta x_i)$ represents the expected value of the within-individual change in x_i (e.g., mutation) weighted by the individual's absolute fitness. Following Hamilton (1975), I am using the non-standard notational convention of leaving the subscripts on the variables inside of the covariance and expectation operators. This practice is required to maintain notational clarity when things get more complicated below.[1] Eq. (1) provides a very general statement

---

[1] For example, using a more standard notational convention, one could write $\text{Cov}(w, x)$ in lieu of $\text{Cov}(w_i, x_i)$.

about any evolutionary system. Later, I will derive and extend its application to cultural evolution and structured populations.

In this model, altruists bestow benefits on other members of their local group j at a cost to themselves. Egoists do not bestow benefits on others, but both egoistic and altruistic group members will benefit from being in groups with many altruists. To investigate this, I specify w_i as follows:

$$w_i = \alpha + \beta_{w_i x_i \cdot x_j} x_i + \beta_{w_i x_j \cdot x_i} x_j + \varepsilon_i \tag{2}$$

Here i's fitness is jointly determined by the effect of her genotype on her fitness, holding her local group composition constant, $\beta_{w_i x_i \cdot x_j} x_i$ (x_j is the average value of x_i in group j), and, the effect of i's local group on her fitness holding the individual's genotype constant $\beta_{w_i x_j \cdot x_i} x_j$. The constant baseline fitness is α and ε_i is the uncorrelated error term.

With Eqs. (1) and (2) we can derive a general statement about the conditions for the genetic evolution of altruism. Because we are interested only in the effects of natural selection, I will ignore non-selective forces such as recombination, drift, meiotic drive and mutation. This means $\Delta x_i = 0$, so the expectation term drops out of (1). Substituting (2) into (1) yields

$$\bar{w}\Delta\bar{x} = \mathrm{Cov}(\alpha, x_i) + \beta_{w_i x_i} \mathrm{Var}(x) + \beta_{w_i x_j}\beta_{x_j x_i} \mathrm{Var}(x) + \mathrm{Cov}(\varepsilon_i, x_i) \tag{3}$$

By definition $\mathrm{Cov}(\alpha, x_i)$ and $\mathrm{Cov}(\varepsilon_i, x_i)$ are zero. Setting $\Delta\bar{x} > 0$ gives the conditions for an increase in the frequency of the altruistic gene in the population (assuming both alleles exist in the population):

$$\beta_{w_i x_i} + \beta_{w_i x_j}\beta_{x_j x_i} > 0 \tag{4}$$

Eq. (4) shows that all solutions to the evolution of altruism—whether they are based on kinship, reciprocity or group selection (more on those later)—are successful according to the degree in which 'being an altruist' predicts that one's partners or group members are also altruistic (Frank, 1998). By definition $\beta_{w_i x_i}$ is negative because, ceteris paribus, altruists have lower fitness than egoists; $\beta_{w_i x_j}$ is always positive because, independent of one's own genotype, it is always better to be in a group with more altruists. This leaves $\beta_{x_j x_i}$, which captures the degree to which 'being an altruist' predicts 'being in an altruistic group'. If groups are randomly remixed every generation, $\beta_{x_j x_i} = 0$, and altruism (even kin-based altruism) cannot evolve. However, if (for whatever reason) altruists can preferentially group with other altruists, $\beta_{x_j x_i}$ will be a positive number between zero and one, and altruism at least has a chance. If egoists can preferentially enter groups with altruists, $\beta_{x_j x_i}$ will be negative. Thus, altruistic genes can only spread if $\beta_{x_j x_i}$ is sufficiently close to one such that the benefits to being in an altruistic group outweigh the costs of bestowing benefits on others. Those with any familiarity with evolutionary theory may recognize (4) as a generalized version of "Hamilton's Rule" (Queller, 1992b). Expressing this in the standard notation, $\beta_{w_i x_j}$ is B, the fitness benefits provided by altruists; $\beta_{w_i x_i}$ is $-C$, the cost to the altruist; and $\beta_{x_j x_i}$ is r, which in the specific case of kin selection is interpreted as the coefficient of relatedness by descent from a recent ancestor. In general, however, $\beta_{x_j x_i}$ is *not* a measure of relatedness by descent from a recent common ancestor—although, as I will explain, that is *one way* to get a positive value of $\beta_{x_j x_i}$. Hereafter, for simplicity, I will refer to $\beta_{x_j x_i}$ as β.

Most importantly, Eq. (4) shows that understanding particular solutions to the evolution of altruism requires analyzing how the model maintains a sufficiently high value of β. Thus, the acceptability or legitimacy of a theoretical solution depends on an evaluation of the constraints (or "special assumptions") that give rise to, and maintain, the statistical association (β). Remember, however, that the greater the value of β, the greater the amount of altruism that can evolve and the greater the selective pressures for mutant genes that can 'beat the system' by exploiting β. Under different circumstances, in different species, different kinds of constraints or special assumptions may be warranted. The details of an organism's social structure, physiology, genome, cognitive abilities, migration patterns or imitative abilities may support some hypotheses and undermine others. Of course, by emphasizing the species-specificity of certain constraints, I do not mean to suggest that each constraint is equally likely to be observed in a randomly selected species. Surely, some constraints are more frequently satisfied in nature than others; nevertheless, the rarely satisfied constraints may provide the most interesting forms of altruism. Unfortunately, despite the need to defend a particular constraint vis-à-vis the details of that species that might justify focusing on that constraint over others, many students of evolution and human behavior regard only kinship- and reciprocity-based models of altruism as legitimate solutions. In contrast, approaches based on quite plausible constraints that may be satisfied only in the human case, such as those based on cultural group selection and culture–gene interaction, are neither widely considered nor understood by non-specialists. However, before dealing with alternative approaches, I will illustrate what I mean by 'constraints', how they facilitate solutions to the altruism dilemma, how these are often submerged in evolutionary models, and why decisions about what constitutes an "acceptable solution" are often arbitrary.

3. Constraints and evolutionary models of altruism

3.1. Green beards

I will begin with the classic example of a solution to the altruism dilemma based on an 'illegitimate' constraint: the 'green beard solution' (Dawkins, 1976). Imagine a gene that causes its bearer to have a green beard and to cooperate with other green-bearded individuals. Because green-bearded cooperators focus their benefits only on other green-bearded cooperators, natural selection will strongly favor this altruistic gene because $\beta = 1$ (unless they are so rare that fellow green beards never find one another). More generally, if altruists can recognize other altruists (with a better than random chance, $\beta > 0$) and preferentially bestow benefits on them, then some amount of altruism can be favored by natural selection. In the situation, the statistical regularity (β) exploited by natural selection is the correlation between green beards (the signal) and altruism. The problem with this solution is that the informational regularity arises from a genetic (or physiological) constraint on the power of mutation to create selfish egoists with green beards (green-bearded defectors). Such a non-altruistic gene will exploit the green beard-cooperator regularity (β) and destroy it in the process. Egoists with green beards will invade until β is driven to 0. Consequently, this model works only if we allow a mutational constraint.

Is such a constraint acceptable? Popular writers on evolution often scoff at the absurdity of such a constraint (Low, 2000; Dawkins, 1976, 1982; Ridley, 1993), despite the fact that the available empirical evidence is somewhat mixed. Keller et al. (1998) have shown empirically that an odor-based signal explains large-scale cooperation in the red fire ant. In humans, experimental work suggests that people have some ability to distinguish cooperators from defectors (Frank et al., 1993; Brown and Moore, 2002), although other work says they cannot (Ockenfels and Selten, 2000; Ekman, 1992; Ekman and O'Sullivan, 1991; Henrich and Smith, in press). In either case however, what should be unacceptable is the common practice of simply assuming that individuals know the strategies of other individuals and assort according, without providing any justification (theoretical or empirical) as to why this should be so (e.g., Wilson and Dugatkin, 1997). This is tantamount to assuming the answer.

The green beard problem can be addressed as a problem in costly signaling (Frank, 1988; Gintis et al., 2001). If it costs less for cooperators to cue their strategy than for non-cooperators to falsely cue, then cooperation can be favored. This occurs because such signals allow the benefits of cooperation to be preferentially delivered to other cooperators. From the perspective of Eq. (4), cooperation can evolve because this signaling allows for the non-random formation of groups such that cooperators find themselves grouped with other cooperators. Egoists cannot fake signals because it is too costly for them, so they find themselves with other defectors.

There are four problems with the costly signaling approach to green beard altruism. First, it relies on the constraint that it is cheaper for cooperators to send the cues of cooperation than for defectors. Mutations that allow defectors to send these signals for the same cost as altruists cannot be permitted—this is a key constraint that makes the whole thing work. Second, as noted above, empirical evidence for an ability to distinguish 'cooperators' from 'defectors' without extensive interaction is contradictory and ambiguous. The ability of good actors (e.g., in the movies) to evoke powerful emotions in us by pretending to be either a valiant 'cooperator' or a malevolent 'defector' suggests there is some variation in the ability to send false signals (so, why are not we all good actors?). Third, we need to explain cooperation and trust among anonymous people (who cannot see one another) in one-shot encounters such as we observe in experimental economic games (Kagel and Roth, 1995; Davis and Holt, 1993; Fehr et al., 2002). In most of these situations, there is no chance for signaling of any kind. So, at best, this approach solves only one kind of altruism problem. Fourth, if such costly signaling explains the degree and kind of cooperation in humans, why has not this mechanism generated much more cooperation in non-human animals, such as chimpanzees, elephants and dolphins?

3.2. Kin-based selection

Kin-selection models actually make use of a variety of different constraints. In some kin-based approaches, the nature of a species' social (or family) structure produces a whole host of possible statistical associations that can be exploited by natural selection. For example, if sisters tend to hang around their mom, genes that cause their bearer to cooperate with others 'hanging around mom' may spread because—if they are really your sisters— they have at least a 50 percent chance of having the same altruistic gene as you. 'Hanging

around mom' merely provides one possible predictor of kinship, and indirectly, of altruism. Consequently, social structures that require individuals to interact with relatives more frequently than non-relatives provide a potentially valuable informational regularity. If 80 percent of those receiving benefits from an altruist are full siblings of the altruist and the rest are non-relatives, then β is approximately 0.40 (assuming the frequency of altruists in the rest of the population is very small).

This social structure constraint is only exploitable if, by whatever means, the underlying non-random association can be maintained. If non-altruistic mutants, posing as fake-family members, can frequently slip into nests or family groupings and obtain the benefits, then kinship solutions will suffer the same problem as green beard models. Among animals lacking the requisite social structure (e.g., some snake species), we should not expect the corresponding forms of altruism (kin or otherwise). Furthermore, the social structure—that often creates the requisite informational regularity—is usually taken as a given, rather than being a product of evolution with its own costs and benefits. Models never endogenize the costs of maintaining a particular family or social structure in analyzing the evolution and maintenance of altruism.

Other kin-based solutions are founded on kin-recognition mechanisms (Fletcher and Michener, 1987) that allow individuals to probabilistically spot their kin and direct altruism towards them. Qualitatively, this solution is the same as the green beard solution, except that having many of the same genes, being reared together, and the potential existence of other evolutionary pressures favoring kin recognition (e.g., inbreeding avoidance) makes the explanation much more plausible—though not fundamentally different—than unjustified green beards. Combinations of kin-cues may make it quite costly for natural selection to produce faker-defectors.

Finally, the rules of inheritance that make kinship a valuable predictor of altruism, are often taken as given despite the fact that certain alleles can violate these rules (e.g., meiotic drive and gametic selection). A non-altruistic gene that is transmitted from parent to offspring with a greater than 50 percent chance (contrary to Mendelian rules) could spread and gradually corrode the informational regularity upon which some forms of kin altruism are based.[2]

Kin-based mechanisms should be designed to focus benefits only on close relatives, and thus kin selection does not help us to solve the problem of cooperation among large groups of unrelated individuals unless our kin-psychology is making a lot of big mistakes by confusing large numbers of non-relatives and strangers with close relatives. This version of the "big mistake hypothesis" (Boyd and Richerson, 2002a) suggests that, because our psychology supposedly evolved in small groups with high degrees of interrelatedness, kin selection

---

[2] Meiotic drive and gametic selection distort (relative to Mendelian rules) the frequencies of different alleles available for recombination. Using a meiotic drive example, imagine two alleles: a C-allele (cooperator) and a D-allele (defector). Diploid individuals who are CC or DD produce all C-gametes and all D-gametes, respectively. When paired with C-alleles, however, D alleles use "chemical warfare" such that CD individuals produce, on-average, a proportion $1 - k$ of C-gametes and k of D-gametes ($k > 0.5$, under Mendelian rules $k = 0.5$). This means that if I am CD and we are brothers, I would be mistaken to assume that you have a 50 percent chance of possessing a copy of my C-allele. Because of meiotic drive we would be less related vis-à-vis the relevant allele than brothers would normally be for other alleles (those not on the Y-chromosome). In mice, the t-allele is transmitted over 90 percent of the time to offspring ($k = 0.90$).

(along with reciprocity, see next section) favored a psychology in humans that is designed to generously bestow benefits on members of their groups. According to this idea, natural selection apparently neglected to provide humans with the ability to distinguish kin and long-term reciprocators from anonymous strangers in ephemeral interactions. Thus, in the novel world of large-scale, complex societies, this once adaptive psychological propensity misfires, giving us large-scale cooperation (Tooby and Cosmides, 1989). Using this line of reasoning, a variety of scholars have attempted to explain why we observe altruism, trust, cooperation, fairness and punishment in one-shot experimental games (Hoffman et al., 1998; Ben-Ner and Putterman, 2000).

This explanation has a number of problems. First, even in small-scale societies, there are plenty of distant relatives, and probably a fair number of strangers, that altruists need to distinguish from close kin—β (relatedness by descent for alleles) decreases geometrically as the circle of kin expands from sibs, to half-siblings, to first cousins, and so on. The idea that individuals in small-scale societies did not have ephemeral encounters with anonymous others is empirically groundless and represents a popular anthropological myth spread by those unfamiliar with the details of the ethnographic and ethno-historical record. Second, although large-scale cooperation is prevalent in many societies, people everywhere favor their kin over non-kin—showing that we can, and do, distinguish these behaviorally (Daly and Wilson, 1988; Sepher, 1983; Wolf, 1970; Westermarck, 1894). Third, lots of non-human primates also live in small-scale societies but show no generalized tendency to cooperate with all members of their group. When non-human primates are placed in larger, evolutionarily novel groups (e.g., zoos and research centers), cooperation does not expand to the enlarged group. It is hard to believe that non-human primates do not make these big mistakes, but humans do. Boyd and Richerson (2002a) further summarizes the evidence against this hypothesis.

3.3. Reciprocity

Reciprocity provides another well-studied solution to the altruism dilemma. In reciprocity-based models with repeated interactions (usually modeled as a repeated prisoner's dilemma), it has been frequently shown that strategies that both bestow benefits on individuals who have bestowed benefits on them in the past and withhold future benefits from strategies that fail to reciprocate can be favored by selection if they are sufficiently common (Axelrod, 1984; Axelrod and Hamilton, 1981; Trivers, 1971). Repeated interaction using these kinds of reciprocating strategies produces the requisite informational regularity. Repeated interactions among the same individuals allows past behavior to guide the flow of benefits preferentially to other reciprocators (β), and more repetitions allow groups of reciprocators to increase their fitness benefits relative to groups of non-reciprocators ($\beta_{w_i x_j}$).

However, making such models work requires a couple of constraints that often go unstated. First, most models analyze only two strategies at a time (e.g., Axelrod, 1984) and do not consider the presence of additional strategies (even one additional strategy) maintained at low frequency by mutation, immigration, or non-heritable environmental variation. It turns out that if the full range of other strategies is occasionally allowed to mutate into existence at low frequencies, then *no strategy* (pure or mixed), such as "tit-for-tat" (TFT)

and "always defect" (ALLD),[3] is evolutionarily stable (Boyd and Lorderbaum, 1987; Lorberbaum, 1994). For example, if the game is repeated a sufficient number of times, tit-for-two-tats (TF2T) can invade a population of mostly TFT if "suspicious tit-for-tat" (STFT) is present at low frequency. TF2T is like TFT but requires two consecutive defections before defecting, while STFT defects on the first round and then plays TFT. Similarly, TF2T can be maintained at high frequency with low frequencies of STFT and TFT, unless ALLD enters the fray at low frequency, in which case it is possible for STFT to invade and become common. Notably, in a population of mostly STFT, we will observe mostly defection. The point here is that whether or not reciprocating strategies can remain stable depends on exactly how the mutational spectrum is restricted (errors can take the place of mutations, but then one needs to know the probability distribution of errors). Curiously, proponents of reciprocity-based altruism criticize green beard explanations because they require restricting the mutational spectrum (e.g. Dawkins, 1976).[4]

Second, it is also important to realize that most theoretical models of reciprocity explore only 2-person interactions. Despite this, many scholars have falsely assumed that the qualitative aspects of the 2-person results can be generalized to *n*-person situations (e.g., Patton, 2000). However, Boyd and Richerson's (1988) analysis of an *n*-person repeated prisoner's dilemma shows that the results do not generalize for groups larger than about 10 individuals. They show that the threshold frequency of reciprocators necessary to maintain stable cooperation in an *n*-person group increases with group size by the power $1/n$. They further show that combining kinship (relatedness by common descent) with reciprocity does not substantially improve the results, as it does in the 2-person case. Reciprocity, on its own, is unlikely to solve the problem of cooperation in large groups (also see Bendor and Mookherjee, 1987; Joshi, 1987).

The final implicit constraint in reciprocity models lies in the fact that individuals must be able to accurately recognize their partner(s) in repeated games. If there are mutants capable of posing as the individual(s) who bestowed benefits in the last turn, then the benefits of reciprocity evaporate (as $\beta \rightarrow 0$). Some might think that although this may be a problem for cooperation in cognitively limited animals, it is not a serious problem for big-brained humans. Perhaps it is simply too costly for natural selection to produce doppelgangers capable of fooling human cognition—a cognition presumably maintained for other evolutionary reasons (which puts the costs of that cognition outside the problem). However, to the contrary, this is a serious problem for the kind and degree of cooperation we observe in many societies. Consider the effects of increasing the group size from 2 to 2000 or of increasing the delay between the time you receive benefits and the time when your next chance to bestow them occurs. In large-scale societies, reciprocators would potentially need to keep track of the acts of hundreds or thousands of individuals for years. Further, greater cognitive capacities seem less likely to promote cooperation via reciprocity than to promote defection by producing more sophisticated deception and strategies of exploitation.

---

[3] TFT cooperates on the first round and then plays whatever its partner played on the previous turn. ALLD always defects on every turn.

[4] If pure strategies sometimes make mistakes by cooperating when their strategy indicated defection, or defecting when they wanted to cooperate, some pure strategies can be evolutionarily stable. Stable pure strategies include both reciprocating and non-reciprocating strategies, but not TFT (Boyd, 1989; Lorberbaum et al., 2002). Interested readers should also see Hirshleifer and Martinez (1988).

As noted above, some scholars have argued that the high levels of prosocial behavior that we observe in contemporary societies represent the "misfiring" of an evolved psychology geared for reciprocity in the small-scale societies of our evolutionary history (Alexander, 1974, 1987; Hamilton, 1975; Tooby and Cosmides, 1989). In large-scale complex societies, this psychology of reciprocity causes us to mistakenly cooperate when we should defect. This explanation is not plausible and suffers the same problems as the kin-selection version of the "big mistake hypothesis" discussed above.

3.4. Indirect reciprocity

By assuming that individuals know something of their partner's previous behavior with other individuals (their reputation or ranking), models of indirect reciprocity attempt to explain generalized cooperation by expanding on the idea of reciprocity (Alexander, 1987). Summarizing results from several different modeling approaches suggests that such a setup only favors cooperation when: (1) populations are small, (2) errors in perception of past behavior are rare, and (3) the number of repeated rounds with the same group are many (Leimar and Hammerstein, 2001; Nowak and Sigmund, 1998; Boyd and Richerson, 1989).[5] Each of these can be understood vis-à-vis Eq. (4). Accurate perceptions and/or small populations guarantee that altruists can use past behavior to preferentially interact with other prosocial strategies, providing a large β. Repeated rounds with the same group mean the benefits experienced by groups with more altruists will be relatively larger than those experienced by groups with fewer altruists (yielding a larger $\beta_{w_i x_j}$). Of course, if a mutant defecting-strategy can generate an inflated reputational signal, perhaps by paying other individuals to lie about the mutant's cooperative tendencies, then the predictive value of reputation will corrode and cooperation will collapse. The only way to prevent the entry of such strategies into the mix is to restrict them by fiat. Such a constraint in the mutational spectrum is exactly what popular writers lampoon when they criticize the green beard model.

Nevertheless, even if we accept this constraint, indirect reciprocity still fails to explain large-scale cooperation and altruism in one-shot, anonymous situations. The amount of cooperation supported by indirect reciprocity declines exponentially with increasing group size (Nowak and Sigmund, 1998; Leimar and Hammerstein, 2001),[6] so large-scale cooperation is not explained. And, if the reputational information is removed from the model or made less inaccurate, indirect reciprocity can only sustain cooperation in very small groups (Boyd and Richerson, 1989). While the empirical evidence indicates that a reputation for being prosocial makes one more likely to receive benefits (Wedekind and Milinski, 2000), it also shows that people behave prosocially in both experimental and real settings without any reputational information, even in large groups (Fehr et al., 2002; Ledyard, 1995; Henrich and Smith, in press). But, as with direct reciprocity and kinship, some evolutionary researchers maintain the "big mistake" hypothesis for indirect reciprocity (Nowak and

---

[5] Nowak and Sigmund's (1998) conclusions about indirect reciprocity on the evolution of cooperation were shown to be overstated and their modeling techniques incomplete by Leimar and Hammerstein (2001).

[6] This claim may not be clear from Nowak and Sigmund (1998), but see the caption of Fig. 3. The authors write, "the averages over time of the frequency of cooperative strategies (defined by $k \leq 0$) are 90, 47, and 18 percent for, respectively, $n = 20$, 50 and 100" (Nowak and Sigmund, 1998, p. 574).

Sigmund, 1998). This suggests that humans cannot tell the difference between small groups of repeated interactants and large groups of strangers—a claim contradicted by experimental evidence (Fehr and Henrich, forthcoming).

3.5. Punishment

Many theorists have attempted to solve the problem of cooperation in large groups by incorporating punishment (e.g., McAdams, 1998; Hirshleifer and Rasmusen, 1989; Fudenberg and Maskin, 1986; Axelrod, 1986). If cooperators punish defectors, then cooperation can be favored. However, if punishment is costly for the punisher (as it must certainly be), then cooperators who do not punish can invade because they avoid both the costs of being punished (for not cooperating) and the costs of punishing defectors (they are 2nd order free riders). If the private benefits derived from punishing are greater than the costs of administering it, punishment may initially increase, but cannot exceed a modest frequency (Boyd and Richerson, 1992). Using punishment, the problem of cooperation becomes one of how to maintain punishers in the population. One way to do this is to limit the mutational spectrum by eliminating strategies that cooperate, but do not punish (e.g., Axelrod, 1986; Axelrod and Dion, 1988; Gintis, 2000). How we evaluate this approach depends on how likely such a mutational constraint is in nature—which means, without further justification, such explanations are on a par with green beard approaches. Another way to solve the problem of punishment is to incorporate a recursive punishing strategy in which punishers punish individuals who do not cooperate and those who fail to punish individuals who do not punish (Boyd and Richerson, 1992). This solution is a kind of mathematical trick that eliminates the cost of punishment by spreading it out over an infinite space. Surely people cannot track defections through a nearly infinite set of stages. Below, I will discuss a culture–gene coevolutionary solution, based on punishment, for the evolution of one-shot, *n*-person, cooperation (Henrich and Boyd, 2001) that builds on a more-plausible informational constraint.

In considering efforts to explain human prosociality, including the above-described approaches to altruism, it is essential to ask why the solution does not apply to non-human animals. Kinship and direct reciprocity, for example, have been quite fruitful for understanding a variety of social behavior in non-human animals. Therefore, if a particular approach claims to explain large-scale and anonymous cooperation in humans, the approach must also either predict it in other species (specifying the conditions) or explain why the model is restricted only to humans. Given that we have not observed human levels of non-kin prosociality in other mammals, approaches should specify why they apply exclusively to humans. None of the standard evolutionary explanations deliver this.

4. The logic of selection within- and between-groups

Before examining the evolution of large-scale prosociality using a coevolutionary approach, it is important to understand how any evolutionary problem—and in particular the evolution of prosociality—can be viewed and described using the logic of multi-level selection. This perspective contrasts *within-group selection*, which acts against altruists who

are exploited by free riders from their groups, to *between-group selection* that favors groups with more prosocial individuals because such groups can outcompete groups dominated by free riders.

Debates about the potential relevance of 'group selection' (or 'multi-level selection'), particularly in its application to human behavior and cooperation, continue despite a fair amount of agreement among those who have seriously explored the question (Wade, 1985; Frank, 1995; Maynard Smith, 1998; Bowles, 2001; Gintis, 2000; Soltis et al., 1995). Since the publication of George Williams' *Adaptation and Natural Selection* in 1966 (Williams, 1966), which quite appropriately laid waste to a particularly naïve form of group- or species-functionalism that had been prominent in biology and anthropology during the preceding decades, a whole generation of biologists and anthropologists learned to scorn any explanation that involves selection among groups or proposes 'group-functional' traits. However, only a few years after Williams' book, Price (1970, 1972) provided an elegant formalization that showed, among other things, how the force of natural selection acting on genes can be partitioned into 'group-level' and 'individual-level' components. Unfortunately, the insight derived from Price's simple demonstration did not spread very far outside of theoretical evolutionary biology and failed to impede the spread of the belief that group-selectionist-thinking is somehow logically flawed, wrong-headed, or merely wishful thinking. This untutored dismissal of group selection, particularly in regard to cultural evolution, has slowed progress in understanding a variety of evolutionary processes.

The way evolutionary scholars often talk about 'group selection' leads people to mistakenly think of it as a separate process, somehow fundamentally different from 'individual selection' or 'natural selection' (e.g., Low, 2000; Ridley, 1993; Dawkins, 1976). Genetic evolution, at least from one perspective, is about changes in the frequency of alternative alleles, not about the frequencies of individual organisms or groups of organisms. Taking that as given, it is both possible and sometimes useful to write down accounting systems that track the frequencies of these alleles by examining their effects on fitness from different points of reference. Useful approaches might involve tracking the fitness of alleles, individuals, families, social groups, genomes, chromosomes, and so on. To show this, I develop a multi-group form of the Price equation that partitions the forces of natural selection into within- and between-group components. With this analytical tool, we can clarify the conditions under which the between-group component may overpower the within-group component and favor group-beneficial alleles or cultural traits.

We start with a population of N individuals subdivided into groups indexed by j, each with n_j members. There are no restrictions on how the groups are composed, except that all groups must contain at least one individual. The variable x_j gives the frequency of the trait/allele in subpopulation j, x'_j represents the same frequency in the next time period (or generation), and $\Delta \bar{x}$ expresses the average change in the frequency of the trait under investigation. Incorporating q_j ($=n_j/N$), the proportion of the total population accounted for in group j, and q'_j (the same proportion in the next time step) means

$$\Delta \bar{x} = \sum_j q'_j x'_j - \sum_j q_j x_j \tag{5}$$

Noting that $\Delta x_j = x'_j - x_j$ gives us

$$\Delta \bar{x} = \sum_j q'_j (x_j + \Delta x_j) - \sum_j q_j x_j$$

We can relate q_j and q'_j by comparing the fitness of group j, w_j, to the mean fitness across all groups: $q'_j = q_j w_j / \bar{w}$. With this, we arrive at the following expression:

$$\Delta \bar{x} = \sum_j \frac{q_j w_j}{\bar{w}} (x_j + \Delta x_j) - \sum_j q_j x_j = \sum_j q_j x_j \left(\frac{w_j}{\bar{w}} - 1 \right) + \sum_j q_j \left(\frac{w_j}{\bar{w}} \right) \Delta x_j$$

Then, using the standard definitions of *Covariance* and *Expectation*, the above simplifies to

$$\bar{w} \Delta \bar{x} = \text{Cov}(w_j, x_j) + E(w_j \Delta x_j) \tag{6}$$

As above, I used a non-standard notational practice and left the subscripts on the variables inside the operators to remind the reader that these are expectations taken over the groups indexed by j and weighted by q_j.

This form of the Price equation is useful because it partitions the forces that influence the dynamics in an intuitively tractable fashion. The covariance term summarizes the effect of selection between groups—group competition. If w_j, the fitness of group j, positively covaries with the frequency of the trait in group j (x_j), then this term will favor an increase in the average value of the trait in the overall population.

The second term, $E(w_j \Delta x_j)$, conceals the effect of selection within groups by inter-mingling it with the effects of transmission (i.e., mutation, non-random mating, etc.). We can further separate out the influence of selection within groups by re-applying the above technique to the term $w_j \Delta x_j$, which itself is the product of two average values: w_j is the average fitness of individuals in group j, and Δx_j is the average change in the frequency of the trait in group j. To do this, individuals within group j are indexed by i. As above, x_{ij} is the frequency of an allele (trait) in individual i and takes on the values of 1 (present) or 0 (absent). Substituting this into (6) yields

$$\bar{w} \Delta \bar{x} = \overbrace{\text{Cov}(w_j, x_j)}^{\text{selection between groups}} + \overbrace{E_j(\text{Cov}(w_{ij}, x_{ij}) + E(w_{ij} \Delta x_{ij}))}^{\text{selection and transmission within groups}} \tag{7}$$

To focus entirely on the effects of selection, we ignore any effects arising from the transmission process (e.g., recombination, mutation, etc.) and migration between groups (more on this below), so $\Delta x_{ij} = 0$, and (7) simplifies to (8):

$$\bar{w} \Delta \bar{x} = \underbrace{\beta_{w_j, x_j} \text{Var}(x_j)}_{\text{selection between-groups}} + \underbrace{E(\beta_{w_{ij} x_{ij}} \text{Var}(x_{ij}))}_{\text{selection within-groups}} \tag{8}$$

Eq. (8) tells us that the change in the frequency of an allele created by natural selection acting on individuals can be partitioned into between-group and within-group components. The magnitude of the between-group component depends on the amount of variation between groups and the size of the partial regression coefficient of group fitness on the frequency

of the trait in the group. The sign of the between-group component depends only on the sign of the partial regression coefficient. If having a higher frequency of the trait predicts higher group fitness (e.g., more cooperation), then the between-group component is positive. Similarly, the magnitude of the within-group component depends on the variation within groups and on the partial regression coefficient of individual allele frequency on individual fitness within groups. The sign of the within-group component depends on the sign of the regression coefficient. Using *partial* regression coefficients reminds us that although fitness is certainly influenced by many factors (too many to consider here), what matters at any given time is the linear relationship between fitness and the trait value, holding all other factors constant. Often the within- and between-group components have the same sign, so the partitioning provides little insight. However, if x happens to track the frequency of an altruistic allele, the within-group regression coefficients will be negative (by definition, altruists are exploited by non-altruists), while the between-group regression coefficient will be positive (groups with more altruists do better than groups with fewer). 'Group selection' or, more accurately, 'genetic group selection' occurs when the between-group component of natural selection acting on gene frequencies[7] overcomes the within-group forces to favor an equilibrium different from that which would be favored by the within-group component acting alone.[8]

4.1. The between-group component is often small relative to the within-group component

The between-group component of selection shown in (8) may be important for understanding a wide range of evolutionary phenomena. However, the real controversy and confusion erupts in its application to the evolution of cooperation/altruism. This is because under a wide variety of conditions, the between-group term in (8) is often small (or rapidly becomes small) relative to the within-group term. This occurs because migration, group formation, and other kinds of genetic mixing among groups deplete the variation among groups ($\text{Var}(x_j)$) while increasing (or at least maintaining) the variation within groups, $\text{Var}(x_{ij})$. To see this, consider a population consisting of two equally sized groups, one initially composed entirely of kindly altruists, who bestow fitness benefits on all other members of their group, and the other of selfish egoists who do not bestow benefits. This condition makes the between-group favoring altruism as big as possible, $\text{Var}(x_j) = 0.25$, while making the within-group variation zero, $\text{Var}(x_{ii}) = 0$. According to Eq. (8), altruistic genes will initially spread, no matter what the relative difference (in magnitude) is between the regression coefficients, because the within-group term is zero. If we keep the groups isolated for a long time and assume no mutation, the population will eventually be dominated entirely by altruists. However, if we allow a small amount of migration between the populations,

---

[7] I specify 'acting on genes' because natural selection can act on any kind of heritable phenotypic variation. Both cultural transmission and the transmission of acquired immunities provide examples of non-genetic, heritable variation that could be subject to natural selection.

[8] Hamilton (1975) re-derived his famous expression (Eq. (4); Hamilton, 1964, 1972) using the logic of selection within and between groups more than a quarter of a century ago, but the theoretical implications of his demonstration are still not widely appreciated by students of evolution and human behavior (also see Queller, 1992b).

say 5 percent exchanged per generation, the variation between groups, $Var(x_j)$, will begin to decline and rapidly approach zero. At the same time, migration will drive the variation within-groups, $Var(x_{ij})$, towards its maximum value of 0.25. Ignoring the fitness bounty reaped by immigrant egoists in the altruistic population, this migration rate alone will reverse the initial values of the between- and within-group variations in about 40 generations. Thus, a great deal of theoretical work shows that genetic group selection will only lead to substantial levels of altruism when groups are very small, migration rates are quite low, and the intensity of selection among groups is high compared to the intensity of selection within groups (Rogers, 1990; Crow and Aoki, 1982; Aoki, 1982; Boorman and Levitt, 1980; Kelly, 1992). In short, models for the evolution of altruism via genetic group selection are theoretically possible under the right constraints; however, most researchers do not expect these to be satisfied very often. With regard to humans, evidence from paleoanthropology and extant small-scale societies does not support such stringent constraints. On the contrary, the mating systems in many small-scale societies favor substantial genetic mixing between groups (Lee, 1979; Richerson and Boyd, 1998) that rapidly deplete $Var(x_j)$. Confirming this deduction (which is based on ethnographic data), the available genetic evidence from small-scale societies shows only small amounts of variation across groups (Hartl and Clark, 1989, pp. 300–301; Brown and Armelagos, 2001).

4.2. Multiple stable equilibria can resist the force of migration and maintain variation between groups

It is important to realize that the evolutionary dynamics associated with many different forms of social interaction, putting cooperation dilemmas aside, do not suffer from the depletion of between-group variation described above, and thus the between-group component is much more likely to be important. This occurs when local interaction between individuals can result in multiple locally stable equilibria. Unlike 'group-selected' solutions in cooperative interactions, neither small group size nor low migration rates are necessary to maintain between-group variation because *within-group selection* can counteract the genetic mixing produced by migration. This allows between-group selection to favor groups at equilibria that produce the highest mean group fitness (the efficient equilibrium). Therefore, when populations are structured and social interactions produce multiple stable equilibria, the between-group component is likely to have an important influence on the final distribution of behaviors/strategies (Boyd and Richerson, 1990). Thus, contrary to popularized claims about the general unimportance of group selection for understanding social behavior (laying aside the specific issue of cooperation), understanding the between-group component may be essential to explaining a wide range of behavior, especially in highly social species.[9] Simple game-theoretic examples of interactions with multiple equilibria include coordination interactions, battle-of-the-sexes games, mutualistic interactions (e.g., the 'stag hunt'

---

[9] Male vs. female exogamy provides an example of a social behavior with multiple equilibria. Assuming that there is a big cost to inbreeding and a polygynous mating system that produces many of unknown paternal half siblings, selection should favor one sex leaving the natal group. It is plausible that in many cases there are two stable equilibria: (1) males leave and (2) females leave. Both equilibria are found in nature.

game, Hirshleifer, 1982) and some models of reciprocity (Axelrod and Hamilton, 1981; Boyd, 1988).

To illustrate, imagine two proto-human groups. Due to random variation, one social group possesses a small number of mutants capable of using gestures (hand signals, body positions and facial expressions) to communicate with other such mutants, while the other group contains a similar small number of mutants capable of using verbalizations to communicate with other mutant-verbalizers; however, gestural communicators cannot communicate with verbalizers, and vice versa. If increased communicative abilities, in any form, are favored by natural selection in both groups, then the relative strength of selection on a particular form of communication (gestural vs. verbal) depends, in part, on the frequency of other individuals in the group who are capable of the same form of communication. Natural selection acting within the first group favors all gestural communicators, while in the second group it favors all verbalizers. Within-group forces will drive the first group towards all-gesturers and the second group towards all-verbalizers. Unlike cooperative dilemmas, migration between groups, which would tend to deplete the variation between groups, will be opposed by within-group frequency-dependent selection. If this within-group selection is strong enough, the variation between groups will remain high. Although verbalizers cannot invade the gestural group because of the coordination problem, groups of verbalizers may have a number of advantages over groups of gestural communicators such as: (1) communication is possible in darkness, through walls, in dense forest, and around topographical features, (2) communication is possible while using the hands for something else, and (3) communicating while injured is easier. As a consequence, the mean fitness of the verbalizer-group may be higher than the gestural group. If so, in the long run verbalizer-genes will be favored over gestural genes because of genetic group selection (i.e., the between-group component). As we will see later, cultural processes are even more likely to generate multiple stable equilibria, so the between-group component of cultural evolution is even more likely to be an important force of selection among culturally evolved equilibria.

4.3. Viewing kinship and reciprocity as group selection

In my view, the best way to view 'kin selection', 'reciprocity' and 'group selection' is as historically derived labels for different types of constraints (or classes of special assumptions). Imposing these constraints generates opportunities for natural selection to favor certain kinds of altruism. Often the labels seem to depend mostly on how the problem was initially set up. For example, why not consider explanations of altruism based on kin-recognition as plausible forms of 'green beard selection'?[10] Many students of evolution and human behavior do not realize that 'inclusive fitness' (which was first used to derive (4) using kinship; Hamilton, 1964), 'individual fitness' (first used to derive (4) for 'reciprocal altruism' by Axelrod and Hamilton, 1981), and 'group selection' (i.e., the partitioned Price equation was used to derive (4) in Price, 1972; Hamilton, 1975) are simply three systems of gene-tracking and fitness accounting from three different perspectives. Any solution can be reformulated from each perspective to yield the identical answer. Hamilton (1975), for

---

[10] Note that sexual recombination does not work against kin recognition in the same way that it does against non-kin green beard genes (unless they are on the Y-chromosome).

example, reformats his kin-selection approach using the Price equation, instead of tracking inclusive fitness.

To demonstrate the relationship between the 'group selection' approach and reciprocity, I will analyze a standard, 2-person, reciprocity model using the partitioned Price equation (4). In doing this, we observe that reciprocity favors cooperation when the between-group component overpowers the individual component. Of course, it is equally possible to re-derive typical 'group selection' solutions (selective emigration, assortative interaction, etc.) using inclusive fitness, or even by tracking individual fitness. None of the accounting techniques can claim general superiority over the others (Frank, 1998; Queller, 1992a; Sober and Wilson, 1998; Wade, 1985). However, despite the fact that it is possible to do this, it is neither easy nor particularly desirable.

In the repeated prisoner's dilemma described here, individuals are paired in groups that play for m rounds each generation. During each round, individuals play either 'cooperate' (C) or 'defect' (D). Table 1 gives the payoffs received by the row player. The index j labels three groups 1, 2, 3. Each group consists of two individuals labeled $i = 1$ or 2. Group $j = 1$ contains two individuals who both play the strategy tit-for-tat. Group $j = 2$ contains one tit-for-tater and one defector (who always plays D). Group $j = 3$ contains two defectors. The variable x_{ij} equals 1 for tit-for-tater's and 0 for defectors. The variable x_j represents the frequency of tit-for-tater's in group j. Eq. (8) gives the change in frequency of tit-for-tater's. Here are the components of (8):

$$E(\beta_{w_{ij}x_{ij}} \text{Var}(x_{ij})) = \tfrac{1}{3}[(0) + (-4)(0.25) + (0)] = -0.33, \qquad \text{Var}(x_j) = 0.167,$$

$$\beta_{w_j,x_j} = m$$

Substituting these into Eq. (8) gives us

$$\bar{w}\Delta\bar{x} = \underbrace{0.167m}_{\text{between-group}} - \underbrace{0.33}_{\text{within-group}}$$

This equation shows that it is the between-group component of natural selection that favors the evolution of reciprocity (in this case tit-for-tat), not 'individual selection'. In fact, the within-group component is always negative—meaning 'individual selection' always selects *against* reciprocity. Within groups, tit-for-tater's never do better than a fitness tie (with another tit-for-tater) and always lose against defectors—thus the within-group regression coefficients, $\beta_{w_{ij}x_{ij}}$, are always zero or negative. The magnitude of the positive

Table 1
Two-person prisoner's dilemma payoff matrix[a]

Player 1 plays	Player 2 plays	
	C	D
C	2	−1
D	3	0

[a] Payoffs are for the row player (Player 1). Players 1 and 2 are interchangeable. 'C' indicates cooperation and 'D' indicates defection.

between-group component (i.e., group selection) depends on m, the number of rounds of play. Assuming groups are randomly remixed every generation, m controls the migration rate between groups. As m increases, groups become more stable. When $m = 1$, groups completely remix after every round and the between-group component is minimized. Reciprocity is favored when m is sufficiently large that teams of reciprocators can run-up their combined fitness total sufficiently to overcome the relative fitness losses they suffer at the hands of defectors, within groups. The fact that the between-group component is what drives reciprocity is often lost because it is substantially easier to solve this problem using the 'individual fitness-accounting' approach—in which a strategy's relative fitness is calculated by averaging across all possible groups (for more on this, see Sober and Wilson).

5. Kinds of cooperation and altruism

We are not limited to picking only one solution to the altruism dilemma, especially in the human case. Our psychology may embody the effects of natural selection having taken advantage of different stable informational regularities over our evolutionary history. For example, the nature of the human family—with its division of labor and stability (Boyd and Silk, 1997; Klein, 1989)—suggests that human ancestors were likely to have repeatedly found themselves in small sub-groups (called 'families') in which other individuals were likely to have the same genes by descent from a recent ancestor. Consequently, like other primates, we are likely to have some cognitive machinery dedicated to bestowing benefits on kin. Similarly, life in small-scale, fairly stable, groups may provide for lots of repeated interaction with unrelated or distantly related individuals. Assuming our cognition/emotional system was complex enough that mutants could not easily pose as friends, then we should expect a psychology that favors helping well-known, reliable reciprocators (i.e., friends)—and for *distinguishing* them from others, especially anonymous strangers. Unfortunately, kin selection can only explain cooperation among close kin, and reciprocity (including indirect reciprocity) is limited to small groups with lots of repeated interaction and high quality reputation information. Experimental findings from many small- and large-scale societies show that people will trust, cooperate and behave altruistically toward anonymous individuals in simple one-shot games (Henrich et al., 2001; Roth et al., 1991; Fehr and Gächter, 2000, 2002; Batson and Shaw, 1991)—a finding further confirmed by lots of ethnographic evidence (Richerson and Boyd, 1998, 2000). As a consequence, none of the above solutions seems capable of: (1) explaining the large-scale cooperation among non-relatives that we observe in both modern and pre-modern societies, including foraging societies,[11] or (2) isolating why evolutionary processes have only led to large-scale cooperation in our species; if all this prosociality is only a product of

---

[11] Two things are worthy of note with regard to foraging societies. First, the nature of sharing, especially of sharing game, cannot be explained by either reciprocity- or kin-based theories (Bliege Bird and Bird, 1997), even in the smallest small-scale societies. Many of the simplest societies such as the !Kung, have quite strong institutions for maintaining cooperation on a scale much larger than the family or the band (Wiessner, 1983). Second, contrary to the view of foraging created by anthropologists studying extant groups, lots of archaeological and ethno-historical data indicates that foraging societies can be politically, economically and socially complex, with large-scale cooperation, social stratification and a substantial division of labor (Arnold, 1996; Roscoe, 2002).

reciprocity and kinship, why do not Japanese macaques forced to live in large zoo populations develop armies, recycling, taxation, storage, and other aspects of human social organization?

6. Cultural group selection, cultural transmission and large-scale cooperation

In many social groups (but not all), humans cooperate on a larger scale than any other species—with the possible exception of eusocial insects. Interestingly, humans are also the most proficient at, and most reliant on, social learning to acquire behavioral practices and strategies. Our cognitive abilities to acquire information via imitation and other forms of direct social learning far exceed those of any other species (Tomasello, 1994, 2000). The operational details of these learning mechanisms appear to have been 'designed' by natural selection to extract useful information from the social world—that is, from the minds of other individuals (Boyd and Richerson, 1985; Henrich and Gil-White, 2001; Henrich and Boyd, 1998). Perhaps it is merely a coincidence that humans are both the most cooperative species and the most reliant on, and proficient at, social learning. However, I argue that the nature of our cultural transmission capacities, and of human psychology more generally, creates stable behavioral equilibria consisting of combinations of cooperation and punishment that are not available to genetic evolutionary processes in acultural species. The existence of these additional, culturally evolved, behavioral equilibria make the group selection component of cultural evolutionary processes much more powerful relative to the within-group component than can occur in genetic evolution. By systematically altering the social environment in favor of prosocial phenotypes, cultural processes create the conditions for natural selection to favor prosocial genes that could not otherwise be favored in mammalian social species because non-humans lack the requisite social learning capacities (e.g., high fidelity imitation) that give rise to gene–culture coevolutionary processes.

We have empirical reasons to believe that culturally transmitted ideas, beliefs and values (i.e., information) are important for understanding human cooperation. First, unlike other animals, the domains of cooperative behavior in humans vary from place to place and from group to group. In some societies, people may cooperate in fishing and house-building, but not warfare. In neighboring groups that inhabit the same physical environment, people may cooperate in warfare and fishing, but not house-building. Thus, unlike cooperation in eusocial insects and kin-based altruism in non-human primates, there is a tremendous amount of variation in the cooperative domains among human groups that is independent of differences in physical environments or local ecologies (Kelly, 1985; Henrich and Boyd, 1998). Second, unlike non-human animals, the scale of human cooperation varies from little cooperation outside the extended family (Johnson, 2003; Johnson and Earle, 1987) to massive cooperation on the level of nation states containing millions of individuals. This observation is confirmed by experimental data using one-shot Ultimatum Games that show big differences across human societies (Henrich et al., 2001). In contrast, the scale of cooperation in non-humans does not vary much among social groups within a species, only among species. Given these two kinds of variation within our species, cultural evolutionary processes seem much more likely than genetic processes to explain the patterns—especially

given the modest genetic heterogeneity found among humans as a species (compared to other species such as chimpanzees) and the relatively recent (last 5000 years) rapid emergence of very large-scale cooperation.

6.1. Cultural group selection

Interestingly, 'group-functional' explanations for cooperation and other forms of group-beneficial behavioral patterns have long been part of anthropology and sociology (Spencer, 1891; Rappaport, 1984; Harris, 1977; Vayda, 1971; Turner and Maryanski, 1979). However, by the end of the 1970s the anti-group selection movement had penetrated cultural anthropology and, by analogy with biological evolution, was used to argue that the individual was the relevant level of analysis in cultural evolution (Harris, 1979, pp. 60–61). The problem was that both biology and anthropology lacked a sufficiently clear understanding of the differences between genetic and cultural evolution to understand why the between-group processes that were unlikely to account for seemingly group-beneficial behavioral traits in genetic evolution could still operate effectively (even rapidly; Boyd and Richerson, 2002b) in cultural evolution. Below, I explain why.[12]

The Price equation derived above, with genetic evolution in mind, turns out to be a very general statement about any evolutionary system, which we will use to frame our thinking about cultural evolution. The 'altruist gene' (x) that we focused on in the earlier derivation could be any characteristic of any evolving system, including the frequency of hydrogen atoms in a cluster of galaxies, or a quantitative phenotypic measure such as IQ, 'cooperativeness in group fishing ventures', managerial success, malaria resistance, or the variance in height of brothers. To discuss cultural evolution using the Price equation, I re-label the variables to avoid confusion. Replacing x, ϕ measures a quantitative phenotypic trait that can be influenced by cultural transmission. As a behavior, preference or strategy, ϕ_{ij} could measure individuals' willingness to die for their country (or tribe) in war, the amount of time or money an individual contributes to charity, or how much of his total 'catch' a hunter brings back to share in camp (as opposed to that portion he eats alone at the kill site). In this context, f_{ij} replaces w_{ij} and gives the cultural fitness for particular values of ϕ_{ij}. Cultural fitness measures the degree to which a particular value of ϕ, representing stuff stored in the head of individual i in group j at time t, affects its proportional representation in the population at time $t + 1$. It may be thought of as replicability, transmissibility or simply influence—note that nowhere in this formulation there is a need for cultural stuff to 'replicate' or be defined as discrete units.[13] The cultural fitness of ϕ_{ij} is jointly determined by the operational details of our social learning psychology and by how that psychology interfaces with the environment (Boyd and Richerson, 1985):

$$\bar{f}\Delta\bar{\phi} = \underbrace{\beta_{f_j,\phi_j}\,\text{Var}(\phi_j)}_{\text{selection between-groups}} + \underbrace{E(\beta_{f_{ij}\phi_{ij}}\,\text{Var}(\phi_{ij}))}_{\text{selection within-groups}} \tag{9}$$

---

[12] In Economics, cultural group selection arguments were foreshadowed in the work of Hayek (1979, 1988).

[13] For a discussion of the confusion created by cultural 'replicators', see Henrich and Boyd (2002) and Boyd and Richerson (2000).

6.2. Cultural transmission, human psychology and between-group variation

As I discussed earlier, the main problem with 'group selection' is the maintenance of between-group variation in the face of migration or other forms of mixing. Unlike genetic evolution, several different mechanisms will act to reduce the within-group term in (9), $Var(\phi_{ij})$, while maintaining the between-group term, $Var(\phi_j)$, even in the face of substantial migration. I will discuss four of these mechanisms. The first two are rooted in the details of how our cognition exploits the distribution of behaviors and ideas among members of a social group in order to 'decide' which of these traits to acquire. As I will show, both evolutionary modeling and empirical data support the existence of these cultural transmission mechanisms. The third and fourth mechanisms are not cultural transmission capacities, but rather psychological 'tastes' or preferences for: (1) avoiding behaviors that deviate from the common pattern and (2) punishing those who do not conform to the expected pattern. In my view, these two 'tastes' are probably either the products of purely cultural evolution (driven by cultural group selection), or coevolved products of genes responding to the novel social environments created by cultural group selection. Thus, they cannot be responsible for initiating the coevolutionary process that led to high levels of cooperation in humans. Nevertheless, once brought into existence by the first two mechanisms, they further catalyze the cultural group selection process by bolstering the forces that create and maintain differences between groups. When applied to social interaction occurring in single groups, these four mechanisms create equilibria that do not exist in genetic systems, including equilibria with high levels of cooperation and punishment in one-shot, *n*-person games (Henrich and Boyd, 2001). With the emergence of multiple equilibria, cultural group selection supplies a process that selects among alternative stable equilibria in favor of the ones best able to compete with other groups (Boyd and Richerson, 1990).

The first mechanism, *conformist transmission*, is a psychological propensity to preferentially copy high frequency behaviors. By biasing individuals in favor of copying common behaviors, preferences or behavioral strategies, this transmission bias tends to homogenize social groups. There are both theoretical and empirical reasons to believe that humans possess a tendency to copy the most common behavior. Theoretically, Henrich and Boyd (1998) have shown that genes favoring a heavy reliance on social learning and *conformist transmission* (copying the majority) can outcompete genes favoring individual learning in both spatially and temporally varying environments. This model predicts two important things: (1) that individuals should increase their reliance on social learning when individual (or environmental) information becomes less certain or as the difficulty of the problem increases, and (2) that individuals should rely on copying the majority (conformist transmission) under a wide range of conditions (also see Boyd and Richerson, 1985; Ellison and Fudenberg, 1993).

Independent experimental work in psychology supports both predictions, as well as a number of other predictions arising from this model. Psychologists studying conformity have shown that as a task's difficulty and financial incentives rise, individuals *increase* their reliance on imitation (vs. individual analysis) regardless of whether others will know how they behave (reducing any fear of social sanctions; see Baron et al., 1996; Insko et al., 1985). Furthermore, with real money on the line, other experiments show that individuals rely on copying the majority in *social dilemmas*, both when self-interest conflicts with

the group-interest and when self-interested choices correspond to group-interested choices (Smith and Bell, 1994; Wit, 1999). Finally, Henrich (2001) shows that the slow take-offs and 'critical mass tipping-point' observed in many empirical studies of the diffusion of innovations are quite consistent with the effects of conformist transmission.

The presence of this adaptive bias on our social learning cognition means that, in the absence of unambiguous information from the environment or other decisive social learning stimuli (such as a prestigious individual, see the next mechanism), individuals will preferentially copy the most common ideas, beliefs, values and practices. Because new immigrants and the offspring of immigrants will preferentially adopt the common practices, conformist transmission can maintain group differences in a way that genetic transmission cannot—because offspring acquire their genes from their parents, not from the group.[14] Consequently, as a byproduct of its evolved design, conformist transmission decreases the phenotypic variation among individuals within groups, thereby depleting the strength of within-group cultural forces. As stochastic forces such as cultural drift (sampling errors in transmission), biological shocks (e.g., plagues) and environmental disasters introduce random variation between groups, conformist transmission will act to maintain this group-level variation—variation that would otherwise be depleted by migration between groups, natural selection and payoff-biased forms of cultural transmission. Thus, by reducing within-group variation and increasing between group variation, conformist transmission provides the raw materials for cultural group selection.[15]

Other cultural transmission mechanisms can create the same effect through other means. If individuals possess a psychological bias to preferentially copy people who are both more successful (get higher payoffs) and similar to themselves in some marker trait such as language or dress, then under a wide variety of conditions (even with substantial migration rates), these cultural transmission mechanisms push the variation within groups towards zero (opposing the force of migration), while sustaining substantial amounts of variation among groups (McElreath et al., 2003; Boyd and Richerson, 1987).

There are both theoretical and empirical reasons to believe that people preferentially copy successful individuals and people like themselves. Recently, Gil-White and I have argued that, with the rise of imitation in the human lineage, natural selection favored cognitive abilities to rank potential models according to their payoffs and to preferentially imitate highly ranked models. Among the cues individuals use to rank potential models is the amount of prestige-deference an individual receives from other people. This deference acts as an honest signal of whom other individuals believe is highly successful or skilled because deference is 'paid' to such individuals in exchange for copying opportunities (high quality information). This rank-based copying bias, which we call *prestige-biased transmission*, allows individuals to shortcut environmental or trial-and-error learning processes, and leap directly to better-than-average skills by imitating successful or skilled cultural models.

---

[14] This is consistent with a substantial amount of work in psychology and behavioral genetics showing that children do not acquire much, via social learning, from their parents (Harris, 1998; Plomin et al., 2000).

[15] This process also provides an explanation of why humans have different 'cultures' and other animals do not. That is, conformist transmission provides one important reason why people in the same social group tend to believe the same things and why these beliefs persist over long periods. Without a conformist component to create 'cultural clumps', social learning models predict (incorrectly) that populations should be a smear of ideas, beliefs, values and behaviors, and that group differences should only reflect local environmental differences.

Further, because the world is a noisy, uncertain place, and it is often not entirely clear why a particular individual acquires great prestige or success, humans have evolved the propensity to copy a wide-range of cultural traits from prestigious individuals, only some of which may actually relate to the individuals' success (Henrich and Gil-White, 2001).

A substantial amount of empirical work from psychology, economics and anthropology confirms a variety of predictions derived from the above theory (Henrich and Gil-White, 2001). However, most importantly for our purposes, these findings confirm that people preferentially imitate the ideas, opinions, beliefs, strategies, and behaviors of prestigious individuals across a wide range of domains (even in domains outside the expertise of the prestigious individual). For example, in a synthesis of the diffusion of innovations literature, Rogers (1995) has shown that the rate of spread of novel technologies and new economic practices into different social groups depends on how quickly prestigious, local "opinion leaders" adopt these innovations. In the laboratory, using a multi-round, market game with substantial incentives in which the results of each player's decisions were posted between sessions, experimental economists unexpectedly found that MBA students tended to mimic the decisions of successful players, even though rewards were distributed on a competitive basis. Allowing imitation also moved the group average substantially closer to the optimal decision predicted by Portfolio Theory (Kroll and Levy, 1992). In a different experiment, Offerman and Sonnemans (1998) showed that subjects making investment decisions tended to copy the beliefs of successful individuals (about the current environment), even when players clearly knew that these individuals had the same information about the current situation (also see Pingle, 1995; Pingle and Day, 1996).

When coupled with prestige-biased transmission, a self-similarity transmission bias makes good adaptive sense. Individuals should imitate successful others only as long as those individuals are not too different from the imitators themselves—otherwise imitated traits may be inapplicable to the copier's particular abilities and circumstances. Interestingly, as an emergent byproduct, this combination of cultural learning mechanisms decreases within-group variation while increasing between-group variation (Boyd and Richerson, 1987).

Following this line of reasoning, theoretical work suggests that individuals in social groups throughout human history needed to coordinate their beliefs, norms and expectations in order to make economic exchanges, marry and raise children. In theoretical work examining the interaction of genes and culture in solving these coordination problems, McElreath et al. (2003) show that natural selection will favor the evolution of a bias to cue off salient symbolic markers when they covary with individuals' underlying norms of interaction—so by using symbolic markers individuals avoid acquiring norms that would produce costly uncoordinated interactions. Although no one has empirically examined the relationship between imitation and ethnic markers, more general evidence for self-similarity learning biases comes from the diffusion of innovations literature (Rogers, 1995, p. 286), laboratory psychology (Rosekrans, 1967; Stotland and Dunn, 1962, 1963) and studies of child development (Harris, 1998).

A third mechanism, *punishment* of non-conformists or norm violators, acts to homogenize social groups. A great deal of ethnographic and experimental research suggests that people in many societies will inflict punishment on individuals who violate group norms of behavior (Sober and Wilson, 1998; Roth et al., 1991; Henrich et al., 2001; Fehr and Gächter, 2002).

If violators of group norms receive costly punishments then either prestige-biased transmission or simply trial and error learning will reduce the variation between groups. Costly punishments will mean lower payoffs for norm violators, so that under prestige-biased transmission, the behaviors of norm violators are less likely to spread. The problem with this norm-based punishment, as I discussed earlier, is the difficulty in explaining how it could evolve in a purely genetic system of inheritance. If punishing norm violators is costly to the punisher, then punishing strategies are unlikely to evolve under natural selection. In the next section, I describe how punishing behaviors can be stabilized if humans have a very small amount of conformist transmission bias in their social learning psychology.

A fourth mechanism, *normative conformity*, arises because individuals want their behavior to match the common behavior in their social group. Normative conformity differs from conformist transmission in that individuals are not using the frequency of a behavior, belief or idea as an indirect indicator of its worth. Instead, normative conformity alters individuals' socially displayed behavior (without necessarily changing their minds) because they want their behavior to match the majority, not because they 'believe' the majority is probably doing the smart thing. A vast amount of experimental work beginning with Asch's famous studies (e.g., Asch, 1951), and including cross-cultural work (Furnham, 1984), show that normative conformity is a robust part of human psychology (Neto, 1995), at least in complex societies (nobody has done work among small-scale societies). Baron et al.'s (1996) experimental work attempted to differentiate normative conformity from the effects of conformist transmission. Their results strongly suggest that people (i.e., university students) have both conformist transmission and normative conformity components to their psychology. These two psychological processes probably evolved for separate reasons. While conformist transmission seems to have evolved genetically as a shortcut means of acquiring useful information, normative conformity is more likely a product of cultural evolution or culture–gene coevolution, and it may have evolved in response to the spread of punishing strategies, as a means to manipulatively cue one's 'similarity' to other group members in order to gain the advantages of in-group membership and avoid punishment. Also, note that normative conformity *may* result in conformist transmission. Under some circumstances, imitators may mistakenly infer the underlying preferences or goals in observing the compliant behavior of models (by assuming the models 'like' to perform certain behaviors), and thereby acquire both the outward behavior and the underlying supporting preferences that were not originally possessed by the models.

Individually, and in combination, these mechanisms increase the importance of cultural group selection by creating a myriad of additional stable equilibria for all kinds of cultural traits, ideas, beliefs, values and practices, including those that govern social and cooperative interactions. As I discussed earlier, the between-group component becomes important in circumstances with multiple stable equilibria because within-group selective processes, acting to push the system back to the locally stable equilibria, oppose the effects of migration that tend to reduce the variation between groups ($\text{Var}(\phi_j)$ in Eq. (9)). The nature of the mechanisms described above means that most of the time these stable equilibria will be monomorphic (everyone possesses the same cultural traits), thus placing $\text{Var}(\phi_{ij})$ near zero.

In thinking about how the difference between cultural and genetic evolution might affect the spread of prosociality, it is also important to keep two other differences in mind. First, cultural evolution is likely to proceed much more rapidly than genetic evolution because

mechanisms such as prestige- and conformist-biased cultural transmission favor rapid transmission and can spread novel behaviors, ideas and practices among populations within a single generation (Boyd and Richerson, 1985, 2002b)—note that substantial evidence indicates that most cultural transmission is not vertical (i.e., not parent to offspring; Plomin et al.). Second, cultural transmission is likely to be more subject to drift and random variation than genetic transmission. The imitative skills of humans are good, qualitatively better than any other animals, but certainly worse than genetic replication. The combination of these two differences means that human groups under cultural evolution will drift into the domains of attraction of alternative equilibria more often than in genetic evolution. Thus, cultural group selection will "see" a greater variety of equilibria than genetic evolution over the same amount of time.

6.3. Conformist transmission can stabilize cooperation by stabilizing punishment

Before discussing cultural group selection, it is necessary to see how cultural transmission mechanisms can combine to create alternative stable prosocial equilibria. Henrich and Boyd (2001) have shown that if human social learning psychology contains both a transmission bias to copy successful individuals ('payoff-biased' or prestige-biased transmission) and a bias to copy high frequency behaviors (conformist transmission), and there are an arbitrary number of 'punishing levels', then highly cooperative equilibria can exist even if conformist transmission is only a weak component of human cultural transmission.

A tendency to copy high frequency behaviors can stabilize costly cooperative strategies without punishment, but only if this conformist transmission is quite strong compared to payoff-biased transmission. All other things being equal, payoff-biased transmission causes higher payoff variants to increase in frequency, and thus cooperation is not evolutionarily stable under plausible conditions because not-cooperating leads to higher individual relative payoffs (within groups) than cooperating. Thus, on its own, payoff-biased transmission suffers the same problem as natural selection in genetic evolution. However, if our social learning psychology contains a combination of conformist and prestige-biased transmission (as I have argued above), then, if cooperation becomes common, conformist transmission will oppose payoff-biased transmission and favor cooperation. When cooperation is not too costly, conformist transmission will maintain cooperative strategies in the population at high frequency. However, because both theory and evidence (Henrich, 2001) suggests that conformist transmission is relatively weak compared to payoff-biased transmission (and the costs of cooperation are probably substantial), it seems unlikely that conformist transmission will be able to maintain cooperation.

A quite different logic applies to the maintenance of punishment. Suppose that culturally transmitted punishing and cooperating strategies are both common and that being punished is sufficiently costly so that cooperators have higher payoffs than defectors. Rare invading 2nd order free riders who cooperate but do not punish will achieve higher payoffs than punishers because they avoid the costs of punishing. Because defection does not pay, the only defections will be due to rare mistakes, and thus the *difference* between the payoffs of punishers and 2nd order free riders will be relatively small compared to that between 1st order free riders and cooperators. Hence, conformist transmission is more likely to stabilize the punishment of non-cooperators than of cooperation itself. As we ascend to

higher order punishing, the difference between the payoffs to punishing vs. non-punishing decreases geometrically towards zero because the occasions that require the administration of punishment become increasingly rare. Second order punishing is required only if someone erroneously fails to cooperate and then someone else erroneously fails to punish that mistake. For third order punishment, yet another failure to punish must occur. As the number of punishing stages (i) increases, conformist transmission, no matter how weak, will at some stage overpower payoff-biased imitation and stabilize common ith order punishment. Once punishment is stable at the ith stage, payoffs will favor strategies that punish at the $i - 1$ order, because common punishers at the ith order will punish non-punishers at stage $i - 1$. Stable punishment at stage $i - 1$ order means payoffs at stage $i - 2$ will favor punishing strategies, and so on down the cascade of punishment. Eventually, common 1st order punishers will stabilize cooperation/altruism at stage 0.[16]

The constraint that allows cooperation to arise and remain stable in this model is the costliness of information. From the gene's point of view, this is a maladaptive side-effect of conformist transmission. If there were genetic variability in the strength of conformist transmission (α) and cooperative dilemmas were the *only* problem humans faced, then conformist transmission might never have evolved. However, human social learning mechanisms were selected for their ability to efficiently acquire adaptive behaviors over a wide range of behavioral domains and environmental circumstances—from figuring out what foods to eat, to deciding what kind of person to marry—precisely because it is costly for individuals to determine the best behavior (Henrich and Boyd, 1998). Thus, we should expect conformist transmission to be important in cooperation as long as distinguishing cooperative dilemmas from other kinds of problems is difficult, costly or error-prone. Looking across human societies we find that cooperative dilemmas come in an immense variety of forms, including harvest rituals among agriculturalists, barbasco (communal) fishing among Amazonian peoples, warfare, irrigation projects, taxes, voting, meat sharing and anti-smoking pressure in public places. It is difficult to imagine a cognitive mechanism capable of distinguishing cooperative circumstances from the myriad of other problems and social interactions that people encounter—this is especially true when information about the payoffs to alternative behaviors or strategies is costly to acquire.

As I have mentioned, natural selection favors the evolution of conformist transmission because it helps to solve individual-level problems (Henrich and Boyd, 1998), and leads to an effective solution in many forms of social interaction, such as coordination games. Consequently, in order for natural selection to favor an ability to "switch-off" the conformist effect upon encountering a cooperation problem, individuals must be able to accurately distinguish cooperative dilemmas from all other dilemmas. To accomplish this, individuals must be able to acquire sufficient information about the relative payoffs received by individuals behaving in alternative ways. However, the amount of payoff information necessary to distinguish a cooperative dilemma from other problems (e.g., coordination interactions) is also sufficient to determine the optimal solution to the problem without any social learning. If people could acquire and process sufficient information to 'know' when to switch-off

---

[16] Boyd et al. (2003) have combined the fact that the payoff difference between punishing and not-punishing is smaller than that between cooperating and not-cooperating with group selection to show that the conditions for the evolution of cooperation and punishment are much more plausible for cultural evolution than genetic evolution.

conformist transmission, then they would be able to determine the optimal choice in the situation, and we would not expect to observe any social learning in social interactions. Nevertheless, when experimentalists have provided opportunities for social learning in economic interactions, subjects usually rely heavily on it (Kroll and Levy, 1992; Ball et al., 2001; Smith and Bell, 1994; Pingle, 1995; Pingle and Day, 1996; Offerman and Sonnemans, 1998; Wit, 1999). Smith and Bell (1994) have tested this explicitly by comparing subjects' strategic use of social information with their imitative use of this information, and found that subjects imitate others *both* when self-interest conflicts with group-interest and when they run in parallel. No difference is observed in the relative strengths of people's imitative tendencies (cf. Weinmann, 1994).[17]

Overall, given the complexity of human cultural adaptations and social environments as well as the lack of information about the costs and benefits of rarely observed alternative social behaviors, this constraint likely holds and is certainly more plausible than, for example, mutational restrictions that confine punishment-alleles and cooperation alleles to the same gene.

6.4. Different processes of cultural group selection

Once stable culturally transmitted differences arise between groups, at least three different forms of cultural group selection may influence the evolution of practices, beliefs, ideas and values: *demographic swamping*, *intergroup competition* and *prestige-biased group selection*. Demographic swamping produces changes in the frequency of cultural traits in an overall meta-population because some social groups (perhaps just one) reproduce new individuals faster than other groups as a consequence of some set of culturally transmitted ideas or practices that are relatively stable in those groups—this is natural selection acting on between-group cultural variation. Demographic swamping probably explains the spread of early agriculturalists into regions once dominated entirely by hunter-gatherers. Agriculturalists gradually replace foragers, increasingly compressing them into tracks of inarable land (Cavalli-Sforza et al., 1994; Young and Bettinger, 1992; Diamond, 1997). Such empirical cases of demographic swamping suggest that this is probably the slowest kind of cultural group selection, operating on time scales of millennia.

In intergroup competition, different cultural groups may also compete directly for access to resources through warfare and raiding. Cultural practices and beliefs that provide a competitive edge to groups in warfare will proliferate at the expense of traits that make groups less effective in competition (and more likely to be defeated, absorbed or dispersed). Such cultural traits might relate to beliefs about patrilocality, heroism, patriotism, economic cooperation (leading to surplus production), the villainy of foreigners, and the proper forms of social or political organization (or all of these). In exploring cultural group selection resulting from intergroup competition, Soltis et al. (1995) calculated evolutionary rates using a model based on group "extinctions" ("extinction" only implies that the group members

---

[17] Further, the availability of the payoff information required to distinguish between cooperative and other kinds of social interactions is often limited. If an individual enters the world and everyone is playing strategy "A", then he has no idea what will happen if he plays "B". If the underlying dilemma is cooperative and A is "cooperate", then everyone may receive the same payoffs, so it will 'look' like (be indistinguishable from) a coordination game.

must be disbanded and scattered, not necessarily killed). Using empirical data from New Guinean horticultural groups, Soltis et al., estimated that a group-beneficial, cultural trait could spread to fixation on time scales of 500–1000 years.

One of the best-documented cases of cultural group selection occurred during the 18th century among the anthropologically famous ethnic groups of the Nuer and the Dinka. Before 1820, the Nuer and Dinka (Kelly, 1985) occupied adjacent regions in the southern Sudan. Despite inhabiting similar environments and possessing identical technology, the two groups differed in significant ways. Economically, both the Dinka and the Nuer raised cattle, but the Dinka maintained smaller herds of approximately nine cows per bull, while the Nuer maintained larger herds with two cows per bull. The Nuer ate mostly milk, corn and millet and rarely slaughtered cows, while the Dinka frequently ate beef. Politically, the Dinka lived in small groups, the largest of which corresponded to their wet season encampment. In contrast, the Nuer organized according to a patrilineal kin system that structured tribal membership across much larger geographic areas. Consequently, the size of a Dinka social group was limited by geography, whereas the Nuer system could organize much larger numbers of people over greater expanses of territory. Despite the similarity of their environments, these two groups showed substantial economic and political differences. Over about 100 years, starting in about 1820, the Nuer dramatically expanded their territory at the expense of the Dinka, who were driven off, killed, or captured and assimilated. As a result, Nuer beliefs and practices spread, fairly rapidly, across the landscape relative to Dinka beliefs and practices despite the fact that the Nuer were soon living in the once "Dinka environment".

Another subtler form of cultural group selection is likely to operate on even shorter time scales than intergroup competition: prestige-biased group selection. Under prestige-biased transmission, people preferentially copy individuals who get higher payoffs. The higher an individual's payoff, the more likely that individual is to be imitated. If individuals occasionally have opportunities to copy people in neighboring groups, then individuals from groups at more efficient (cooperative) equilibria will be preferentially imitated by individuals from groups at less efficient (non-cooperative) equilibrium because the average payoffs to individuals from cooperative groups are much higher than the average payoff of individuals in non-cooperative groups (Henrich and Boyd, 2001). Boyd and Richerson (2002b) have shown that, under a wide range of conditions, this form of cultural group selection will spread group-beneficial behaviors from a single group (at a group-beneficial equilibrium) through a meta-population of other groups that were previously stuck at more individualistic equilibria. Furthermore, this process allows more efficient equilibria (institutional forms) to recombine and spread. For example, in an ecology of many groups, one group may have drifted onto an efficient system of patrilineal inheritance (suppose this is more efficient than matrilineal or bilateral inheritance); another group may have evolved cooperative norms specifically for the construction of large buildings; and, a third group may have developed a segmentary lineage system that allows them to organize large numbers of warriors. The above process allows these three different kinds of behavioral equilibria to individually spread, and eventually recombine in a single social group, making this group the most successful/competitive group by far. Notably, this process is fast and can probably occur on time scales of decades or centuries.

7. Genes may respond to the changed social environment created by cultural evolution

By systematically altering the selective environment faced by genes, cultural evolution via cultural group selection may lead to the subsequent spread of prosocial genes—genes that would not otherwise be favored without the action of cultural processes. As cultural group selection spreads punishment, prosocial genes (genetically encoded "preferences" for in-group altruism and punishment) may be favored if they allow individuals to avoid the costs of being punished for not cooperating. Because such prosocial genes would have evolved in a world with substantial amounts of culturally evolved, between-group variation, such genes would foster prosocial psychologies adapted to cue off local behavioral patterns. This would allow human social groups to vary in both their degree of prosociality and their domains of cooperation (house-building, warfare, fishing, etc.) in a manner not possible in non-cultural species. One way to think about these "preferences" is as learning biases that allow individuals to rapidly acquire prosocial behaviors, values or strategies by observing others, or with a minimum of punishment. Elsewhere, Henrich and Boyd (2001) have shown that prosocial genes will spread under a wide range of conditions in the altered environments produced by cultural group selection.

8. Conclusions

Besides providing a plausible evolutionary foundation for the empirically observed and experimentally measured patterns of human prosociality that cannot easily be accounted for by standard evolutionary models, the culture–gene coevolutionary approach solves two other problems in understanding cooperation and punishment that have not even been addressed by standard explanations.

Problem 1. If large-scale cooperation is a product of purely natural selection acting on genes to favor something such as indirect reciprocity, why is not large-scale cooperation more wide-spread in nature? Why do not other social species like baboons, sea lions, chimpanzees, and dolphins (which show evidence of kinship and maybe reciprocity) cooperate at human levels? The above approach, by rooting the development of large-scale cooperation in the details of human social learning, addresses this challenge. Other mammals do not cooperate to the degree humans do because they lack the social learning abilities that produce cultural evolution and behavioral equilibria not available to genetic transmission alone. Of course, this answer only pushes the question back to why more animals do not have human-style cultural capacities. In answering this question, Boyd and Richerson (1996) have shown that there is an adaptive valley in the evolution of cultural abilities that inhibits the spread of these abilities when rare. Thus, such cultural abilities should be rare in nature, but once a species crosses the 'cultural threshold', whole new evolutionary vistas open up (also see Henrich and Gil-White, 2001).

Problem 2. Unlike other animals, human cooperation varies in both scale and behavioral domains across social groups. As I mentioned, many social groups inhabit the same physical environment and possess the same technology, but cooperate to differing degrees and in different domains (e.g., the Nuer and Dinka). Standard evolutionary approaches struggle

with these uniquely human observations—in other animals, cooperation and sociality do not vary much from group to group within a species. However, the coevolutionary approach suggests that rapid within-group cultural forces will drive different social groups to one of a myriad of different quasi-stable equilibria, where they may remain for long periods until the between-group component of cultural evolution sorts among the various equilibria. In selecting among alternative equilibria, cultural group selection processes may operate on much longer time scales than within-group processes, especially when alternative equilibria have similar effects on cultural group fitness. The rates of cultural evolution under group selection (decades, centuries and millennia) have about the right time scale to explain human history (Diamond, 1997), whereas genetic explanations are too slow and rational choice ones too fast.

Acknowledgements

References

Alexander, R., 1974. The evolution of social behavior. Annual Review of Ecology and Systematics 5, 325–383.

Alexander, R., 1987. The Biology of Moral Systems. Aldine De Gruyter, New York.

Aoki, K., 1982. A condition for group selection to prevail over counteracting individual selection. Evolution 36 (4), 832–842.

Arnold, J., 1996. The archaeology of complex hunter-gatherers. Journal of Archaeological Method and Theory 3 (2), 77–126.

Asch, S.E., 1951. Effects of group pressure upon the modification and distortion of judgments. In: Guetzkow, H. (Ed.), Groups, Leadership and Men. Carnegie, Pittsburgh, pp. 177–190.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, US.

Axelrod, R., 1986. An evolutionary approach to norms. American Political Science Review 80 (4), 1095–1111.

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

Axelrod, R., Dion, D., 1988. The further evolution of cooperation. Science 242, 1385–1389.

Ball, S., Eckel, C., Grossman, P., Zame, W., 2001. Status in markets. Quarterly Journal of Economics 155 (1), 161–181.

Baron, R., Vandello, J., Brunsman, B., 1996. The forgotten variable in conformity research: impact of task importance on social influence. Journal of Personality and Social Psychology 71 (5), 915–927.

Batson, D., Shaw, L., 1991. Evidence for alturism: toward a pluralism of prosocial motives. Psychological Inquiry 2 (2), 107–122.

Bendor, J., Mookherjee, D., 1987. Institutional structure and the logic of ongoing collective action. The American Political Science Review 81 (1), 129–154.

Ben-Ner, A., Putterman, L., 2000. Some implications of evolutionary psychology for the study of preferences and institutions. Journal of Economic Behavior and Organization 43 (1), 91–99.

Bliege Bird, R., Bird, D.W., 1997. Delayed reciprocity and tolerated theft: the behavioral ecology of food-sharing strategies. Current Anthropology 38 (1), 49–78.

Bolton, G.E., Katok, E., Zwick, R., 1998. Dictator game giving: rules of fairness versus acts of kindness. International Journal of Game Theory 27, 269–299.

Boorman, S., Levitt, P., 1980. The Genetics of Altruism. Academic Press, New York.

Bowles, S., 2001. Individual interactions, group conflicts, and the evolution of preferences. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. MIT Press, Cambridge, MA, pp. 155–190.

Boyd, R., 1988. Is the repeated prisoner's dilemma a good model of reciprocal altruism? Ethology and Sociobiology 9, 211–222.

Boyd, R., 1989. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. Journal of Theoretical Biology 136, 47–56.

Boyd, R., Lorderbaum, J.P., 1987. No pure strategy is evolutionarily stable in the repeated Prisoner's dilemma game. Nature 32 (6117), 58–59.

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

Boyd, R., Richerson, P.J., 1987. The evolution of ethnic markers. Cultural Anthropology 2 (1), 27–38.

Boyd, R., Richerson, P.J., 1988. The evolution of reciprocity in sizable groups. Journal of Theoretical Biology 132, 337–356.

Boyd, R., Richerson, P.J., 1989. The evolution of indirect reciprocity. Social Networks 11 (3), 213–236.

Boyd, R., Richerson, P.J., 1990. Group selection among alternative evolutionarily stable strategies. Journal of Theoretical Biology 145, 331–342.

Boyd, R., Richerson, P.J., 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology 13 (3), 171–195.

Boyd, R., Richerson, P., 1996. Why culture is common, but cultural evolution is rare. Proceedings of the British Academy 88, 77–93.

Boyd, R., Richerson, P.J., 2000. Memes: universal acid or a better mouse trap. In: Aunger, R. (Ed.), Darwinizing Culture: The Status of Memetics as a Science. Oxford University Press, Oxford, pp. 143–162.

Boyd, R., Richerson, P., 2002a. Group beneficial norms can spread rapidly in a structured population. Journal of Theoretical Biology 215, 287–296.

Boyd, R., Richerson, P., 2002b. Solving the puzzle of human cooperation. In: Levinson, S. (Ed.), Evolution and Culture. MIT Press, Cambridge, MA.

Boyd, R., Silk, J., 1997. How Humans Evolved. W.W. Norton & Company, New York.

Boyd, R., Gintis, H., Bowles, S., Richerson, P.J., 2003. The evolution of altruistic punishment. PNAS 100 (6), 3531–3535.

Brown, R., Armelagos, G., 2001. Apportionment of racial diversity: a review. Evolutionary Anthropology 10, 34–40.

Brown, W.M., Moore, C., 2002. Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: an evolutionary analysis. Advances in Psychological Research 11, 59–78.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.

Crow, J., Aoki, K., 1982. Group selection for a polygenic behavioral trait: a differential proliferation model. Proceedings of the National Academy of Sciences of the United States of America 79 (8), 2628–2631.

Daly, M., Wilson, M., 1988. Homicide. Aldine de Gruyter, New York.

Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press, Princeton, NJ.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, Oxford.

Dawkins, R., 1982. The Extended Phenotype. Oxford University Press, Oxford.

Diamond, J.M., 1997. Guns, Germs, and Steel: The Fates of Human Societies. W.W. Norton & Company, New York.

Ekman, P., 1992. Telling Lies. Norton, New York.

Ekman, P., O'Sullivan, M., 1991. Who can catch a liar? American Psychologist 46, 913–920.

Ellison, G., Fudenberg, D., 1993. Rules of thumb for social learning. Journal of Political Economy 101 (4), 612–643.

Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90 (4), 980–995.

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

Fehr, E., Henrich, J. (forthcoming). Is strong reciprocity a maladaption? In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge, MA.

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

Fletcher, D.J.C., Michener, C.D., 1987. Kin Recognition in Animals. Wiley, New York.

Frank, R., 1988. Passions within Reason: The Strategic Role of the Emotions. W.W. Norton & Company, New York.

Frank, S., 1995. George price's contributions to evolutionary genetics. Journal of Theoretical Biology 175, 373–388.

Frank, S.A., 1998. Foundations of Social Evolution. Princeton University Press, Princeton, NJ.

Frank, R., Gilovich, T., Regan, D.T., 1993. The evolution of one-shot cooperation: an experiment. Ethology and Sociobiology 14, 247–256.

Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–556.

Furnham, A., 1984. Studies of cross-cultural conformity: a brief and critical review. Psychologia: An International Journal of Psychology in the Orient 27 (1), 65–72.

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

Gintis, H., Smith, E.A., Bowles, S., 2001. Costly signaling and cooperation. Journal of Theoretical Biology 213 (1), 103–119.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. Journal of Theoretical Biology 7, 1–16.

Hamilton, W.D., 1972. Altruism and related phenomena, mainly in the social insects. Annual Review of Ecology and Systematics 3, 192–232.

Hamilton, W.D., 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox, R. (Ed.), Biosocial Anthropology. Malaby Press, London, pp. 133–156.

Harris, M., 1977. Cannibals and Kings: The Origins of Cultures. Random House, New York.

Harris, M., 1979. Cultural Materialism: The Struggle for a Science of Culture. Random House, New York.

Harris, J.R., 1998. The Nurture Assumption: Why Children Turn Out the Way They Do. Touchstone, New York.

Hartl, D.L., Clark, A.G., 1989. Principles of Population Genetics. Sinauer Associates, Sunderland, MA.

Hayek, F., 1979. The political order of a free people. Law, Legislation and Liberty, vol. 3. University of Chicago Press, Chicago.

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. American Anthropologist 103, 992–1013.

Henrich, J., Boyd, R., 1998. The evolution of conformist transmission and the emergence of between-group differences. Evolution and Human Behavior 19, 215–242.

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

Henrich, J., Boyd, R., 2002. On modeling cultural evolution: why replicators are not necessary for cultural evolution. Journal of Cognition and Culture 2 (2), 87–112.

Henrich, J., Gil-White, F., 2001. The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior 22 (3), 165–196.

Henrich, J., Smith, N., in press. Comparative experimental evidence from Peru, Chile and the US shows substantial variation among social groups. In: Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C. (Eds.), Cooperation, Punishment and Self Interest: Experimental and Ethnographic Evidence from 15 Small Scale Societies. Oxford University Press, New York, and University of Michigan.

Henrich, J., Boyd, R., Bowles, S., Gintis, H., Camerer, C., Fehr, E., McElreath, R., 2001. In search of Homo economicus: experiments in 15 small-scale societies. American Economic Review 91, 73–78.

Hirshleifer, J., 1982. Evolutionary models in economics and law: cooperation vs. conflict strategies. Research in Law and Economics 4, 1–60.

Hirshleifer, J., Martinez, C., 1988. What strategies can support the evolutionary emergence of cooperation. Journal of Conflict Resolution 32 (2), 367–398.

Hirshleifer, D., Rasmusen, E., 1989. Cooperation in a repeated prisoner's dilemma with ostracism. Journal of Economic Behavior and Organization 12, 87–106.

Hoffman, E., McCabe, K., Smith, V., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Economic Inquiry 36, 335–352.

Insko, C.A., Smith, R.H., Alicke, M.D., Wade, J., Taylor, S., 1985. Conformity and group size: the concern with being right and the concern with being liked. Personality and Social Psychology Bulletin 11 (1), 41–50.

Johnson, A., 2003. Families of the Forest: Matsigenka Indians of the Peruvian Amazon. University of California, Berkeley, CA.

Johnson, A., Earle, T., 1987. The Evolution of Human Societies: From Foraging Group to Agrarian State. Stanford University Press, Stanford.

Joshi, N.V., 1987. Evolution of cooperation by reciprocation within structured demes. Journal of Genetics 66 (1), 69–84.

Kagel, J.H., Roth, A.E., 1995. The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ.

Keller, M., Edelstein, W., Schmid, C., Fang, G., Fang, F.-X., 1998. Reasoning about relationships and obligations in close relationships: a comparison across two cultures. Developmental Psychology 34 (4), 731–741.

Kelly, R.C., 1985. The Nuer Conquest. University of Michigan Press, Ann Arbor.

Kelly, J.K., 1992. Restricted migration and the evolution of altruism. Evolution 46 (5), 1492–1495.

Klein, R.G., 1989. The Human Career: Human Biological and Cultural Origins. University of Chicago, Chicago.

Kroll, Y., Levy, H., 1992. Further tests of the separation theorem and the capital asset pricing model. American Economic Review 82 (3), 664–670.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ, pp. 111–194.

Lee, R.B., 1979. The !Kung San: Men, Women, and Work in a Foraging Society. Cambridge University Press, Cambridge.

Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. Proceedings: Biological Sciences 268 (1468), 745–753.

Lorberbaum, J., 1994. No strategy is evolutionarily stable in the repeated prisoner's dilemma. Journal of Theoretical Biology 168, 117–130.

Lorberbaum, J.P., Bohning, D.E., Shastri, A., Sine, L., 2002. Are there really no evolutionary stable strategies in the iterated prisoner's dilemma. Journal of Theoretical Biology 214, 155–169.

Low, B.S., 2000. Why Sex Matters: A Darwinian Look at Human Behavior. Princeton University Press, Princeton, NJ.

Maynard Smith, J., 1998. The origin of altruism. Nature 393 (18), 639–640.

McAdams, R.H., 1998. The origin, development and regulation of norms. Michigan Law Review 96, 338.

McCabe, K., Smith, V., LePore, M., 2000. Intentionality detection and "mindreading": why does game form matter? Proceedings of the National Academy of Science 97 (8), 4404–4409.

McElreath, R., Boyd, R., Richerson, P., 2003. Shared norms can lead to the evolution of ethnic markers. Current Anthropology 44 (1), 123–129.

Neto, F., 1995. Conformity and independence revisited. Social Behavior and Personality 23 (3), 217–222.

Nowak, M.A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393 (6685), 573–577.

Ockenfels, A., Selten, R., 2000. An experiment on the hypothesis of involuntary truth-signaling in bargaining. Games and Economic Behavior 33 (1), 90–116.

Offerman, T., Sonnemans, J., 1998. Learning by experience and learning by imitating others. Journal of Economic Behavior and Organization 34 (4), 559–575.

Patton, J.Q., 2000. Reciprocal altruism and warfare: a case from the Ecuadorian Amazon. In: Cronk, L., Chagnon, N., Irons, W. (Eds.), Adaption and Human Behavior: An Anthropological Perspective. Aldine de Gruyter, New York, pp. 417–436.

Pingle, M., 1995. Imitation vs. rationality: an experimental perspective on decision-making. Journal of Socio-economics 24, 281–315.

Pingle, M., Day, R.H., 1996. Modes of economizing behavior: experimental evidence. Journal of Economic Behavior and Organization 29, 191–209.

Plomin, R., Defries, J., McLearn, G. E., 2000. Behavioral Genetics. Freeman, New York.

Price, G., 1970. Selection and covariance. Nature 227, 520–521.

Price, G., 1972. Extensions of covariance selection mathematics. Annals of Human Genetics 35, 485–490.

Queller, D.C., 1992a. A general model for kin selection. Evolution 46 (2), 376–380.

Queller, D., 1992b. Quantitative genetics, inclusive fitness, and group selection. The American Naturalist 139 (3), 541–558.

Rappaport, R.A., 1984. Pigs for the Ancestors. Yale University Press, New Haven.

Richerson, P., Boyd, R., 1998. The evolution of ultrasociality. In: Eibl-Eibesfeldt, I., Salter, F.K. (Eds.), Indoctrinability, Ideology and Warfare. Berghahn Books, New York, pp. 71–96.

Richerson, P., Boyd, R., 2000. Complex societies: the evolutionary dynamics of a crude superorganism. Human Nature 10, 253–289.

Ridley, M., 1993. The Red Queen: Sex and the Evolution of Human Nature. Penguin Books, New York.

Rogers, A., 1990. Group selection by selective emigration: the effects of migration and kin structure. The American Naturalist 135 (3), 339–413.

Rogers, E.M., 1995. Diffusion of Innovations. Free Press, New York.

Roscoe, P., 2002. The hunters and gatherers of new guinea. Current Anthropology 43 (1), 153–162.

Rosekrans, M.A., 1967. Imitation in children as a function of perceived similarity to a social model and vicarious reinforcement. Journal of Personality and Social Psychology 7 (3), 307–315.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study, Pittsburgh and Tokyo: an experimental study. American Economic Review 81 (5), 1068–1095.

Sepher, J., 1983. Incest, the Biosocial View. Academic Press, New York.

Smith, J.M., Bell, P.A., 1994. Conformity as a determinant of behavior in a resource dilemma. Journal of Social Psychology 134 (2), 191–200.

Sober, E., Wilson, D.S., 1998. Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge.

Soltis, J., Boyd, R., Richerson, P.J., 1995. Can group-functional behaviors evolve by cultural group selection? An empirical test. Current Anthropology 36 (3), 473–494.

Spencer, H., 1891. Essays: Scientific, Political, and Speculative. Williams & Northgate, London.

Stotland, E., Dunn, R., 1962. Identification, "oppositeness," authoritarianism, self-esteem, and birth order. Psychological Monographs: General and Applied 76 (9), 1–30.

Stotland, E., Dunn, R., 1963. Empathy, self-esteem and birth order. Journal of Abnormal and Social Psychology 66 (6), 532–540.

Tomasello, M., 1994. The question of chimpanzee culture. In: Wrangham, R., McGrew, W.C., de Waal, F.B.M., Heltne, P.G. (Eds.), Chimpanzee Cultures. Harvard University Press, Cambridge, MA, pp. 301–317.

Tomasello, M., 2000. The Origins of Human Cognition. Harvard University Press, Cambridge.

Tooby, J., Cosmides, L., 1989. Evolutionary psychology and the generation of culture: theoretical considerations. Ethology and Sociobiology 10, 29–49.

Trivers, R.L., 1971. The evolution of reciprocal alturism. The Quarterly Review of Biology 46, 34–57.

Turner, J., Maryanski, A., 1979. Functionalism. Benjamin/Cummings, Menlo Park, CA.

Vayda, A.P., 1971. Phases of the process of war and peace among the Marings of new Guinea. Oceania 42, 1–24.

Wade, M., 1985. Soft selection, hard selection, kin selection, and group selection. The American Naturalist 125 (1), 61–73.

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

Weinmann, J., 1994. Individual behavior in a free riding experiment. Journal of Public Economics 54 (2), 185–200.

Westermarck, E., 1894. The History of Human Marriage. Macmillan, London.

Wiessner, P., 1983. Style and social information in kalahari san projectile points. American Antiquity 48 (2), 253–275.

Williams, G., 1966. Adaption and Natural Selection: A Critique of Some Current Evolutionary Theory. Princeton University Press, Princeton, NJ.

Wilson, D.S., Dugatkin, L., 1997. Group selection and assortative interactions. The American Naturalist 149 (2), 336–351.

Wit, J., 1999. Social learning in a common interest voting game. Games and Economic Behavior 26, 131–156.

Wolf, A.P., 1970. Childhood association and sexual attraction: a further test of the Westermarck hypothesis. American Anthropologists 72, 503–515.

Young, D., Bettinger, R.L., 1992. The numic spread: a computer simulation. American Antiquity 57 (1), 85–99.

JOURNAL OF
Economic Behavior
& Organization

Discussion

Comments

Gordon Tullock*

3301 N Fairfax Drive, Arlington, VA 22201, USA

Henrich's contribution (it is too long to be called an article and, indeed, is almost a book) is excellent, combining as it does a general survey of the existing literature together with original contributions. My objective here is not to directly comment on his work, which, to repeat, I find excellent, but to add a few additional points. These points are more speculative than his work and being heavily pushed for time, I will not fully footnote them. I merely want to add some things which may help future researchers. Since this is done rather hastily, I certainly hope I will help them and not lead them down false trails.

The first point is the green beards. There is a sort of green beard in both primate and human behavior. Suppose that chimpanzees have a hereditary gene which leads them to occasionally beg for food from other members of their band. They also have a gene to at least occasionally respond when food is begged from them by giving some food. The begging for the food is the green beard in that it would be possible for the animal to have a gene for begging for and accepting food while another gene led to never giving food to another member of the band who asked for it. This combination of genes, however, would be genetically dominated and eventually replaced by a gene which instructed their bodies to make modest gifts to other members of the band who beg for them, but only if the recipient of the gift did not refuse to make gifts under similar circumstances.

Human social life is an example of this kind of behavior. I buy your lunch today and expect you to buy mine tomorrow. There are many other cases where casual social life would cease if reciprocal kindnesses were denied. Apparently chimpanzees are genetically driven to behave in the same way.

But let me turn to a more exciting subject: war. A number of our relatives live in tribal groups with specific territory under their control. Territoriality is quite common in the animal kingdom. Indeed there seem to be plants that to some extent exercise dominion over the land around them. That a band of chimpanzees would have a specific territory which they protected and mainly stayed inside is not surprising. Monopolizing resources of an

_____
* Tel.: +1-703-993-4908; fax: +1-703-993-4935.
E-mail address: gtulloc1@gmu.edu (G. Tullock).

area sufficient to support the band, even in a bad year, presumably would be sensible. And to repeat, territoriality is not particularly uncommon among animal species. Intrusion by other members of the same specie is resented and resisted. Normally this resistance simply involves jumping up and down and shouting, but it sometimes leads to genuine fighting. Jane Goodall saw a case in which a tribe of 11 chimpanzees were systematically hunted down and killed by the larger tribe from which they had broken off. So far as I know there are no other cases of similar genocide but chimpanzees and Bonobos frequently carry the scars of battle on their persons. Presumably the scars will mainly be earned in intra-tribal squabbles over mates and food but some may have been the honorable scars of intertribal warfare.

If it is true, however, that the tribe will occasionally have major squabbles over real estate, why would not individuals hold back and thus preserve themselves from injury while sharing the spoils of victory. Of course there would not be many victories if such genes became common, but that is not the kind of argument normally accepted in evolutionary discussions. That an individual who held back would be punished by the others raises the similar problem of who will punish the non-punisher? Given the amount which has been written on this topic, I will leave it here, but simply point out that it does not solve the problem in the view of most evolutionary biologists.

I have however another solution. Presumably courage is gene controlled. If we simply assume that courage is courage, and is not courage in one situation combined with a tendency to flee in another, then the problem is solvable. There are many disputes within the tribe, partially over food, but more commonly over status in the tribe or mates. A brave chimpanzee would tend to take a leading roll in intertribal warfare as well as the internal disputes over, say, mates. Thus a courage gene would have survival potential and would mean that those bands which have many such genes will tend to prevail in intertribal warfare. From the economic standpoint, a big male entering into combat in a leading role is generating a favorable externality for the other members of the tribe at a cost to him. If however the same gene also provides for courage in inter-tribal fighting it may have both individual and group advantages.

Once again the same problem arises with human beings. The draft dodger has a higher likelihood of surviving to old age, but if there are many of them his society has a lower probability of survival. Most societies have recognized this and provided private incentives for joining in the armies of the society. I myself was drafted, and I can assure you I would not have volunteered. Nonetheless there were volunteers. Perhaps the number of genuine volunteers was greater than that of the draft dodgers. Whether this coincided with success in perpetuating their genes I do not know, but at least in common parlance it is thought that "all the girls love a soldier". With us, in this case, as in my hypothetical discussion of the chimpanzees, courage in battle carries with it courage in internal squabbles with a result that the genes are perpetuated in spite of the casualty rate in battle.

The willingness of people to take risks for their "tribe" even if there is no formal governmental organization to force them into it was illustrated in the breakup of Yugoslavia. Most of the constituent states were able to obtain independence without serious fighting; Bosnian however was not a single ethnic group but a set of three living more or less separately. The Serbs had inherited most of the formal organization of Yugoslavia and certainly almost all of the arms. They decided to expand the Serbian part of Bosnia, apparently intending to

take over the whole of it. They were opposed by a disorganized group of volunteers with, at first almost no weapons. American newspaper readers may recall the desperate situation of the Bosnians in the early part of a war. They were defended mainly by pickup groups of volunteers with poor or no weapons. Naturally the Serbian regulars with their better equipment were, on the whole, successful, but they did not win a total victory and today hold only about a third of Bosnia. This I think is a good illustration of human beings responding to attack on their tribe more or less spontaneously. I certainly hope that the genes of the soldiers who defended the Bosnian area will be perpetuated.

I would like to turn now to a related area although not directly involved in Henrich's paper. This is the rather peculiar behavior of almost all human beings in making unsolicited gifts to people living far away. If there is an earthquake in southern Mexico or a typhoon strikes in the coast of South China and this gets publicity in the United States many Americans will send gifts to the people injured usually through something like the red cross. It is extremely hard to see how this would promote the survival of genes in the Americans. Since Dawkins' most persuasive book, most biologists have thought that it is a gene that is selfish not the entity bearing it. To quote the standard example I would risk my life to save my brother if his likelihood of death was certain if I did not, and the risk to my life was less than 50 percent. The gene which directed me to do this would be statistically more frequent in the population after I had so acted then it would have been had I not done so. Thus the action serves the selfish need of the gene but not of its human bearer.

That we would make sacrifices for relatives who have varying chances of bearing our genes is thus not surprising. Further it seems likely that the prehumans and early humans lived in bands much like modern chimpanzees. If so each one would have some degree of relationship to other members of the tribe and hence genes ordering various levels of help would have survival value for the gene. All of the primates who live in bands of some size practice exogamy and it seems likely that early humans did so too. Incest is likely to lead to generic difficulties. It is even possible to argue that our incest taboos are survivals of much the same phenomenon. But if the tribes practice exogamy, then it would follow that an individual would have at least some gene overlap with individuals in the next tribe over. Thus a gene which provided lesser aid to individuals in the next tribe would also have survival value. Further the next tribe no doubt practices exogamy so that there would be a weaker relationship, but nevertheless a real one, with the tribe lying beyond it. The likelihood that any given gene that one individual holds would be duplicated in another individual in another tribe would gradually decline as the geographical distance increases. We should not then be surprised at individuals helping members of their local community, their relatives even more so, but to a smaller extent people living farther away who have less likelihood of bearing the same gene.

If we consider primitive conditions, however, knowledge of other tribes would fall off rapidly as you move away. There would be no prospect of the individual tribesman even knowing about people in India. Thus we would anticipate that this kind of charitable aid would eventually reach a low level, but stay at that level as you move even farther away. There will be no reason for genetic control for cutting out people who are so far away you will never have any knowledge of them. Thus my gift to typhoon survivors in China is explained by the fact that our far distant ancestry did not know such people existed and hence the genes did not cut off charitable gifts to people who were beyond their cognizance.

This is not strictly speaking a modification of Henrich's article but I think it is in the spirit of his work. In any event it is clear that we do engage in cooperation and charity when it is not to the direct benefit of the individual engaging in the activity. Henrich deals in this subject in a very competent way. I have added a little bit and I hope that many other people will contribute too. This is an area where we need information badly. We have a lot of knowledge already, but I think that everyone will agree that we need more.

Discussion

Cultural group selection, coevolutionary processes, and large-scale cooperation: a comment

Richard B. Norgaard*

Energy and Resources Group, 310 Barrows Hall, University of California at Berkeley, Mail Code 3050, Berkeley, CA 94720-3050, USA

Joseph Henrich provides us with a very thoughtful, critical review of the literature on gene-culture coevolution with respect to the existence and maintenance of cooperation and sociality in general. He also presents us with an original argument for introducing cultural group selection to explain the persistence of individual behaviors that favor the cultural evolution of prosociality. I am sure that his critical analysis will advance this literature significantly.

At the same time, I remain generally puzzled by the literature on gene-culture coevolution. To argue that people evolve at a cultural level, as well as at the genetic level, and that the two levels coevolve is a major step from the earlier literature that tied all traits, biological and social, to genetic evolution alone. Henrich, by bringing in group selection, takes one more step away from the modern synthesis of evolutionary and genetic thinking that arose during the 1930s and "hardened," according to Gould (2002, Chapter 7), by the 1960s. Having taken these steps, I am curious why the tie to biological evolution still remains so strong in the explanation of social phenomena.

I am very comfortable with the concept that cultures are made up of ideas (or memes, idenes, sociogenes, culturgens, etc.) that reproduce, though imperfectly, that people think up new ideas, that new ideas move between cultures, and that some ideas prove less fit than others. What puzzles me is why the scholars working in this literature see the process as so symmetrical. Culture is treated as a single unit, like an organism, that evolves directly while also coevolving with genes. The symmetry keeps the arguments tight and facilitates clarity in thinking. Yet, I have difficulty thinking of culture as a single evolutionary unit, and I see no reason why evolutionary thinking must always be tied so closely to genes.

Might culture be divisible into multiple units with each unit evolving on its own while coevolving with each other? In my own work (Norgaard, 1994), I have argued that modern

---

* Tel.:+1-510-642-3465; fax: +1-510-642-1085.
E-mail address: norgaard@socrates.berkeley.edu (R.B. Norgaard).

culture can be divided into value systems, knowledge systems, organizational systems, and technological systems. Each of these can be thought of as reproducing itself imperfectly, experiencing innovations, and being influenced through introductions from other cultures. At the same time, each of these systems puts selective pressure on the other systems. For my own work, I found no reason to delve into whether these coevolutionary processes within culture affected human genetic evolution. The rates of change are quite different. Allowing for asymmetry between genes and culture, allowing evolutionary processes to take place within and between aspects of culture, weakens the ties to our genes while opening up our thinking to evolutionary explanations of why things social are both so tightly interlocked and yet can change.

Genes are tied to the organism, a structure, making an evolutionary unit. In the gene-culture literature, culture is without adequate structure, it has too little to attach the ideas to. As we divide culture up, however, structure emerges. What is the organismal equivalent, the structure, of culture? It is a lot easier to see the structures of our value systems, our knowledge systems, our organizational systems, and our technological systems, to see the structure that is being reproduced and the processes of reproduction. Our educational and research system is clearly made up of schools and universities and funding agencies and the National Academy of Sciences, and we can see how they reproduce themselves while also changing. Capitalism, markets, and democracy are "forms" of social organization that reproduce themselves while also changing. And we can see how they coevolve. Bernstein (2001), for example, illustrates how the rise of economics as an academic discipline in our knowledge system has coevolved with social organization and vice versa. Of course, these cultural subdivisions are still fuzzy, but they are certainly not as fuzzy as culture as a whole. And it is good to keep in mind that species as a category also has fuzzy edges, increasingly so as we move to microbial life.

To speak of culture as a unit is analogous to speaking of all life, to amass all of the plants and animals together. We see the structure of life when we look at species separately, it is the parts that evolve, largely in response to each other, and in this sense life as a whole evolves Ditto culture. And as we accept the evolution and coevolution of subsystems of culture, the explanation of cooperation and prosociality in general becomes much more social, much less genetic, but no less evolutionary.

I suspect the differences in the perspective that I am taking and that of the gene-culture research community are historic. The gene-culture literature has taken a long view. It has strong roots in anthropology, most notably the work of Boyd and Richerson (1985 and numerous works since) and Durham (1991). It is harder to see the structures of historic societies or the structural units of cultures clearly different from our own. Anthony Giddens (1990) argues that modern society is special for both its highly developed structures and that the separateness of these structures leads to fragmented lives. The gene-culture literature also has strong roots in the works of geneticists including Cavalli-Sforza and Feldman (1981) and in evolutionary biology, especially the work of Lumsden and Wilson (1981). These founders of the literature moved beyond biology, but they certainly were not in a strong position to characterize culture in a very interesting way. Thus, treating organisms and culture as symmetric can be seen as an artifact of the actors, script, and stage at hand. I sense, however, an additional historic reason. Gene-culture theorists have crossed disciplinary boundaries, for some no doubt at considerable cost. Treating genes and culture symmetrically may

have facilitated crossing disciplinary boundaries by closing off the morass of contradictory theory and evidence we have on social systems. If this is also the case, then the symmetrical construct has been at the cost of closing off the larger application of evolutionary thinking in the social sciences. Let us hope this too will soon be historic.

References

Bernstein, M., 2001. A Perilous Progress: Economists and Public Purpose in Twentieth-Century America. Princeton University Press, Princeton, NJ.

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

Cavalli-Sforza, L., Feldman, M., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton, NJ.

Durham, W., 1991. Coevolution: Genes, Culture, and Human Diversity, Stanford University Press. Stanford, CA.

Giddens, A., 1990. The Social Consequences of Modernity. Stanford University Press, Stanford, CA.

Gould, S. J., 2002. The Structure of Evolutionary Theory. Harvard University Press, Cambridge, MA.

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

Norgaard, R.B., 1994. Development Betrayed: The End of Progress and a Coevolutionary Revisioning of the Future. Routledge, London.

Discussion

How to measure social interactions via group selection? Cultural group selection, coevolutionary processes, and large-scale cooperation: a comment

Yuji Aruka*

Faculty of Commerce, Chuo University, 742-1 Higashinakano Hachioji-shi, Tokyo 192-0393, Japan

Abstract

It is interesting to know that Henrich's skillful challenge of the Price equation of biology in Price [Nature 227 (1970) 520; Annals of Human Genetics 35 (1972) 485] to locate large-scale cooperation of human beings certainly gives a straight way to argue social interactive mechanism in economic theory. In fact, Glaeser and Scheinkman [E.L. Glaeser, J.A. Scheinkman, Measuring social interactions, in: S.N. Durlauf, H.P. Young (Eds.), Social Dynamics, Brooking Institute, Washington, DC, 2001, pp. 83–131] exhibits a limited form of the Price equation in economic theory without any explicit reference to it. This suggests a common analytical way to measure interactive effects of between-groups and within-group whether in biology or economics. This way however is to be accompanied by some difficulties. Some overarching mechanism between between-groups and within-group will be necessary to establish a definite direction of group selection. The idea of Hildenbrand [W. Hildenbrand, Market Demand, Princeton University Press, Princeton, 1994] may be helpful to do this, although his trial has been done in a qualitatively different field.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Heterogeneities; Social interactions; Macroscopic order; Overarching mechanisms

* Tel.: +81-426-74-3617; fax: +81-426-74-3651.
E-mail address: aruka@tamacc.chuo-u.ac.jp (Y. Aruka).

1. Social interactions in utility theory

1.1. New issues in utility function

We begin to notice that Henrich's insights of prosocial genes under the cultural transmission mechanism could in general be compatible with a new understanding of evolutionary adaptive economic systems.

A new movement to incorporate social interaction in utility functions of economics has notably begun during these days. Some economists are working to capture the effects of social interaction among the agents. It is not difficult for us to mention a series of new attempts to introduce social interactions and heterogeneities of agents. One set may come about from the random preferences approach given by Hildenbrand (1971), Föllmer (1974), followed by Durlauf (1997, 2000) who borrows the idea of Gibbs distribution in statistical mechanics; while another set is of Weidlich and Haag (1983), and Helbing (1995) who are concerned with application of master equations to social interactions. Aoki (1996) also is to be mentioned.

1.2. Cooperative systems under a macroscopic order

In these models, a macroscopic field on which particular agents interact could bind themselves by the field itself. In other words, some positive feedback (i.e. reinforcement mechanism, passing through a critical point of social situations) may give a limiting system of order, which might be a cooperative system. One type of reinforcement mechanism can come towards bringing cooperation into play via copying or imitating the reference agents (their neighbors), as Arthur (1994) has argued. A source of copying behavior by individuals may come from learning. Voluntary learning by individuals can however not always lead a group to a cooperative situation. Glaeser and Scheinkman (2001, p. 94), similarly as in Henrich, has referred to social interaction through punishment mechanism:

> Social interactions occur because through learning this ideal behavior, individual influences each other. Natural examples of this type of effect occur in crowd behavior where individuals seem to completely forego what is commonly thought of as civilized behavior because they are sanctioned by the crowd (see, e.g. the extensive literature on the motivation of Nazis).

1.3. Sub-grouping and the memes of subgroups

Another important feature in social interaction of agents is found in partitioning or labeling individuals into their sub-groups. Group selection through its cultural transmission is relevant to the concept of meme or social norms. In this context, Friedman and Sighn (2001, pp. 106–107) argued vengeance as a negative reciprocity:

> All known groups of humans maintain memes that prescribe appropriate behavior toward fellow group members, and typically prescribe different appropriate behavior toward individuals outside the group. . . . The success of the meme, as with any other adaptive

unit, is measured by its ability to displace alternatives, by its fitness. ... Groups affect individual fitness in several ways. ... they provide gains from internal cooperation and some gains from external cooperation.

2. A simple model of local interactions

2.1. Glaeser and Scheinkman

With the above comprehensive background in line with Henrich's stated circumstances, Glaeser and Scheinkman (2001) gives a simple model of local interactions for their practical purpose to measure social interactions, although in a very restricted model.[1]

People respectively live in different cities. θ is a distribution of quality across people, which has mean zero. Its variance is denoted σ_θ^2. θ_i is the individual i's quality, assumed to be constant across cities. X_i represents individual-level characteristics like age or gender. The idiosyncratic tastes of individual i may then be written:

$$\Theta_i = \theta_i + f(X_i) \tag{1}$$

"[T]he marginal utility of the action for individual i is directly by an idiosyncratic taste shock Θ_i, and his neighbor's action A_{i-1}." (Glaeser and Scheinkman, 2001, p. 95):

$$U(A_i, A_{i-1}, \Theta_i) = \Theta_i A_i - \frac{1}{2}(1-\alpha)A_i^2 - \frac{1}{2}\alpha(A_i - A_{i-1})^2 \tag{2}$$

More specifically, the actions space is supposedly restricted on the real line. Agents may live in their respective cities j on the real line. The choice of action by individual i is based either on his own taste for the action or by imitating his predecessor $i-1$ on the domain.[2]

Individual i chooses an action A_i based on their idiosyncratic tastes Θ_i, and their predecessors' action level A_{i-1}. The individual i's action A_i thus is defined:

$$A_i = \theta_i + f(X_i) + \alpha A_{i-1} \tag{3}$$

Let $\overline{A}_j$ be the mean action level in city j. Equivalently:[3]

$$A_i - \overline{A}_j = \theta_i + f(X_i) - \overline{f(X)}_j + \alpha(A_{i-1} - \overline{A}_j) \tag{4}$$

Here α is regarded as the regression coefficient of A_i with respect to A_{i-1}. It is also assumed that "there is no sorting across neighborhoods within cities." (Glaeser and Scheinkman, 2001, p. 95) Notice that offsprings of i never break out in this line.

2.2. Local interactions

We can then calculate the variance of $A_i - \overline{A}_j$. In equilibrium, it holds:

$$\mathrm{Var}(A_i - \overline{A}_j) = \mathrm{Var}(A_{i-1} - \overline{A}_j) \tag{5}$$

---

[1] The next two subsections are a brief summary on Section 3 of Glaeser and Scheinkman (2001, pp. 94–100).

[2] The benefit from copying their predecessor will be measured by $(1-\pi)$.

[3] $\overline{f(X)}$ denotes the mean level of the function $f(X)$; $\overline{f(X)}_j$ denotes the mean level of the function $f(X)$ in city j.

$\text{Var}_j^{f(X)}$ represents the variance of $f(X)$ within city j.[4] In addition, $\text{Var}_j^{f(X)}$ is assumed to be constant across cities:

$$\text{Var}(A_i - \overline{A}_j) = \underbrace{\sigma_\theta^2 + \text{Var}_j^{f(X)}}_{\text{within-city}} + \underbrace{\alpha^2 \, \text{Var}(A_{i-1} - \overline{A}_j)}_{\text{between-cities}} = \frac{\sigma_\theta^2 + \text{Var}_j^{f(X)}}{1 - \alpha^2}. \qquad (6)$$

In the limit, Glaeser and Scheinkman (2001) gives the regression coefficient:

$$\alpha = \frac{\text{Var}_{\text{agg}} - \text{Var}_{\text{ind}}}{\text{Var}_{\text{agg}} + \text{Var}_{\text{ind}}}. \qquad (7)$$

Here Var_{ind} shows the individual-level variance. Var_{agg} shows the adjusted variance of the city aggregate rate.[5]

2.3. Price equation in Henrich's context

"The Price equation is a simple statistical statement" on the expected change of the frequency of a gene x. In this gene, x_i denotes an altruistic allele if individual i retains $x_i = 1$, while an egoistic allele if $x_i = 0$. x_i can then express the current frequency of this gene. Let w_i be the number of offsprings of i, namely, the absolute fitness of i while $\bar{w}$ the average fitness of the population. In Henrich's contribution, "we ignore any effects arising from the transmission process (e.g. recombination, mutation, etc.)."[6] (Henrich's drafts, p. 18). It then follows the Henrich Eq. (8):

$$\bar{w}\,\Delta\bar{x} = \underbrace{\beta_{wj,xj}\,\text{Var}(x_j)}_{\text{selection between-groups}} + \underbrace{E(\beta_{wjxij}\,\text{Var}(x_{ij}))}_{\text{selection within-group}}. \qquad (8)$$

Here $\beta_{w_{ij}x_{ij}}$ means the within-group regression coefficients of w_{ij} with respect to x_{ij} in group j. This type of Price equation "tells us that the change in the frequency of allele created by natural selection acting on individuals can be partitioned into between-groups and within-group components." (Henrich's drafts, p. 18).

2.4. Difficulties of Price equation

It is easily seen that our Price equation shares the same idea as the simple model of Glaeser and Scheinkman (2001). In those kinds simply a statistical relationship is expressed to measure a social interaction by decomposing both in between-groups effects and in within-group effects. This way is quite rightful in practice in measuring contributions of

---

[4] "[I]n the equilibrium of this model two agents who are separated by K other agents will ... do exactly the same thing if there are no fixed agents between them." (Glaeser and Scheinkman, 2001, p. 95).

[5] We set

$$\text{Var}_{\text{agg}} = \text{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(A(i) - \frac{\overline{f(X)_j}}{1-\alpha}\right)\right]$$

.

[6] In the Glaeser–Scheinkman model, no sorting occurs within city. Cities may be replaced with groups in the Price equation. $A_i - \overline{A}_j$ may be x_{ij}.

interactive factors due to between-groups and within-group as shown generally in a form of:

$$\bar{w}\,\Delta\bar{x} = \underset{\text{selection between-groups}}{\text{Cov}(w_j, x_j)} + E_j(\text{Cov}(w_{ij}, x_{ij}) + \underset{\text{selection and transmission within-group}}{E(w_{ij}\Delta x_{ij}))} \tag{9}$$

But this kind of formulation does not entirely mean to bring an overarching mechanism working between between-groups and within-group. Without elucidating the overarching mechanism, the signs of between-groups/within-group effects cannot be definitely ascertained.

> [S]election on group level, as a rule, is slow and much less effective than on the level of individual (Hofbauer and Sigmund, 1989, p. 109).

The assumption of a mutant gene to promote altruistic behavior for his group may be rejected by the reason that it reduces the fitness of its carrier. In order to make group selection well workable, transmission mechanism is indispensable, as Henrich states. Cultural group selection based on such a transmission could be viable if and only if a certain macroscopic order should be implemented in the interactive field. In other words, a promising way to take a cooperative system formation into account is to introduce a macroscopic order in the concerned field in which individual agents act bilaterally with their other groups. It may be the key issue to find such a binding condition in group selection to form a macroscopic order. Putting another way, we must look for a macroscopic microeconomic linkage (i.e. an overarching mechanism to illustrate the processes of a disperse social system).

In the next section, we can only refer to the binding conditions for a macroscopic order in the theory of consumer demand, although it is not yet a binding condition on covariances of group selection for cooperation.

3. A digression: an overarching mechanism working in market demand

3.1. Slutsky equation on price changes

Some digression may be required for a while. It is very instructive to learn how a macroscopic binding condition works to organize a kind of economic law. In the theory of consumer demand, we have a well-known equation called Slutsky equation on demand variation due to the changes of price p. $f^i(p, x^i)$ denotes a demand function of individual $i \in I$ who belongs to the cloud (income class) x^i. The Slutsky equation system requires the following hypothesis in order to derive the law of demand.

Hypothesis (i). The average income effect matrix A is positive semi-definite.

Hildenbrand then proved quite ingeniously that Hypothesis (i) can be equivalent to Hypothesis (ii), provided the demand functions satisfy the budget identity (see Hildenbrand, 1994, Appendix 5).

Hypothesis (ii). [Increasing spread and expanding dispersion of household's demand.] For every sufficiently small $\Delta > 0$, the cloud $\{f^i(p, x^i + \Delta)\}_{i \in I}$ is more spread than $\{f^i(p,$

x^i)}$_{i \in I}$. In other words, the second moments of the former are greater than those of the latter.

Dispersion on the clouds can be defined to measure heterogeneity of households. To remove the so-called Giffen effects, the average income effect matrix A must be positive semi-definite. According to a new theory of Hildenbrand, this corresponds to the fact that "dispersion in each income to measure heterogeneity of households may increase, as income size increases." In short, it then holds:

$$\text{Ell}(\text{Cov } \nu(x^i + \Delta)) \supset \text{Ell}(\text{Cov } \nu(x^i) \tag{10}$$

Here Ell means ellipsoid of dispersion in the covariance matrix Cov ν. Our empirical test thus comes down to calculate the matrix of covariance around the mean on the clouds.

3.2. A macroscopic microeconomics

As Hildenbrand (1994) smartly proved, thus, it is verified that individualistic demand behavior could be well-defined to make income effects always positive, provided that macroscopic variables like variances and covariances on spending among consumer goods are pertinently bound under a certain set of hypotheses.[7] In a sense that a microeconomic behavior on demand could be solved by introducing a macroscopic condition, Hildenbrand called this approach "macroscopic microeconomic approach." (Hildenbrand, 1994, p. 74).

Hildenbrand in cooperation with Härdle (1990) working in the field of non-parametric testing, in fact, estimated the matrices of second moments as well as covariances using the available data like *UK Family Expenditure Survey* (FES), *French Enquete Budget de Famille* (*EBF*) and the surveys in other developed countries but except for Japan.[8]

Now we return our main topics. It seems us difficult to find a binding condition on group selection towards a cooperative system in the case of altruistic and egoistics alleles. First of all, we must be keen to promote some intensive efforts to achieve such empirical analysis as done in Glaeser and Scheinkman (2001).

References

Aoki, M., 1996. New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects. Cambridge University Press, Cambridge, UK.

Arthur, W.B., 1994. Increasing Returns and Path Dependence in the Economy. University of Michigan Press, Ann Arbor.

Aruka, Y., 2001. The law of consumer demand in Japan: a macroscopic microeconomic view. In: Takayasu, H. (Ed.), Empirical Science of Financial Fluctuations: The Advent of Econophysics. Springer, Tokyo, pp. 294–303.

Durlauf, S.N., 1997. Statistical mechanics approaches to socioeconomic behavior. In: Arthur, W.B., Durlauf, S.N., Lane, D.A. (Eds.), The Economy as an Evolving Complex System II. Addison-Wesley, Reading, MA, pp. 81–104.

---

[7] Subgroups may be interpreted as income classes in the society; fitness of the subgroup may be measured in terms of spending on consumer goods.

[8] According to the above method Aruka (2001) estimated the matrix of second moment, as well as the matrix of covariance in the family expenditure data in Japan of the period 1979–1998.

Durlauf, S.N., 2000. A Framework for the Study of Individual Behavior and Social Interaction. SSRI Working Paper Series. Social Systems Research Institute of the Economic Department, The University of Wisconsin, Madison.

Föllmer, H., 1974. Random economies with many interacting agents. Journal of Mathematical Economics 1, 51–62.

Friedman, D., Sighn, N., 2001. Evolution and negative reciprocity. In: Aruka, Y. (Ed.), Evolutionary Controversies in Economics: A New Transdisciplinary Approach. Springer, Tokyo, pp. 103–114.

Glaeser, E.L., Scheinkman, J.A., 2001. Measuring social interactions. In: Durlauf, S.N., Young, H.P. (Eds.), Social Dynamics. Brooking Institute, Washington, DC, pp. 83–131.

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge, UK.

Helbing, D., 1995. Quantitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes. Kluwer Academic Press, Dordrecht.

Hildenbrand, W., 1971. Random preferences and equilibrium analysis. Journal of Economic Theory 3, 414–429.

Hildenbrand, W., 1994. Market Demand. Princeton University Press, Princeton.

Hofbauer, J., Sigmund, K., 1989. The Theory of Evolution and Dynamical Systems. Cambridge University Press, Cambridge, UK.

Weidlich, W., Haag, G., 1983. Quantitative Sociology. Springer, Heidelberg.

Discussion

A whole new ball game[☆]

David Sloan Wilson[*]

Department of Biological Sciences, Binghamton University, State University of New York,
Binghamton, NY 13902-6000, USA

At a broad scale, Henrich's article is right on target, forcefully making a number of major points that should be of central interest to the readership of JEBO: (a) human social behavior needs to be understood from an evolutionary perspective; (b) the principles of kin selection and reciprocity are not up to the task; (c) all evolutionary theories of social behavior can be understood as various forms of selection within and among groups; (d) numerous mechanisms that fall under the umbrella term "culture" reduce the importance of within-group selection and increase the importance of among-group selection, even in very large groups of unrelated individuals; (e) the uniqueness of human social behavior therefore can be explained within the context of a theory that explains social behavior in all species.

Because evolutionary biologists have relied so heavily on kin selection and reciprocity in the past, and because they prematurely rejected group selection without realizing that it was embedded in the structure of their own favored ideas, the theoretical framework outlined by Henrich constitutes a whole new ball game for understanding the nature of human social behavior. Understandably, much remains to be learned about the new game and its rules. I therefore have a number of comments that should be regarded as constructive, without detracting from my overall admiration of the article and the broad vision that I share with Henrich.

Relative fitness and the importance of random variation among groups. Most selection at the individual level takes place in randomly mating populations, which means that genetic variation among individuals is random. Inbreeding and other forms of assortative mating can create nonrandom variation, but this is not regarded as essential for the process of individual-level selection to take place. It follows that random genetic variation among groups should be sufficient for group-level selection to take place, but Henrich seems to

---

imply the opposite when he states (p. 4) that "If groups are randomly re-mixed every generation, $B_{xjxi} = 0$, and altruism (even kin-based altruism) cannot evolve."

This statement is based on defining altruism as increasing the absolute fitness of others while decreasing the absolute fitness of the altruist. However, multilevel selection theory in general and the Price equation in particular are based on relative fitness comparisons.

To appreciate the difference, consider a behavior that increases the fitness of everyone in the group at no cost to the actor. The actor increases its own absolute fitness but not its relative fitness within the group. However, if groups vary in the frequency of this behavior— if only at random—then groups with a higher frequency will disproportionately contribute to the total gene pool. In the Price equation, the within-group component of gene frequency change is zero while the between-group component is positive. This example is comparable to standard individual selection, in which a beneficial gene increases the fitness of all the genes in an individual at no special cost to itself.

It is more realistic to assume that some cost in required to benefit the whole group, such as individuals donating money to public radio, which they enjoy along with those who do not donate. In this case the within-group component of the Price equation becomes negative and the behavior begins to take on the appearance of altruism. However, it is important to realize that the private cost to the altruist appears in the within-group component of the Price equation while the benefit appears in the between-group component, because it is merely the individual's share of the public good that has been provided. This means that a degree of altruism, defined in terms of relative fitness rather than absolute fitness, can still evolve with random variation among groups. In fact, random variation is sufficient to evolve this kind of altruism whenever the altruist increases its own absolute fitness, even though the cost is private and the benefit is public (see Wilson, 1998 for an interesting exception to this rule).

Behind this somewhat technical discussion is a fundamental point that needs to be appreciated at a more intuitive level: multilevel selection theory requires thinking in terms of relative rather than absolute fitness. Psychological egoists are motivated to maximize their utility *regardless of their effects on others*. Thus, psychological egoism is inherently grounded in the concept of absolute fitness. It is common to assume as an article of faith that psychological egoism is what evolves by individual selection. The Price equation shows that this assumption is mistaken. Some egoistic acts evolve by within-group selection, when they increase the fitness of the actor more than others in the same group, but other egoistic acts actually decrease relative fitness within groups and require all the ingredients of group selection to evolve (e.g. multiple groups, variation among groups). Psychological egoists might ignore their effects on others in their own minds, but as theorists and scientists we must know about effects on others to determine whether a given behavior counts as an individual-level or a group-level adaptation. Thinking in terms of absolute fitness is so strongly ingrained, and the distinction between absolute and relative fitness is sufficiently subtle, that it will probably take a long time to fully appreciate the significance of multilevel selection theory's focus on relative fitness within and among groups.

Complex interactions and the genotype–phenotype relationship. The vast majority of evolutionary models of social behavior assume a simple relationship between genes and behaviors. For example, most altruism models assume that altruistic and selfish behaviors are coded by alternative alleles at a single locus. This is obviously intended as a simplifying assumption to make the models tractable, without claiming realism. The hope is that more

complicated and realistic genotype–phenotype relationships will not alter the basic results. It turns out that this hope is unfounded, and for a very basic reason. When genes code directly for behaviors, the partitioning of phenotypic variation within and among groups becomes closely linked to the partitioning of genetic variation. For example, the only way for a group to become behaviorally uniform is for it to be genetically uniform. The importance attributed to kin selection can be traced to this assumption. However, when genes code for component traits that interact with each other and the environment to produce the phenotypic traits that influence fitness, the partitioning of phenotypic variation within and among groups can become very different than the partitioning of genetic variation. For example, a group can become phenotypically uniform even when it is genetically diverse. Of course, this is what Henrich attributes to culture but the result can be generalized to other kinds of complex interactions.

Some examples will make these ideas less abstract. In Wade's (1977) classic group se-lection experiment, he created groups of $N = 16$ flour beetles at random, allowed them to develop for 37 days, and selected on the basis of the phenotypic trait "group size". A sim-ple genotype–phenotype relationship would involve high- and low-fecundity genes, which create high- and low-fecundity individuals, which in turn create high- and low-fecundity groups in an additive fashion. If so, then high-fecundity individuals would be favored by selection within groups and random samples of $N = 32$ genes would not produce much opportunity for selection among groups. In reality, the group-level phenotypic trait "group size" was caused in part by a complex interaction between stage-specific cannibalism (larvae eat eggs) and synchrony of egg laying. Groups in which all females laid their eggs together became large because there were few eggs for the larvae to eat. Groups in which egg laying was staggered became small because the first larvae to hatch ate the eggs that were yet to hatch. These component traits may have been genetically influenced, but they interacted in a way that converted random genetic variation into highly nonrandom phenotypic variation among groups. Furthermore, the phenotypic variation was partially heritable, as revealed by the response to selection during the course of the experiment. Complex interactions altered the balance between levels of selection without anything that could be called cultural (see Sober and Wilson, 1998 for a fuller discussion of this example).

As an even more extreme example, Swenson et al. (2000a,b) inoculated laboratory mi-crocosms with literally millions of microbes from a single well-mixed source. Based on sampling error, the initial variation among the microcosms was nil, yet profound differences developed over a period of 4 days (many microbial generations). The reason is due to the well-known butterfly effect; since microbial ecosystems are complex, even the tiniest initial differences caused by sampling error provided the basis for ever larger differences as the mi-crocosms developed along separate trajectories. Furthermore, a response to ecosystem-level selection proved that some of the phenotypic variation among microcosms, initiated by sam-pling error and magnified by complex interactions, was heritable.

To summarize, some of what Henrich attributes to culture can be attributed more generally to complex genotype–phenotype relationships. This does not diminish the importance of the cultural mechanisms discussed in the target article but it does encourage us to look for additional possibilities in human and nonhuman complex systems alike.

Low-cost punishment and other details of human multilevel selection. Thinking about the rules of "the new ball game" is so recent that many of the details are understandably

speculative. I heartily concur with Henrich's critique of the "big mistake hypothesis", in which prosocial behaviors that are difficult to explain in terms of kin selection and narrow reciprocity are attributed to the misfiring of psychological mechanisms that were adaptive in ancestral environments. It is absurd to suppose that there was no need to discriminate degrees of relatedness or probability of return benefit in ancestral social environments, especially since modern hunter-gatherers and other species that live in groups have these abilities. On the other hand, I think that Henrich makes a serious error when he states (p. 14) that "... punishment is costly for the punisher (which it certainly must be)." Human social groups are remarkable for their ability to severely punish deviant members at low cost to the punishers. Simple gossip can destroy a person's reputation at no cost to the gossipers—as long as their talk is justified by the target's deviant behavior (Wilson et al., 2000). The ethnographic literature is full of examples of low-cost and even no-cost social controls, such as forcing someone into "voluntary exile" for violating a norm or placing a "blot" upon their genealogy, which prevents advantageous marriages from taking place for the deviant individual's entire lineage (see Sober and Wilson, 1998 for a fuller discussion of these and other examples). The ability of human groups to control their upstart members at low cost is termed "reverse dominance" by Boehm (1993, 1999) and should be acknowledged as one of the major factors favoring the large-scale cooperation that Henrich is trying to explain.

Other important details can be challenged while remaining within the spirit of "the new ball game". Boyd and Richerson's (1985) conformist transmission model assumes that conformity evolves as an individual-level adaptation to varying environments and that its effect on cultural group selection is initially fortuitous. It is equally plausible to assume that conformist cultural transmission evolved directly by genetic group selection (Wilson and Kniffin, 1999). The model of Henrich and Boyd (2001), in which a dash of conformity stabilizes the cost of enforcement within groups, is clever and potentially important but also potentially unnecessary, since enforcement can evolve by group selection in randomly varying groups as long as the absolute fitness threshold is passed, as discussed previously. The consequences of imitating success depend radically on the scale of comparison. Imitating success within groups causes selfishness to be emulated, while imitating success at a wider scale can cause prosocial behaviors to be emulated. To my knowledge, no model explicitly examines the evolution of alternative imitation rules that vary in their scale of comparison. Rational choice and explicit group-level decision making can have an important effect on human cultural evolution (Boehm, 1996; Wilson, 1997).

The social physiology of human groups. The physiological processes that enable organisms to survive and reproduce are often mind-boggling in their sophistication and complexity when understood in detail. Much of the fascination of biology resides in working out the mechanistic details of adaptive design. If human groups are adaptive units, they should also have a complex and sophisticated "social physiology", to use a term that is already current among social insect biologists. As Henrich notes (p. 26), social scientists were accustomed to interpreting various features of social organizations as "for the good of the group" during the first half of the 20th century, but this tradition was eclipsed by methodological individualism at about the same time that group selection was rejected in biology. In both cases, the early group-functional approaches were naïve and easy targets for criticism. However, now that group selection is on a firm theoretical foundation, especially in the case of human evolution, the tradition of functionalism in the social sciences needs to be revived, albeit in

a more sophisticated form. Richerson and Boyd (1999) have taken a step in this direction by comparing the "social physiology" of American, Russian, and German military groups during World War II. I have taken another step by analyzing religious groups from a multilevel evolutionary perspective (Wilson, 2002). It is almost obvious in retrospect that the many explicit behavioral prescriptions, beliefs about supernatural agents and their relationships with people, and social practices employed by religions around the world and throughout history are largely designed to bind individuals into functional groups, much as Durkheim (1912) broadly envisioned ninety years ago. Indeed, the very word 'religion' is derived from the Latin word 'religio', which means 'to unite or bind together'. At the same time, religions are frequently 'corrupted' from within, which can be interpreted as within-group selection gaining the upper hand. Human evolution has truly been a multilevel process in which group selection is a strong but by no means the only force. In the future, much of the fascination of the social sciences will reside in working out the mechanistic details of group-level adaptive design as a partial solution to the ever-present threat of within-group selfishness.

It should be clear from the target article and my comments that much remains to be learned about multilevel selection and human evolution. I hope the target article has the impact it deserves and will stimulate the readership of JEBO to join the new game.

References

Boehm, C., 1993. Egalitarian society and reverse dominance hierarchy. Current Anthropology 34, 227–254.
Boehm, C., 1996. Emergency decisions, cultural selection mechanics and group selection. Current Anthropology 37, 763–793.
Boehm, C., 1999. Hierarchy in the Forest. Harvard University Press, Cambridge, MA.
Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.
Durkheim, E., 1912/1995. The Elementary Forms of Religious Life (Karen E. Fields, Trans.). The Free Press, New York.
Henrich, J., Boyd, R., 2001. Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. Journal of Theoretical Biology 208, 79–89.
Richerson, P.J., Boyd, R., 1999. Complex societies: the evolutionary origins of a crude superorganism. Human Nature 10, 253–290.
Sober, E., Wilson, D.S., 1998. Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge, MA.
Swenson, W., Arendt, J., Wilson, D.S., 2000a. Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. Environmental Microbiology 2, 564–571.
Swenson, W., Wilson, D.S., Elias, R., 2000b. Artificial Ecosystem Selection. Proceedings of the National Academy of Sciences 97, 9110–9114.
Wade, M.J., 1977. An experimental study of group selection. Evolution 31, 134–153.
Wilson, D.S., 1997. Incorporating group selection into the adaptationist program: a case study involving human decision making. In: Simpson, J., Kendricks, D. (Eds.), Evolutionary Social Psychology. Erlbaum, Mahwah, NJ, pp. 345–386.
Wilson, D.S., 1998. Hunting, sharing and multilevel selection: the tolerated theft model revisited. Current Anthropology 39, 73–97.
Wilson, D.S., 2002. Darwin's Cathedral: Evolution, Religion, and the Nature of Society. University of Chicago Press, Chicago.
Wilson, D.S., Kniffin, K.M., 1999. Multilevel selection and the social transmission of behavior. Human Nature 10, 291–310.
Wilson, D.S., Wilczynski, C., Wells, A., Weiser, L., 2000. Gossip and other aspects of language as group-level adaptations. In: Heyes, C., Huber, L. (Eds.), Cognition and Evolution. MIT Press, Cambridge, MA, pp. 347–365.

Discussion

Comment on "cultural group selection, coevolutionary processes and large-scale cooperation" by Joseph Henrich

Alex Possajennikov[*]

University of Mannheim, SFB 504, Mannheim, Germany

The paper provides a very good overview of recent theoretical and empirical research in anthropology and theoretical biology focusing on explaining large-scale cooperation in human societies. The author surveys an impressive body of results, criticizes the purely genetic approach to the evolution of cooperation, and promotes a gene-culture coevolutionary approach instead. Since most of the results and theories surveyed are published in anthropology and biology journals, and are probably not well known to economists, it is of great interest to read such a survey. I congratulate the author on having done a good job in providing the survey, and with explaining the gene-culture coevolution of social learning and cooperation.

The following general picture emerges from the paper. Pure genetic evolution cannot explain the degree of cooperation in humans and the difference in the domains of cooperation; humans (almost) uniquely among animals possess the (genetic) ability for social learning; social learning may stabilize any behavior via, e.g. conformist transmission; a more cooperative behavior in a group makes the group more viable; cooperation spreads through cultural group selection; in a cooperative environment genetic evolution also favors cooperation. This general picture looks quite clear and intuitive. However, some statements and steps need more analysis.

Altruism: Before analyzing the steps of the above reasoning, I would like to make a comment that is not directly related with the paper but more with the topic of this special issue, on the use of the notion "altruism". Though an action that reduces own fitness and increases the fitness of others may look altruistic, the reasons for taking the action may be the desire to conform, or the fear of punishment, and thus be purely selfish. Punishment, in turn, may have the same reasons, or even be out of spite. At least in the references cited in

---

[*] Present address: School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Tel.: +44-115-951-5461; fax: +44-115-951-4159.
E-mail address: a1ex.possajennikov@nottingham.ac.uk (A. Possajennikov).

the paper, I could not find any that makes a clear case that punishment is altruistic. Altruism, as the innate inclination to take a socially beneficial action at an individual cost, appears only at the last stage of the process described in the previous paragraph. A large part of the paper concerns the evolution of large-scale cooperation; (genetic) evolution of altruism follows it and is in fact not much elaborated in the paper. For example, in the models of reciprocity and punishment, playing a cooperative action is not altruism, but rather selfish behavior, hoping for future benefits or avoiding costs.

Large-scale cooperation and genetic evolution: As I am not an expert in anthropology and biology, I take as given the assertion of the author that humans cooperate on a much larger scale, and in a larger variety of domains, than animals do. The demonstration that purely genetic models of the evolution of cooperation may produce some cooperation but not on a large scale also looks convincing; however, see the end of the next paragraph for possible problems with it.

Social learning: I also take as given the assertion that humans are (almost) unique in having the (genetic) ability for social learning. That social learning is useful in a wide range of varying environments for one-person decision problems and for coordination problems is convincing. That social learning is rare among animals because of a cost of maintaining such an ability (Boyd and Richerson, 1996) is a possible explanation that is not elaborated in sufficient detail in the paper, which focuses instead on the advantages of social learning. Less convincing is the assumption that it is difficult to distinguish cooperative dilemmas from other problems. Though it may be a more plausible kind of "big mistake" hypothesis than the inability to distinguish small scale repeated interactions from large-scale one-shot ones, it puts the evolution of cooperation via social learning on a different ground than the genetic evolution of cooperation. In genetic models of the evolution of cooperation the situation was unambiguously identified as a cooperative dilemma. If a genotype that leads to cooperative behavior in social dilemmas leads somehow to a more beneficial behavior in coordination problems, and the situations are non-distinguishable, the cooperative genotype can be favored by genetic evolution. Perhaps cultural evolution explains cooperation better but then cultural and genetic evolution should be compared on a level field of the uncertain world of complex non-distinguishable situations that could be social dilemmas, coordination problems, or perhaps something else.

Conformist transmission and cooperation: That once the ability for social learning, and with it such mechanisms as conformist transmission, evolved, it can stabilize any behavior, looks intuitively plausible. The devil is, as usual, in the details. For example, the results in the example of the stability of cooperation and punishment hinge on the assumption that conformist transmission has the same strength for any punishment stage. As defections from the common pattern of cooperation and punishment are rare, the payoff advantage of non-punishing on the ith stage is small, allowing conformist transmission to overpower it. However, if conformist transmission is based on imitation, or observational learning, there are very few opportunities to see the punishment actually carried out if cooperation is common, so conformist transmission could also be weaker for each stage of punishment. Though the result of the stability of cooperation and punishment might still hold, possibly a larger degree of conformist transmission is needed to support it.

Cultural group selection: That more cooperation in a group makes the group more viable is a plausible assumption. More problematic, in my view, is cultural group selection. Richerson

and Boyd (1998), Boyd and Richerson (2002a,b) develop models where whole groups become extinct, and new groups are formed by fission from existing groups. In their models a group is the unit of selection. On the other hand, the Price Eq. (8), even generalized for groups, has an individual as the unit of selection. Though helpful, once groups are there, the Price equation does not provide us with information about how groups are formed. A formal model of group formation from individual actions is missing, though the discussion of different processes of cultural group selection in the paper is illuminating in this respect.

Focusing on inter-group variation, or taking a group as the unit of selection, is justified if social learning within groups is faster than cultural group selection. This, in turn, may require fairly stable groups, which brings us back to the problem of cooperation in one-shot anonymous interactions that genetic evolution could not explain. For cultural transmission, be it via conformist transmission or punishment, it is also necessary to observe the behavior of others in a similar situation or to have the possibility of punishment, which means that the situation is repeated, or the identities of individuals do not change too rapidly from one interaction to the next. People may behave cooperatively in artificial one-shot anonymous social dilemmas in experiments because they socially learned to behave like this in complex, possibly repeated, possibly requiring coordination, real life environments.

Genetic evolution of cooperation: The final step, that once cooperative behavior is common, genes that favor altruism may spread, is problematic, and is not elaborated in the paper. If altruism is understood as preference for taking a socially beneficial action, and genes for social learning are there, once altruists are common, the state is stable. However, altruists are behaviorally indistinguishable from egoists who socially learn to play the action because of fear of punishment or normative conformity. The population of altruists is only neutrally stable. If the environment is complex and changing, rigid genetic altruism may never be as good as the ability to learn. Each new situation is difficult to evaluate, and the ability to learn what is good to do may be better than the approximate (because of the complexity of the situation) ability to play the socially beneficial action.

The paper by Henrich endeavors to explain large-scale cooperation and the difference in domains of cooperation among humans. It makes a convincing case in favor of gene-culture coevolution of social learning and cooperation though some steps of the reasoning require closer attention; in particular, cultural group selection lacks a formal model. However, the evolution of altruism, as innate preference for socially beneficial action, is not well explained. What is well explained is the possible evolution of the ability to cooperate in social dilemmas. This ability is acquired by a process that may incorporate genetic evolution of reciprocity and punishment, as well as social learning via conformist or prestige-biased transmission, or again punishment. Social learning may people actually like cooperation, but not necessarily, especially where punishment is involved. The extent to which altruism (preference for socially beneficial action), as opposed to the ability to cooperate, actually evolves, is still an open question.

References

Boyd, R., Richerson, P., 1996. Why culture is common, but cultural evolution is rare. Proceedings of the British Academy 88, 77–93.

Boyd, R., Richerson, P., 2002a. Solving the Puzzle of Human Cooperation. In: Levinson, S. (Ed.), Evolution and Culture. MIT Press, Cambridge MA, in press.

Boyd, R., Richerson, P., 2002b. Group beneficial norm can spread rapidly in a structured population. Journal of Theoretical Biology 215, 287–296.

Richerson, P., Boyd, R., 1998. The Evolution of Ultrasociality. In: Eibl-Eibesfeldt, I., Salter, F.K. (Eds.), Indoctrinability, Ideology and Warfare. Berghahn Books, New York, pp. 71–96.

Discussion

Cooperation and wealth[☆]

Oded Stark[*]

University of Bonn, Bonn, Germany
University of Vienna, Vienna, Austria
ESCE Economic and Social Research Center, Cologne, Germany and Eisenstadt, Austria

Abstract

We calculate the equilibrium fraction of cooperators in a population in which payoffs accrue from playing a single-shot prisoner's dilemma game. Individuals who are hardwired as cooperators or defectors are randomly matched into pairs, and cooperators are able to perfectly find out the type of a partner to a game by incurring a recognition cost. We show that the equilibrium fraction of cooperators relates negatively to the population's level of wealth.
© 2003 Published by Elsevier B.V.

JEL classification: A13; C70

Keywords: Equilibrium fraction of cooperators; Population's level of wealth; Single-shot prisoner's dilemma game

1. Introduction

An example illustrates that the level of wealth of a population and the equilibrium fraction of cooperators in a population are inversely related. It has been argued that the fraction of cooperators in a large society can be expected to be smaller than the fraction of cooperators in a small society (Binmore, 1998; Cook and Hardin, 2001). To the extent that a large society (say a city) is wealthier than a small society (say a town), the size effect may conceal a wealth effect.

2. The game and the payoffs

Consider the following two-player, two-strategy game in which a player who cooperates gets a payoff of *R* if his opponent cooperates, and *S* if the opponent defects. A player who defects gets *T* if his opponent cooperates, and *P* if the opponent defects. The game is a prisoner's dilemma game: $T > R > P > S$. Hence defection is the dominant strategy for each player.

Let there be a large population of players consisting of individuals who are hardwired to be cooperators and individuals who are hardwired to be defectors. Individuals are randomly matched into pairs. An individual does not know the type of the individual with whom he is matched, but he can obtain such information at a cost, $0 < K < \bar{K}$, where $\bar{K}$ will be defined below. The type-recognition test is perfect. Thus, if an individual chooses to incur the cost and administer the test, the individual finds out whether he is matched with a cooperator or with a defector. The individual can then decide to play or not to play. If the individual decides not to play, he randomly picks another individual from the population and administers the type-recognition test in the new match. If individuals agree to play, they play their hardwired strategies, receive their respective payoffs, and leave the partner-seeking population to be replaced by new individuals. In equilibrium (to be characterized below) the flow of individuals of each type who enter the population exactly replaces the flow of individuals of each type who exit the population.

3. The types and their expected payoffs

Following Stark (1999, chapter 5), we study a population that consists of three types: defectors who play without incurring a recognition cost, cooperators who play after incurring the recognition cost, and cooperators who play without incurring the recognition cost. While there can be an equilibrium with all three types present and an equilibrium with defectors only, (i) there cannot be an equilibrium without defectors; and (ii) there cannot be an equilibrium with only defectors and non-testing cooperators. The rationale for (i) is that there cannot be an equilibrium with only non-testing cooperators because defectors will do better than cooperators; there cannot be an equilibrium with only testing cooperators because non-testing cooperators will do better; and there cannot be an equilibrium with only both types of cooperators because the non-testing cooperators will do better than the testing cooperators. The rationale for (ii) is that there cannot be an equilibrium with only defectors and non-testing cooperators because defectors will do better than the non-testing cooperators.

Let the steady-state fractions of testing cooperators, non-testing cooperators, and defectors be π_t, π_{nt}, and π_d, respectively, $\pi_t + \pi_{nt} + \pi_d = 1$. Given the manner in which a testing cooperator acts and plays, his expected payoff is

$$V_t = R - \frac{K}{1 - \pi_d}. \tag{1}$$

The proof is as follows: the expected net payoff from administering the cost *K* (exactly once) and encountering a cooperator in the first match is $R(1 - \pi_d) - K(1 - \pi_d)$; from failing

to encounter a cooperator in the first match but encountering one in the second match is $R\pi_d(1-\pi_d) - 2K\pi_d(1-\pi_d)$; from failing to encounter a cooperator in the first two matches but succeeding in encountering one in the third match is $R\pi_d^2(1-\pi_d) - 3K\pi_d^2(1-\pi_d)$; and so on. Thus,

$$
\begin{aligned}
V_t &= R(1-\pi_d) - K(1-\pi_d) + R\pi_d(1-\pi_d) - 2K\pi_d(1-\pi_d) \\
&\quad + R\pi_d^2(1-\pi_d) - 3K\pi_d^2(1-\pi_d) + \cdots \\
&= \frac{R(1-\pi_d)}{1-\pi_d} - K(1-\pi_d)(1+2\pi_d+3\pi_d^2+\cdots) \\
&= R - K(1-\pi_d)[(1+\pi_d+\pi_d^2+\cdots) + (\pi_d+\pi_d^2+\cdots) + (\pi_d^2+\cdots) + \cdots] \\
&= R - K(1-\pi_d)\left(\frac{1}{1-\pi_d} + \frac{\pi_d}{1-\pi_d} + \frac{\pi_d^2}{1-\pi_d} + \cdots\right) \\
&= R - K(1-\pi_d)\frac{1/(1-\pi_d)}{1-\pi_d} = R - \frac{K}{1-\pi_d}. \qquad \square
\end{aligned}
$$

The expected payoff of a non-testing cooperator who plays the game with whomever he is paired with in the first match is

$$
V_{nt} = (1-\pi_d)R + \pi_d S. \tag{2}
$$

Since a defector always plays, that is, he plays when matched either with a non-testing cooperator or with a defector, his expected payoff is

$$
V_d = \frac{1-\pi_t-\pi_d}{1-\pi_t}T + \frac{\pi_d}{1-\pi_t}P = T - \frac{\pi_d}{1-\pi_t}(T-P). \tag{3}
$$

4. Equilibrium with defectors and testing cooperators but without non-testing cooperators

From the discussion in the preceding section it follows that an equilibrium with defectors and testing cooperators but without non-testing cooperators is feasible. If there are no non-testing cooperators, $\pi_t + \pi_d = 1$; the expected payoff of testing cooperators is $V_t = R - (K/\pi_t)$; and the expected payoff of defectors (who can play only with defectors) is $V_d = P$. In equilibrium, testing cooperators receive the same expected payoff as defectors. Thus,

$$
R - \frac{K}{\pi_t} = P
$$

or

$$
\pi_t = \frac{K}{R-P}, \tag{4}
$$

assuming that $K < R - P \equiv \bar{K}$.

To help unravel the nature of the equilibrium, consider alternative values of K. Suppose that K were equal to $R - P$. π_t would then be equal to one. But having a population with only

testing cooperators cannot be an equilibrium because in that case the non-testing cooperators will do better. Thus, we have a contradiction. Suppose that $K \to 0$. It follows that $\pi_t \to 0$. Yet suppose the opposite, that is, that $\pi_t \to 1$. If such were the case, the population would consist of only testing cooperators which, from (i) in Section 3, cannot hold. As K assumes values that increasingly move it away from being close to $R - P$ toward close to zero, the associated values of π_t must become *smaller*. To see the reason for this result, suppose that an equilibrium holds at $R - (K_0/\pi_{t_0}) = P$ and consider the opposite, that is, as K declines from K_0 to K_1, π_t *increases* from π_{t_0} to π_{t_1}. But then $(K_1/\pi_{t_1}) < (K_0/\pi_{t_0})$, rendering it impossible to restore equilibrium at $R - (K_0/\pi_{t_0}) = P$. As long as R and P are given, observing the equilibrium requires that π_t and K move in tandem.

To complete the characterization of the equilibrium we note that in order for there to be no non-testing cooperators in the population, it has to be the case that if a non-testing cooperator were to enter the population, he will receive a lower payoff than that received by the testing cooperators and the defectors, that is, $\pi_t R + (1 - \pi_t) S < P$. Substituting $\pi_t = K/(R - P)$ and rearranging terms we get

$$K < \frac{(P - S)(R - P)}{R - S} = \frac{P - S}{R - S} \bar{K} < \bar{K}$$

since $(P - S) < (R - S)$. Hence, exclusion of non-testing cooperators requires that

$$K < \frac{P - S}{R - S} \bar{K} \equiv \bar{\bar{K}}.$$

5. The relationship between the equilibrium fraction of cooperators in a population and a population's level of wealth

Suppose we compare two populations that are equal in all respects except that one population is uniformly wealthier than the other population. By "uniformly" we mean that there are no distributional differences in the payoffs to strategies; the only difference between the two populations is that in one population the payoffs are uniformly higher than in the other population, say by a factor of $\mu > 1$. Holding K constant, $\pi_t^w = K/\mu(R - P)$ of the wealthier population is smaller than $\pi_t = K/(R - P)$ of the less wealthy population: the equilibrium fraction of cooperators in a wealthy population is smaller than the equilibrium fraction of cooperators in a (uniformly) less wealthy population.[1]

To appreciate the nature of this outcome consider the case of $\pi_t = K/\mu(R - P)$ where $\mu \to \infty$. It follows that $\pi_t \to 0$. The implication of a rising μ is that the absolute difference

---

[1] To rule out the possibility that, in spite of the payoffs to every cooperator and to every defector being higher in the wealthier population, the payoff per capita (and, since population size is held constant, total wealth) will be lower in the wealthier population, the sufficient condition that $\mu > \underline{\mu} \equiv \pi_t/\pi_t^w$ can be added. This condition arises from the requirement that the per capita payoff in the wealthier population will be higher than the per capita payoff in the less wealthy population:

$$\pi_t^w \left(\mu R - \frac{K}{\pi_t^w} \right) + (1 - \pi_t^w) \mu P > \pi_t \left(R - \frac{K}{\pi_t} \right) + (1 - \pi_t) P.$$

between the payoffs R and P becomes increasingly larger. With K held constant, if π_t were, alternatively, to *rise*, the expected payoff of testing cooperators will increasingly distance itself from the expected payoff of defectors (who, it will be recalled, play only with defectors) and equilibrium will not be restored.

Two comments regarding recognition costs are in order. First, for the equilibrium to hold, K can take a wider range of values than before since the constraint pertaining to K, which is now $K < \mu \bar{\bar{K}}$, is less stringent. Second, the inverse relationship between the equilibrium fraction of cooperators and the level of wealth holds even when K increases with wealth, provided that the increase is less than μ. An increase in wealth is due to and entails a first order increase in the payoffs from trade and exchange but, at most, a second order increase in the cost of conducting trade. Indeed, in a population whose level of wealth is higher, the recognition cost could be lower (for example, a computerized credit inquiry could replace a lengthy interview). If $K = K(\mu)$ and $K'(\mu) < 0$, then $\pi_t^w = K(\mu)/\mu(R - P)$ and

$$\frac{\partial \pi_t^w}{\partial \mu} = -\frac{K(\mu)}{\mu^2(R - P)} + \frac{K'(\mu)}{\mu(R - P)} < -\frac{K(\mu)}{\mu^2(R - P)};$$

the adverse effect of a higher level of wealth on the equilibrium fraction of cooperators is stronger.

6. Robustness of the cooperation–wealth relationship when the testing cooperators are somewhat adventurous

Suppose that a testing cooperator acts in the following manner: with probability $0 < q \leq 1$ he administers the type-recognition test. With probability $1 - q$ he does not resort to the test and plays with whomever he happens to be paired with. (We know that q cannot be equal to zero because then we will have only defectors and non-testing cooperators which, from (ii) in Section 3, cannot be the case in equilibrium.) We seek to find out whether the result of Section 5 holds in this setting too.

The expected payoff of an adventurous testing cooperator is

$$V_t^a = \frac{(1 - \pi_d)R + (1 - q)\pi_d S - qK}{1 - q\pi_d}. \tag{5}$$

The proof is as follows: when testing occurs with probability q, a match will confer a payoff either when the test is applied (at a cost K) and the partner in the match is found to be a cooperator, a case in which the play yields $q[-K + (1 - \pi_d)R]$, or when the test is not applied, a case in which the payoff received is $(1 - q)[(1 - \pi_d)R + \pi_d S]$. In the event that the test is applied and the partner to the match is found not to be a cooperator, which occurs with probability $q\pi_d$, no payoff is received and the entire procedure is repeated thereby yielding V_t^a. Thus,

$$\begin{aligned} V_t^a &= q[-K + (1 - \pi_d)R] + (1 - q)[(1 - \pi_d)R + \pi_d S] + q\pi_d V_t^a \\ &= \frac{(1 - \pi_d)R + (1 - q)\pi_d S - qK}{1 - q\pi_d}. \end{aligned} \qquad \square$$

Since the combined population share of testing cooperators who happen not to administer the test and of defectors is $1 - q\pi_t$, the expected payoff of a defector is

$$V_d = \frac{1 - q\pi_t - \pi_d}{1 - q\pi_t} T + \frac{\pi_d}{1 - q\pi_t} P$$

or

$$V_d = T - \frac{\pi_d}{1 - q\pi_t}(T - P). \tag{6}$$

In equilibrium, adventurous cooperators receive the same expected payoff as defectors. Thus, from (5) and (6),

$$\frac{(1 - \pi_d)R + (1 - q)\pi_d S - qK}{1 - q\pi_d} = T - \frac{\pi_d}{1 - q\pi_t}(T - P).$$

Of course, $q\pi_t + (1 - q)\pi_t = \pi_t$ and hence $\pi_t + \pi_d = 1$. We therefore have that

$$\frac{\pi_t R + (1 - q)(1 - \pi_t)S - qK}{1 - q(1 - \pi_t)} = T - \frac{1 - \pi_t}{1 - q\pi_t}(T - P). \tag{7}$$

Evaluating this last equality at $q = 1$ yields

$$\frac{\pi_t R - K}{\pi_t} = T - (T - P)$$

or

$$\pi_t = \frac{K}{R - P}.$$

By continuity this last equality holds for values of q in (7) that are in the small neighborhood of 1. Hence, the cooperation–wealth relationship alluded to in Section 5 holds also when testing cooperators apply the test with a probability that is less than, but close to, one.

7. Conclusion

We calculate the equilibrium fraction of cooperators in a population in which payoffs are received upon playing a two-person single-shot prisoner's dilemma game; individuals who are hardwired as cooperators or as defectors are paired randomly; cooperators check, at a cost, the type of individual with whom they are paired prior to executing a game and play only with cooperators; and defectors play with whomever they happen to be paired. Measuring the wealth of a population by the level of the payoffs in the prisoner's dilemma game, we show that the wealthier the population, the lower the equilibrium fraction of cooperators.

Acknowledgements

References

Binmore, K., 1998. Game Theory and the Social Contract. II. Just Playing. MIT Press, Cambridge, MA.

Cook, K.S., Hardin, R., 2001. Norms of cooperativeness and networks of trust. In: Hechter, M., Opp, K.D. (Eds.), Social Norms. Russell Sage Foundation, New York, pp. 327–347.

Stark, O., 1999. Altruism and Beyond: An Economic Analysis of Transfers and Exchanges Within Families and Groups. Cambridge University Press, Cambridge.

Discussion

Evolutionary parallelism versus co-evolution: a comment on Joseph Henrich

Werner Güth [a,*], Hartmut Kliemt [b]

[a] *Max-Plank-Institute for Research into Economic Systems, Strategic Interaction Unit,*
von Wirtschafts systemen, Wernor Gth Kahlaische Strasse 10, D 07745 Jena, Germany
[b] *Philiosophy Department GMU, Duisburg, Germany*

Accounts of the astonishing extent to which human behavior tends to be unselfish even in anonymous large-scale interaction (see Kropotkin, 1919) have always provoked explanations in terms of group interest, group selection and group functions. But since the late sixties the use of such terms has become almost a Lakatosian "do not" (see Lakatos, 1978) among evolutionary theorists. Only recently there have been several attempts to rejuvenate concepts of group selection in evolutionary biology and beyond (for a discussion that captured broader attention (Sober and Wilson, 1998)). That these attempts now have made their way into economics (see for an overview Bergstrom, 2002) and that they are related to findings from the experimental economics and other empirical literature is a promising development.

Joseph Henrich's paper is a case in point. He is clearly right to insist that the idea of group selection has often been dismissed too lightly on a priori rather than empirical grounds in the past. At the same time Henrich's criticisms of many approaches to explaining the extent of "pro-social behavior" in human societies are well put. One of us toyed some 20 years ago with the "big mistake hypothesis" (see Kliemt, 1985). The severe difficulties of that hypothesis are clearly laid out by Henrich (in the section on kin based selection). Likewise he evenhandedly evaluates the mixed evidence on the "green beard" signaling mechanism. We have used a related idea of "signaling and detecting commitments" when analyzing trust in large-scale anonymous interactions (see Güth and Kliemt, 2000; Güth, Kliemt and Peleg, 2000). We still believe that Henrich's criticism does not render our ideas useless. But it is clearly very useful to put economic views in a broader perspective here. In any event, Henrich's cautionary remarks are well put: In evolutionary settings all depends on the complexities of the situations in which evolution takes place. Bearing this in mind let

---

* Corresponding author.: Tel.: +49-3641-686-620; fax: +49-3641-686-623.
E-mail address: gueth@mpiew-jena.mpg.de (W. Güth).

us probe somewhat further into Henrich's account of the interaction between "genes, mind and culture" (Lumsden and Wilson, 1981).

1. Group selection, assortative matching and mediating structures

Henrich points out that in Eq. (4) of the paper the crucial term is $\beta_{x_j x_i}$ "which captures the degree to which 'being an altruist' predicts 'being in an altruistic group'." But literally speaking this is not true. What the term captures is the likelihood of another bearer of the same allele being around in any interaction. What Henrich describes as the possibility that "(for whatever reason) altruists can preferentially group with other altruists" could as well be described as their faculty to "interact" preferentially with other altruists. Using the description that individuals are "grouped" is not a neutral expression for preferential interaction since it transports the notion of a group as a distinct entity.

Nobody can reasonably deny that the details of the frequency of "matching and mating" as well as other details of interaction influence evolutionary success. If there were no factors—including "green beard" and other signaling and detection mechanisms—enhancing the likelihood of interaction between kin as opposed to non-kin, even kin selection could not work. Therefore, we need to study whatever mechanisms can bring about differentials in the likelihood of interaction. But to use the term "group selection" so broadly as to capture all evolutionary effects on the composition of a gene pool brought about by differentials in the likelihood that an individual bearer A of a genetic endowment × will interact with another bearer B of the same genetic endowment × seems far-fetched. At least we are not necessarily talking about "groups" in the common sense of "sets of permanently interacting individuals". All the more or less moral connotations of altruism that are normally wheeled in by relying on group selection should be kept out here.

We presume that Henrich would insist that his own, like Sober's and Wilson's well known approach, draws attention to mediating structures to an extent that nevertheless justifies the use of the term "group selection". But it should not be forgotten that in his approach it is relative shares in the gene pool that count as the single currency of evolutionary success. Henrich's broad use of the term "group selection" certainly does not imply that the old idea of "serving the interest" of one's kind per se plays a role. It is not, as older theorists seemed to think, the general interest in the sense of the interest of all members of some biological natural kind that is furthered. In short, Henrich's point that mediating structures are important is well taken. We fully agree with him but also insist that one should be careful not to allude to the stronger classical sense of the old group selection idea.

2. Genes, mind and culture

Economists have often tried to model competitive processes as a form of social selection completely analogous to biological selection (see early on Alchian, 1950). In a broader perspective, Hayek and his followers tried to look at society as a whole from an evolutionary point of view (see Vanberg, 1986). Such analogies may be useful in themselves, but they do not integrate biological and cultural evolution in an effort to understand the interaction

between the evolutionary processes. We think that Henrich is entirely right in trying to go beyond analogy. He is correctly aiming at an integrated approach, but we do not think that his paper fully succeeds in going beyond parallelism of genetic and cultural evolution.

His extended discussion of the Price equation makes this obvious. After presenting it along with some of its modifications in the first part of the paper, the equation is reformulated for the process of cultural evolution. This is pure parallelism. How cultural and genetic evolution interact is not systematically taken into account. The central argument about second-order free riding suffers from the same deficiency. Henrich argues "that the nature of our cultural transmission capacities, and of human psychology more-generally, creates stable behavioral equilibria, consisting of cooperation and punishment, that are not available to genetic evolutionary processes in acultural species" (beginning of the section on "cultural group selection, cultural transmission and large-scale cooperation"). The description of the causal processes that in Henrich's view explain why the second-order free riding problem can be solved is not without plausibility. Human individuals may not be able to discriminate between social dilemma situations and other situations in their imitation behavior. Indeed, it may well be that there is some adaptive advantage to imitating the prevalent forms of behavior and role models. We do not deny this. Neither do we deny that the process may be mistake driven (see on the role of trembles or mistakes in a biological evolutionary context (Selten, 1983, 1988)). We think that Henrich presents a plausible potential explanation that may lead to a plausible solution of the problem of second-order free riding in punishment. But we do not see how this solution of the free rider problem connects with group selection and the like as biological phenomena. Nor do we think that the connection with the empirical and in particular experimental literature in economics and psychology is made in a sufficiently systematic way to go beyond evolutionary story telling (a business we also like to engage in).

For instance, Henrich alludes to studies of the biological advantages of certain forms of imitation behavior. But this insight is not systematically linked with his argument about cultural evolution in a model of the co-evolutionary process. His argument that punishment can be sustained due to certain limitations of the human cognitive apparatus seems independent of the biological argument of the first part. Proximate causes like the structure of the cognitive apparatus, however it came about, form the basis of Henrich's explanations (and explanatory ideas like Coleman's "free riders and zealots" (see Coleman, 1988) might be factored in here as well). Relying on categories like "conformist and prestige-biased transmission" the paper contributes considerably to our understanding of why cultural variety between internally homogeneous groups can be sustained. But the link between ultimate causes in biological evolution and cultural aspects remains to a large extent in the dark and the relationship to group selection rather obscure.

To combine "conformist and prestige-biased transmission" with punishment mechanisms seems quite convincing. This may put a premium on a genetic proclivity to show group serving behavior regardless of the group in which it takes place and regardless of the specific form it adopts. But this mechanism operates without relying on any differential advantage in propagating genetic material by membership in alternative groups. Selecting the cooperators by culturally evolved punishment mechanisms is not—at least not primarily—a matter of group selection but rather an individual selection process brought about by a mediating structure of norm enforcement. Moreover, the related argument that conformity with group

norms may be self-enforcing in the presence of multiple equilibria and equilibrium selection problems has been made frequently (see Binmore, 1998; Bergstrom, 2002). What would be interesting is a specific model describing how the levels of genetic and cultural evolution interact in bringing about the effects of which Henrich quite plausibly speaks in his story.

3. Concluding remark

The traditional question of how "nurture" and "nature" hang together cannot be ignored if we intend to understand to what extent human nature is adapted to living in a "great society" and to what extent this form of life is "artificially" created by human institutions. Which forms of social life are viable over the long haul depends on whether interaction at large is parasitic on social capital created in small groups and whether or not any genetic prerequisites of large-scale human cooperation are selected for or weeded out by life in the modern world. In particular whether there is a natural biological basis for altruism and non-opportunistic behavior in large-scale interaction even among strangers is a most interesting issue.

How these issues are related to the quarrel between adherents of group selection and their adversaries is not clear from the paper. Whether or not "groups" form another ("emergent") stratum of reality or a mediating structure governed by its own evolutionary dynamics is important in itself but an answer to the issue is not needed for the argument of the second part of the paper. The "missing link" of a specific evolutionary model of the feedback between cultural evolution, cognitive abilities, psychological constraints and the like must still be provided. Our own favorite approach in this regard would be an indirect evolutionary one (see Güth and Kliemt, 1998; and originally Güth and Yaari, 1992). But we hasten to admit that we are quarrelling with the same almost unsurmountable difficulties of coping with the details of an overly rich social reality that Henrich tries to deal with in his inspiring discussion.

References

Alchian, A.A., 1950. Uncertainty, evolution, and economic theory. Journal of Political Economy 58, 211 ff.

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

Binmore, K., 1998. Game Theory and Social Contract. Just Playing, vol. 2. MIT Press, Cambridge, London.

Coleman, J.S., 1988. Free riders and zealots: the role of social networks. Sociological Theory 6, 52–57.

Güth, W., Kleimt, H., Peleg, B., 2000. Co-evolution of preferences and information in simple games of trust. German Economic Review 1, 83–110.

Güth, W., Kliemt, H., 1998. Towards a fully indirect evolutionary approach. Rationality and Society 10, 377–399.

Güth, W., Kliemt, H., 2000. Evolutionarily stable co-operative commitments. Theory and Decision 49, 197–221.

Güth, W., Yaari, M., 1992. An evolutionary approach to explaining reciprocal behavior in a simple strategic game. In: Witt, U. (Ed.), Explaining Process and Change—Approaches to Evolutionary Economics. The University of Michigan Press, Ann Arbor, 23 ff pp.

Kliemt, H., 1985. Moralische Institutionen. Karl Alber, Freiburg und München.

Kropotkin, P., 1919. Mutual Aid: A Factor of Cooperation. A.A. Knopf, New York.

Lakatos, I., 1978. The Methodology of Scientific Research Programmes. Cambridge University Press, Cambridge.

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

Selten, R., 1983. Evolutionary stability in extensive two-person games. Mathematical Social Sciences 5, 269–363.

Selten, R., 1988. Evolutionary stability in extensive two-person games—correction and further development. Mathematical Social Science 16, 223–266.

Sober, E., Wilson, D.S., 1998. Unto Others. The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge.

Vanberg, V., 1986. Spontaneous Market Order and Social Rules. A Critical Examination of F.A.V. Hayek's Theory of Cultural Evolution. Economics and Philosophy, vol. 2. 75 ff pp.

Discussion

Altruism may be sexy
Comment on cultural group selection, coevolutionary processes and large-scale cooperation

Kenichi Aoki*

Department of Biological Sciences, The University of Tokyo, Hongo 7-3-1, Bunkyoku, Tokyo 113-0033 Japan

The target article by Joseph Henrich provides much food for thought. I have enjoyed reading it and recent papers on experimental economics (Henrich et al., 2001; Sigmund et al., 2002) that describe the observations that Henrich is trying to explain. My comment comprises some prefatory remarks followed by two substantive ones.

We suggested that indiscriminate altruism could have evolved in hominids by group selection if the cost–benefit ratio was less than 1/6. The value of 1/6 was our estimate of the average relatedness in multi-male hominid groups (Aoki, 1982; Crow and Aoki, 1982, 1984; Aoki and Nozawa, 1984). Our approach was entirely genetic and ignored the cultural component to human altruism. Undoubtedly as Henrich argues here, "culturally-transmitted ideas . . . are important for understanding human cooperation." However, the first "empirical reason" given in support of this claim, namely "the domains of cooperative behavior in humans vary. . . . In some societies, people may cooperate in fishing and house-building. . . . In neighboring groups . . . people may cooperate in warfare," does not rule out a genetic propensity to be altruistic—a direct bias, perhaps. For example, language acquisition is made possible by a genetically determined language function (shared by all healthy individuals), but there are a tremendous variety of spoken and signed languages in the world (Kyle and Woll, 1985; Crystal, 1987; Ruhlen, 1994).

Nevertheless, there is great merit to the model of cultural group selection acting on the large between-group differences generated by conformist transmission (Boyd and Richerson, 1982, 1985). When individuals copy the majority behavior, cultural dynamics

---

* Tel.: +81-3-5841-4485; fax: +81-3-3818-7547.
E-mail address: kenaoki@biol.s.u-tokyo.ac.jp (K. Aoki).

within groups may yield two stable equilibria. Specifically, there may be an equilibrium at which altruists predominate in addition to one at which egoists are in the majority. Because of natural selection against altruists, the egoist equilibrium will be more easily reached. Moreover, when the tendency to conform is weak, the altruist equilibrium vanishes unless there are spiteful individuals who are willing to punish egoists, altruists who do not punish egoists, etc. at a cost to themselves. However, if groups at the altruist equilibrium do better in between-group competition, altruism may enjoy a net advantage. Lumsden and Wilson (1981) summarize the evidence for conformist transmission in humans. Henrich and Boyd (1998) have conducted a thorough theoretical analysis of the conditions under which conformist transmission might evolve.

The point I want to make first is that substantial between-group differences can also be generated by runaway sexual selection, whereby a male trait and a female preference for that trait become genetically correlated and coevolve (Fisher, 1958). Even a male trait that lowers viability—altruism, in particular—can evolve, provided the female preference is strong enough. Differentiation of groups occurs when different combinations of traits and their associated preferences coevolve as a result of subtle differences in the initial conditions. Sexual selection also applies to female (and sexually monomorphic) traits. However, in birds and mammals sexual selection usually acts more forcefully on a trait expressed in males, because a male that is chosen by many females can sire offspring by each of those females, whereas the fertility of a female is limited more by physiological constraints than by the number of males she attracts. Traits that first evolve in males can later be expressed in females because the two sexes share most of their genes.

Much theoretical (and empirical) work on sexual selection has followed on the seminal papers by Lande (1981) and Kirkpatrick (1982). Here, I adapt Kirkpatrick's two-locus haploid model to the altruism problem by setting the viability of egoistic males to $1 + bt_2$ and the viability of altruistic males to $1 - c + bt_2$, where t_2 is the frequency of altruistic males. Clearly, c and b are the cost and benefit of altruism. There are also two kinds of females, one that mates at random, and the other that shows a sexual preference for altruistic males. The latter are a (>1) times more likely to mate with an altruistic male than an egoistic male in a two-way choice. The frequency of such choosy females is p_2.

Fig. 1 is a schematic that illustrates the curve of neutral equilibria and the directions of change in the (t_2, p_2)-variable plane. Note that t_2 increases in the region above the curve and
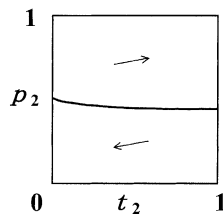

Fig. 1. The curve of neutral equilibria is a hyperbola: $p_2 = (c/b)\{1 + [(1-c+b)/(1-c+bt_2) - a/(1+bt_2)]/(a-1)\}$. When (as in the present case) this curve transects the (t_2, p_2)-variable plane, the frequency of altruistic males, t_2, increases in the region above and decreases in the region below the curve. There is a correlated change in the frequency of choosy females, p_2, due to the positive linkage disequilibrium between the two loci.

decreases in the region below, as indicated by the arrows. Hence, the evolutionary outcome depends on the initial value of p_2. The figure was drawn for the parameter values $c = 1/3$, $b = 1/2$, and $a = 2$. The cost and benefit are such that their ratio exceeds the upper limit at which altruism can evolve by kin selection in a dioecious (separate sexes) random mating group, namely $c/b = (2/3) > (1/2)$ (Hamilton, 1964). There is no empirical basis for the strength of preference assumed here. Although Jensen-Campbell et al. (1995) show that women find altruistic men sexually more attractive than egoistic men, it is not clear how the attraction measures used by these authors relate to the actual occurrence of mate choice. To conclude my account of sexual selection, genetic dynamics within groups may yield altruist as well as egoist equilibria, providing the between-group variation on which group selection can act. (Sexual selection can also be modeled with cultural transmission of trait and preference (see Richerson and Boyd, 1989; Nakajima and Aoki, 2002; Ihara et al., 2003).)

My second point concerns the process of cultural group selection involved in the spread of early farming in Europe. Henrich suggests that this occurred by "demographic swamping," but I believe that "prestige-biased group selection" is closer to the truth. To see why let us briefly review the arguments and the evidence. Ammerman and Cavalli-Sforza (1971, 1973) invoke Fisher's wave-of-advance model (Fisher, 1937) to argue that early farming spread through Europe by the range expansion of intrusive Mid-Eastern farmers. As supporting evidence they note that the rate of advance is constant, as predicted by Fisher's model. These authors are also of the view that the gene frequency clines observed in Europe were formed during the spread of early farming—the clines are parallel to the direction of spread (Ammerman and Cavalli-Sforza, 1984). Unfortunately, the two proposals are contradictory, since Fisher's model implies replacement of the indigenous populations of Europe by the people of Mid-Eastern origin.

We have extended Fisher's model by incorporating the possibility that indigenous hunter-gatherers may be converted to farming—prestige-biased transmission (Aoki et al., 1996). Our model predicts a wave front of converted farmers advancing at a constant speed, in the wake of which the converted farmers will gradually mix with the Mid-Eastern farmers.

References

Ammerman, A.J., Cavalli-Sforza, L.L., 1971. Measuring the rate of spread of early farming in Europe. Man 6, 674–688.

Ammerman, A.J., Cavalli-Sforza, L.L., 1973. A population model for the diffusion of early farming in Europe. In: Renfrew, C. (Ed.), The Explanation of Culture Change. Duckworth, London, pp. 343–357.

Ammerman, A.J., Cavalli-Sforza, L.L., 1984. The Neolithic Transition and the Genetics of Populations in Europe. Princeton University Press, Princeton, NJ.

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

Aoki, K., Nozawa, K., 1984. Average coefficient of relationship within troops of the Japanese monkey and other primate species with reference to the possibility of group selection. Primates 25, 171–184.

Aoki, K., Shida, M., Shigesada, N., 1996. Travelling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. Theoretical Population Biology 50, 1–17.

Boyd, R., Richerson, P.J., 1982. Cultural transmission and the evolution of cooperative behavior. Human Ecology 10, 325–351.

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

Crow, J.F., Aoki, K., 1982. Group selection for a polygenic behavioral trait: a differential proliferation model. Proceedings of the National Academy of Sciences of the United States of America 79, 2628–2631.

Crow, J.F., Aoki, K., 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proceedings of the National Academy of Sciences of the United States of America 81, 6073–6077.

Crystal, D., 1987. The Cambridge Encyclopedia of Language. Cambridge University Press, Cambridge.

Fisher, R.A., 1937. The wave of advance of advantageous genes. Annals of Eugenics 7, 355–369.

Fisher, R.A., 1958. The Genetical Theory of Natural Selection. Dover, New York.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. Journal of Theoretical Biology 7, 1–16.

Henrich, J., Boyd, R., 1998. The evolution of conformist transmission and the emergence of between-group differences. Evolution and Human Behavior 19, 215–242.

Henrich, J., Boyd, R., Bowles, S., Camberer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of Homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review 91, 73–78.

Ihara, Y., Aoki, K., Feldman, M.W., 2003. Runaway sexual selection with paternal transmission of the male trait and gene-culture determination of the female preference. Theoretical Population Biology 63, 53–62.

Jensen-Campbell, L.A., Graziano, W.G., West, S.G., 1995. Dominance, prosocial orientation, and female preferences: do nice guys really finish last. Journal of Personality and Social Psychology 68, 427–449.

Kirkpatrick, M., 1982. Sexual selection and the evolution of female choice. Evolution 36, 1–12.

Kyle, J.G., Woll, B., 1985. Sign Language. Cambridge University Press, Cambridge.

Lande, R., 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the United States of America 78, 3721–3725.

Lumsden, C.J., Wilson, E.O., 1981. Genes, Mind, and Culture. Harvard University Press, Cambridge, MA.

Nakajima, A., Aoki, K., 2002. On Richerson and Boyd's model of cultural evolution by sexual selection. Theoretical Population Biology 61, 73–81.

Richerson, P.J., Boyd, R., 1989. The role of evolved predispositions in cultural evolution. Ethology and Sociobiology 10, 195–219.

Ruhlen, M., 1994. The Origin of Language. Wiley, New York.

Sigmund, K., Fehr, E., Nowak, M.A., 2002. The economics of fair play. Scientific American January, 80–85.

Discipline note

The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions

Herbert Gintis*

15 Forbes Avenue, Northampton, MA 01060, USA

Received 17 July 2002; accepted 12 September 2002

Joseph Henrich's "Cultural Group Selection, Coevolutionary Processes and Large-scale Cooperation" is an insightful exposition of a very important research agenda in the theory of human cooperation, to which he has made important contributions. In these remarks, I will attempt to add to Henrich's argument for conformist cultural transmission, to weaken the group selection requirements for gene-cultural coevolution, and to strengthen the argument for the genetic side of the evolutionary dynamic. My analysis has three stages. First, as Boyd and Richerson (1990) stressed in a related context, the evolutionary stability of altruism does not necessarily involve a within-group fitness deficit for the altruistic trait. Cultural group selection can take the unproblematic form of groups with higher expected payoffs for all agents displacing groups with lower expected payoffs. Price's equation and the thorny issues surrounding classical group selection theory can thereby be avoided.

Second, the analysis of conformist cultural transmission can be strengthened by recognizing the uniquely human capacity to *internalize norms*, which then become *goals* (arguments in the agent's objective function) rather than *techniques—means* of attaining other goals—or *beliefs*—factual statements concerning states of affairs and causal relations.

Third, human beings have *prosocial emotions*, including shame, guilt, and empathy, that equip the individual with rewards for altruistic behavior and penalties for self-regarding behavior. Prosocial emotions are predicated upon highly developed, physiologically-based, genetically-grounded, arousal systems. Prosocial emotions heighten the subjective utility associated with internal norm compliance. To see this, note that when a self-regarding goal is not achieved we feel 'disappointed,' 'frustrated,' or 'angry,'—emotions we share with many vertebrates—but when an other-regarding (altruistic) goal is not realized, we feel 'guilty,' 'ashamed,', or 'remorseful.' Prosocial emotions also predispose us to internalize

---

* Tel.: +1-413-545-6363; fax: +1-413-586-6014.
E-mail address: hgintis@attbi.com (H. Gintis).

some norms rather than others. For instance, as Smith (2000[1759]) long ago noted, we are predisposed to experience empathy for other members of our social group, unless they have specifically acted to harm us. Which individuals and types of individuals count as members of "our social group" is highly culturally specific, but the expression of empathy for members of one's own group is probably a human universal.

Gene-culture coevolution, then, includes the following mechanism. Cultural complexity and the rapidity of cultural change (Boyd and Richerson, in press) render the internalization of norms fitness-enhancing, so the genetic predisposition to internalize norms is an evolutionary adaptation. Moreover, agents who internalize norms tend to punish norm violators. Thus, conformity to social norms becomes fitness-enhancing, which renders the genes for prosocial emotions fitness-enhancing adaptations. The internalization of norms and the prosocial emotions permit large-scale cooperation among non-kin, setting the stage for the technological and cultural evolution characteristic of modern civilization.

1. A model of cultural equilibrium with altruism

Consider a group in which members can adopt either altruistic cultural norm A, or fail to adopt, in which case they are self-interested.[1] Self-interested types (we shall call them type B) have baseline fitness $g(\alpha)$, where α is the fraction of altruists in the group, while altruistic types (those who have trait A) have fitness $(1-s)g(\alpha)$, where $0 < s < 1$. We assume in each period (a generation), members of the group pair off randomly and have offspring in proportion to their fitness. Families pass on their cultural norms to their offspring, so offspring of AA parents are altruists, offspring of BB parents are self-interested types, and half of the offspring of AB-families are altruists. We also assume that the self-interested offspring of AB- and BB-families are susceptible to influence by salient altruists in the community and community institutions promoting altruistic values, a fraction $\alpha\gamma$ of such offspring becoming altruists. We call this *oblique transmission* (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985).

First, agents pair off randomly, form families, mate, and have children. Suppose there are n males and n females at the beginning of the period. If the fraction of altruists is α, there will be $n\alpha^2$ AA-families, who will produce $n\alpha^2(1-s)^2\beta$ offspring, all of whom are altruists, where we choose β so that population grows at rate $g(\alpha)$. There will also be $2n\alpha(1-\alpha)$ AB-families, who will have $2n\alpha(1-\alpha)(1-s)\beta$ offspring, half of whom are altruists. Finally, there will be $n(1-\alpha)^2$ BB-families who will have $2n(1-\alpha)^2\beta$ offspring. Adding up the number of offspring, we see that we must have $\beta = g(\alpha)/(1-s\alpha)^2$. The frequencies of AA, AB, and BB offspring are thus given by

$$f_{AA} = \frac{\alpha^2(1-s)^2}{(1-\alpha s)^2}, \qquad f_{AB} = \frac{2\alpha(1-\alpha)(1-s)}{(1-\alpha s)^2}, \qquad f_{BB} = \frac{(1-\alpha)^2}{(1-\alpha s)^2}. \qquad (1)$$

Second, a fraction $\alpha\gamma$ of offspring of AB- and BB-families who are self-interested types switch to being altruists under the influence of the oblique transmission of the altruistic

---

[1] For a deeper analysis of this and related models, see Gintis (2003a,b).

norm A. It is easy to check that the change in the fraction of altruists in the next generation is given by

$$\dot{\alpha} = f(\alpha) = \frac{a(1-\alpha)(\gamma - s)}{1 - s\alpha}. \tag{2}$$

Third, suppose in every time period, each group member i learns the fitness and the type of a randomly chosen other member j, and changes to j's type if j's fitness is perceived as higher. However, information concerning the difference in fitness of the two strategies is imperfect, so the larger the difference in the payoffs, the more likely the agent is to perceive it and change. Specifically, we assume the probability p that an altruistic agent shift to being self-interested is proportional to the fitness difference of the two types, so $p = \sigma s$ for some proportionality constant $\sigma > 0$. Then it is easy to show (Gintis, 2000b, Chapter 9) that the fraction α of altruists follows the *replicator dynamic*

$$\dot{\alpha} = -\sigma\alpha(1-\alpha)s. \tag{3}$$

We now combine the cultural transmission and replicator sources of change in the fraction of altruists, giving $\dot{\alpha} = f(\alpha) - \sigma\alpha(1-\alpha)s$, which reduces to

$$\dot{\alpha} = \frac{\alpha(1-\alpha)}{1 - s\alpha}(\gamma - s - s\sigma(1 - s\alpha)), \tag{4}$$

where σ now represents the relative speed of the socialization and biologically adaptive processes.

We call the situation $\dot{\alpha} = 0$, $\alpha \in [0, 1]$ a *cultural equilibrium* of the dynamical system. We then have the following theorem.

Theorem 1. (*a*) *If*

$$0 < s < s_{\min} = \frac{\gamma}{1 + \sigma}, \tag{5}$$

$\alpha = 1$ *is a globally stable cultural equilibrium*: *self-interested behavior is driven to extinction.*

(*b*) *If*

$$s_{\min} < s < s_{\max} = \frac{1}{2\sigma}\{1 + \sigma - \sqrt{(1+\sigma)^2 - 4\gamma\sigma}\}, \tag{6}$$

both $\alpha = 0$ *and* $\alpha = 1$ *are stable equilibria of the system and there is third unstable equilibrium* $\alpha^* \in (0, 1)$ *separating the basins of attraction of the two stable equilibria*: *both self-interested and altruistic equilibria are stable.*

(*c*) *If* $s > s_{\max}$, *then* $\alpha = 0$ *is the only stable equilibrium of the system*: *the altruistic norm is driven to extinction.*

This theorem shows that very costly altruistic norms will not be adopted, moderately costly altruistic norms form a stable equilibrium, but this equilibrium cannot be reached starting from low frequency in the population. Finally, low-cost altruistic norms can be supported in a cultural equilibrium, provided there is a positive rate of oblique transmission of this trait.

We also have

Corollary 1.1. *If the strength of the replicator dynamic σ satisfies*

$$\sigma < \frac{\gamma - s}{s},$$

the altruistic cultural equilibrium is globally stable. If

$$\frac{\gamma - s}{s} < \sigma < \frac{\gamma - s}{s(1 - s)},$$

both the self-interested and the altruistic cultural equilibria are stable, and the basin of attraction of the altruistic equilibrium shrinks as a increases. Finally, if

$$\sigma > \frac{\gamma - s}{s(1 - s)},$$

the self-interested cultural equilibrium is globally stable.

This corollary shows that if cultural transmission is sufficiently strong by comparison with fitness-based updating (the replicator dynamic), either the self-interested type will be driven to extinction within the group, or there will be multiple equilibria. *In equilibrium, altruists never suffer a fitness deficit within the group*. Thus, Price's equation and the whole apparatus of within-group versus between group dynamics can be dispensed with.

If there is little oblique cultural transmission (small γ), however, the last part of the corollary applies, so an altruistic trait will not spread if it is associated with a positive fitness cost s. This is why altruism towards non-kin is restricted to *Homo sapiens*, which is the only species with a high level of oblique cultural transmission.

We have said nothing about the shape of the fitness function $g(\alpha)$. If 'altruism' has its usual connotation, $g(\alpha)$ will be increasing in α. Moreover, it is easy to see that the above analysis applies just as well to a set of altruistic cultural traits $A_1, \ldots, A_m$, each associated with a fitness deficit $s_1, \ldots, s_m$. An obvious argument shows that if there are several such groups competing for a limited set of resources, the group with the highest altruism benefit, which we can write $g(\alpha_1, \ldots, \alpha_m)$, where α_i is the frequency of altruistic trait i in the population, will displace the other groups.

2. The psychology of norm internalization

As Henrich suggests, conformist transmission occurs in part because it is often reasonable to follow a majority rather than undergoing the experimentation costs involved in discovering the most cost-effective alternative. However, when it comes to cultural forms representing basic values and non-experiential beliefs, this argument may well not apply, since experimentation may be infeasible, prohibitively costly, or impossible in principle.

An important prosocial psychological mechanism prevalent in humans is to promote some norms from *means* to *goals*. This process is called the *internalization of norms*.

An *internal norm* is a pattern of behavior that is intrinsically desired as a personal goal, rather than simply being used to achieve some other goal. Agents conform to an internal norm because so doing is an end to itself, and not because of the external social sanctions, such as material rewards that follow from norm compliance or punishments that follow from norm violation. For instance, an individual who has internalized the value of "speaking truthfully" will do so even in cases where the net payoff to speaking truthfully would otherwise be negative.[2] It follows that where people internalize a norm, the frequency of its occurrence in the population will be higher than if people follow the norm only instrumentally (i.e. when they perceive it to be in their interest to do so).

Society's values are transmitted through the internalization of norms (Parsons, 1967; Grusec and Kuczynski, 1997). Successful societies also promote altruistic norms that subordinate the individual to group welfare, fostering such behaviors as bravery, honesty, fairness, willingness to cooperate, and sympathy for the plight of others (Brown, 1991).[3] Included among the norms that are commonly internalized are standards of personal hygiene, appropriateness of dress and demeanor, ritual religious observances, as well as such altruistic traits as caring about the well-being of others, voluntarily cooperating in collective tasks, punishing those who violate norms, and sacrificing on behalf of one's ethnic, racial, or geographically defined "insider group."

Humans internalize norms through *socialization* by parents (*vertical transmission*), by extraparental conspecifics who control educational and religious practices (*oblique transmission*), and informal organizations of friends and neighbors (*horizontal transmission*). There is no documented capacity to internalize norms in non-human species, and its universal presence in the most varied human societies would indicate a strong genetic predisposition. Indeed, the capacity to internalize norms is diminished or lacking in some so-called 'sociopaths,' who otherwise appear to suffer no mental deficit. There is a genetic predisposition to sociopathy, and sociopaths comprise a large fraction of repeatedly incarcerated individuals in the United States (Mealey, 1995).

If the norms promoted by internalization are prosocial, then individuals will behave prosocially. Similarly, if the norms promoted by internalization are altruistic, then individuals will behave altruistically. Since prosocial and altruistic behavior enlarges the range of fitness-enhancing cooperative activities possible in human groups, a straightforward cultural evolution model (any one of those suggested by Henrich) would predict that prosocial and altruistic norms will proliferate in human societies.

If this argument is correct, then the key questions are (a) what accounts for the human predisposition to internalize norms? (b) what are the limits of internalization, or more precisely what cultural and genetic forces account for the mix of self-interested and altruistic behaviors observed in human societies? I address these questions in Gintis (2003a,b). In closing here, I will address the relationship between prosocial emotions and the internalization of norms.

---

[2] From Theorem 1 above, if the fitness cost *s* of speaking truthfully in some situation is too high, the norm of speaking truthfully in that situation will not be expressed in a cultural equilibrium.

[3] There are, of course, many anti-social norms that persist in equilibrium as well in particular societies (Edgerton, 1992).

3. A bioeconomic explanation of norm internalization

Why do we have the generalized capacity to internalize norms? The capacity to internalize is certainly curious, something akin to the capacity of a digital computer to be programmed, albeit only within certain strict limits. From a biological standpoint, internalization may be an elaboration of imprinting and imitation mechanisms found in several species of birds and mammals, but its highly developed form in humans indicates it probably had great adaptive value during our evolutionary emergence as a species. Moreover, from an economic standpoint, the everyday observation that people who exhibit a strongly internalized moral codes lead happier and more fulfilled lives than those who subject all actions to a narrow calculation of personal costs and benefits of norm compliance suggests it might not be 'rational' to be self-interested.

Gintis (2003a) shows that *if* internalization of *some* norms is personally fitness-enhancing (e.g. preparing for the future, having good personal hygiene, positive work habits, and/or control of emotions), *then* genes promoting the capacity to internalize can evolve. Given this genetic capacity, we have seen above, altruistic norms will be internalized as well, provided their fitness costs are not excessive. In effect, altruism 'hitchhikes' on the personal fitness-enhancing capacity of norm internalization. Altruistic behavior, then, is an *exaption*, in the sense of Gould and Vrba (1981).

But, why should the internalization of *any* norms be individually fitness-enhancing? The following is a possible explanation, based on the observation that internalization alters the agents' *goals*, whereas instrumental and conventional cultural forms merely aid the individual in attaining *pre-given* goals. In humans, as much as in other species, these goals are related to, but not reducible to, biological fitness.

Biological fitness is a theoretical abstraction unknown to virtually every real-life organism. Organisms therefore do not, in any circumstance, literally maximize fitness. Rather, organisms have a relatively simple state-dependent objective function that is itself subject to selection according to its ability to promote individual fitness (Alcock, 1993). In a slowly-changing environment, this objective function will track fitness closely. In a rapidly-changing environment, however, natural selection will be too slow, and the objective function will not track fitness well.

The development of cultural transmission, in the form of instrumental techniques and conventions, and the ensuing increase in social complexity of hominid society, doubtless produced such a rapidly-changing environment, thus conferring high fitness value on the development of a *non-genetic mechanism for altering the agent's objective function*. Internalization is adaptive because it allows the human objective function to shift in directions conducive to higher personal fitness. The internalization of norms is thus adaptive because it facilitates the transformation of drives, needs, desires, and pleasures (arguments in the human objective function) into forms that are more closely aligned with fitness maximization. Internalization is limited to our species, moreover, because no other species places such great emphasis on cultural transmission.

We humans thus have a 'primordial' objective function that does not well serve our fitness interests, and which is more or less successfully 'overridden' by our internalized norms. This primordial objective function knows nothing of 'thinking ahead,' but rather satisfies immediate desires. Lying, cheating, killing, stealing, and satisfying short-term

bodily needs (wrath, lust, greed, gluttony, sloth) are all actions that produce immediate pleasure and drive-reduction, at the expense of our overall well-being in the long run. This fact explains the congenital weakness of human nature in its tendency to succumb to the unruly temptations of the flesh.

This evolutionary argument is meant to apply to the long period in the Pleistocene during which the human character was formed. Social change since the agricultural revolution some 10,000 years ago has been far too swift to permit even the internalization of norms to produce a close fit between utility and fitness. Indeed, with the advent of modern societies, the internalization of norms has been systematically diverted from *fitness* (expected number of offspring) to *welfare* (net degree of contentment) maximization. This, of course, is precisely what we would expect when humans obtain control over the content of ethical norms. Indeed, this *misfit* between welfare and fitness is doubtless a necessary precondition for civilization and a high level of *per capita* income. This is true because were we fitness maximizers, every technical advance would have been accompanied by an equivalent increase in the rate of population growth, thus nullifying its contribution to human welfare, as predicted long ago by Thomas Malthus. The demographic transition which has led to dramatically reduced human birth rates throughout most of the world is a testimonial to the gap between welfare and fitness. Perhaps the most important form of prosocial cultural transmission in the world today is the norm of having few, but high quality, offspring.

4. Shame, guilt, empathy, and other prosocial emotions

"How selfish soever man may be supposed," notes Adam Smith in *The Theory of Moral Sentiments* (2000 [1759]), "there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it" Smith here alludes to the role of *sympathy* in promoting agreeable relations among men. Later, he analyzes the contribution of shame to the enforcement of covenants: "Our imagination therefore attaches the idea of shame to all violations of faith."

The fact that adherence to social norms is underwritten by emotions has too often been ignored in the years since Smith wrote his great work. The experience of shame, guilt, and other visceral reactions plays a central role in sustaining cooperative relations. Prosocial emotions function like the basic emotion, "pain," in providing guides for action that bypass the explicit cognitive optimizing process that lies at the core of the standard behavioral model in economics. Damasio (1994, p. 173) calls these "somatic markers." A somatic marker is a bodily response that "forces attention on the negative outcome to which a given action may lead and functions as an automated alarm signal which says: Beware of danger ahead if you choose the option that leads to this outcome . . . the automated signal protects you against future losses." Emotions thus contribute to the decision-making process, not necessarily by clouding reason, but in beneficial ways as well. Damasio continues: "suffering puts us on notice . . . it increases the probability that individuals will heed pain signals and act to avert their source or correct their consequences" (p. 264). while some psychologists (e.g. Zajonc, 1980) have stressed emotions as important influences in behavior, and a few economists

have taken up these ideas (Loewenstein, 1996; Laibson, 1996), the role of emotions in behavior has been little studied by economists.

The specific social circumstances that evoke emotions are culturally specific, but the capacity to express and experience emotions are human universals (Ekman and Davidson, 1994; Ekman, 1999), involving complex chains of hormonal, visceral, and non-cognitive neural responses that are rooted in our genetic constitution as a species. While many human emotional responses have counterparts in vertebrates (Darwin, 1998), and especially our closest relatives, the chimpanzees (de Waal, 1998), the prosocial emotions in non-humans are extremely rudimentary by comparison. From this, and their intimate association with the cultural forms that evoke them, we can conclude that human prosocial emotions coevolved with human culture.

Dramatic confirmation of the importance of emotions is the well-known *ultimatum game*, in which, under conditions of anonymity, two player are shown a sum of money, say US$ 10. One of the players, called the "proposer," is instructed to offer any number of dollars, from US$ 1 to 10, to the second player, who is called the "responder." The responder, again under conditions of anonymity, can either accept the offer or reject it. If the responder accepts the offer, the money is shared accordingly. If the responder rejects the offer, both players receive nothing. Since the game is played only once and the players do not know each other's identity, a self-interested responder will accept *any positive amount of money*. Knowing this, a self-interested proposer will offer the minimum possible amount, US$ 1, and this will be accepted. However, when actually played, *the self-interested outcome is almost never attained or even approximated*. In fact, as many replications of this experiment have documented, under varying conditions and with varying amounts of money, proposers routinely offer respondents very substantial amounts (50% of the total generally being the modal offer), and respondents frequently reject offers below 30% (see Gintis, 2000b, Chapter 11 for citations and analysis).

Economists were at first astonished at this behavior. Why would people reject a positive amount of money? They suggested that perhaps players did not understand the game. But by changing the rules a little bit, experimenters quickly showed that this was not the case. For instance, if the responder is told that the offer is generated by a computer, the rejection rate become very low, however small a share is offered to the responder. Or, in another variation, the players were told that if the responder rejected the offer, he would get nothing, but the proposer would still get the share he proposed for himself. In this case, responders never rejected offers.

In fact, responders reject positive offers because they are *angry* at being treated unfairly, and they are willing to pay (in the form of getting nothing when they could have gotten something) to hurt the person who perpetrated this unfairness. Rejecting unfair offers is part of what we call *strong reciprocity* (Gintis, 2000a; Bowles and Gintis, 2003), which is a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid. Strong reciprocity is thus a combination of *altruistic cooperation* and *altruistic punishment*.

How generalizable are these studies? Most such studies, after all, use college students as subjects, and though the results are similar around the world, this could be because college students are pretty much the same around the world. Are the patterns described above

evidence of universal patterns of behavior, or do the subjects' economic and social environments shape behavior? If the latter, which economic and social conditions are involved? Is strong reciprocity better explained statistically by individuals' attributes such as their sex, age, or relative wealth, or by the attributes of the group to which the individuals belong? Are there cultures that approximate the self-interested actor of standard economics and biology?

For answers, I and several collaborators (including Joe Henrich, who undertook the first such field experiment, and has led the work on analyzing the results of our larger study) undertook a large cross-cultural study of behavior in societies very different from ours. Twelve experienced field researchers, working in 12 countries on four continents, recruited subjects from 15 small-scale societies exhibiting a wide variety of economic and cultural conditions. Our sample consists of three for-aging societies, six who practice slash-and-burn horticultural, four nomadic herding groups and three sedentary, small-scale agriculturists. Our results are described in Henrich et al. (2001).

We found, first, that the self-interested actor model is not supported in any society studied. Second, there is considerably more behavioral variability across groups than had been found in previous cross-cultural research and the self-interested actor model fails in a wider variety of ways than in previous experiments. Third, group-level differences in the degree of cooperation in production and the extent of contact with market economies explain a substantial portion of the behavioral variation across societies: the higher the extent of contact with market economies and the higher the level of cooperation in production, the larger was the offer in the ultimatum game. Finally, behavior in the experiments is generally consistent with economic patterns of everyday life in these societies.

These and related studies show that the emotions involved in social interaction—the empathy we feel towards those who have been kind to us and the hostility towards those who have not—often counteract the purely material incentives involved, producing outcomes that are more cooperative than would have been expected from a dispassionate tallying up the material costs and benefits of various actions.[4]

If we have such highly developed prosocial emotions, it is very likely that they served to enhance our individual and group fitness in the social conditions under which our species evolved biologically. The most obvious purpose served by such emotions as pride, self-esteem, empathy, guilt, and shame, is that they induce individuals to obey social norms in situation where the temptation of immediate gratification might induce individuals to break social rules, thereby incurring the wrath of the group to the extent of reducing one's long-run fitness. Thus, to the extent that humans evolved cultural forms that depend on individual compliance with norms, and developed the genetic predisposition to (altruistically) punish norm violators, the fitness-enhancing value of prosocial emotions were enhanced. This, of course, is a perfect case of the coevolution of genes and culture.

The economist might reasonably respond that a self-interested "rational actor" would have no need of such emotions, but would rather directly tally up the costs and benefits of a particular action, thus choosing a level of "anti-social" behavior that maximized his personal payoff. Prosocial emotions, which are in effect a *subsidy* on prosocial behavior and

---

[4] These same emotional capacities and predispositions, of course, are equally key to understanding conflict and war in human societies, but space limitations prevent me from exploring this "anti-social" side of prosocial emotions.

a *tax* on anti-social behavior, should only serve to reduce the payoff to norm-compliance. To see that such reasoning is suspect' note that it applies equally to *pain*, which no one doubts is an evolved mechanism for recognizing and avoid bodily damage: a rational actor should *register* bodily damage, and choose actions that optimally trade off such damage against other gains; but pain itself is an additional *tax* on bodily harm, and should be suboptimal.

What, then, is wrong with the argument that prosocial emotions are inefficient? First, there is a simple argument from bounded rationality: the process of optimizing one's prosocial behavior may simply be too informationally costly and error-prone. Second, subjective pleasure and comfort are important elements in the objective function that humans maximize, and there is no reason to believe that these correlate well with reproductive fitness, since we may not have been subject to evolutionary selection for sufficiently long to ensure a close fit between the two. The prosocial emotions, being directly pleasurable, might have evolved to correct an imbalance due to our imperfect decision-making criteria. Finally, there is much evidence that human beings are excessively present-oriented and use hyperbolic discounting, which lowers the present value of future as compared with current rewards (Akerlof, 1991; Ainslie and Haslam, 1992; Loewenstein and Prelec, 1992; Laibson, 1997; Kirby and Marakovic, 1997; O'Donoghue and Rabin, 1999, 2000, 2001). There is at present no plausible model explaining the adaptive value of hyperbolic discounting, but it is known to be present in non-human animals, so presumably there is some evolutionary explanation of the phenomenon. Be that as it may, if humans are excessively present-oriented, then such prosocial emotions as shame and empathy will contribute to fitness, because they either increase the present cost of harming others (as in the case of empathy) or they increase the future expected cost of norm violation (as in the case of shame).

References

Ainslie, G., Haslam, N., 1992. Hyperbolic discounting. In: Loewenstein, G., Elster, J. (Eds.), Choice Over Time. Russell Sage, New York, pp. 57–92.

Akerlof, G.A., 1991. Procrastination and obedience. American Economic Review 81, 1–19.

Alcock, J., 1993. Animal Behavior: An Evolutionary Approach. Sinauer, Sunderland, MA.

Bowles, S., Gintis, H., 2003. The Evolution of Strong Reciprocity, Cooperation in Heterogeneous Populations, Theoretical Population Biology (in press).

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

Boyd, R., Richerson, P.J., 1990. Group selection among alternative evolutionarily stable strategies. Journal of Theoretical Biology 145, 331–342.

Boyd, R., Richerson, P.J. The Nature of Cultures. Department of Anthropology, UCLA, in press.

Brown, D.E., 1991. Human Universals. McGraw-Hill, New York.

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

Damasio, A.R., 1994. Descartes' Error: Emotion, Reason, and the Human Brain. Avon Books, New York.

Darwin, C., 1998. In: Ekman, P. (Ed.), The Expression of Emotions in Man and Animals. Oxford University Press, Oxford.

de Waal, F., 1998. Chimpanzee Politics: Sex and Power Among the Apes. Johns Hopkins University Press, Baltimore, MD.

Edgerton, R.B., 1992. Sick Societies: Challenging the Myth of Primitive Harmony. Free Press, New York.

Ekman, P., 1999. Basic emotions. In: Dalgleish, T., Power, T. (Eds.), The Handbook of Cognition and Emotion. Wiley, New York, pp. 45–60.

Ekman, P., Davidson, R., 1994. The Nature of Emotion: Fundamental Questions. Oxford University Press, Oxford.

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

Gintis, H., 2000b. Game Theory Evolving. Princeton University Press, Princeton, NJ.

Gintis, H., 2003a. The hitchhiker's guide to altruism: genes, culture, and the internalization of norms. Journal of Theoretical Biology 220, 407–418.

Gintis, H., 2003b. Solving the puzzle of human prosociality. Rationality and Society, 15.

Gould, S.J., Vrba, E., 1981. Exaption: a missing term in the science of form. Paleobiology 8, 4–15.

Grusec, J.E., Kuczynski, L., 1997. Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory. Wiley, New York.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fenr, E., Gintis, H., McElreath, R., 2001. Cooperation, reciprocity and punishment in fifteen small-scale societies. American Economic Review 91, 73–78.

Kirby, K.N., Marakovic, N.N., 1997. Modeling myopic decisions: evidence from hyperbolic delay-discounting with subjects and amounts. Organizational Behavior and Human Decision Processes 64, 22–30.

Laibson, D., 1996. A Cue-Theory of Consumption. Harvard University Press, Cambridge, MA.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quarterly Journal of Economics 112, 443–477.

Loewenstein, G., Prelec, D., 1992. Anomalies in intertemporal choice: evidence and an interpretation. Quarterly Journal of Economics 57, 573–598.

Loewenstein, G.F., 1996. Out of control: visceral influences on behavior. Organizational Behavior and Human Decision Processes 65, 272–292.

Mealey, L., 1995. The sociobiology of sociopathy. Behavioral and Brain Sciences 18, 523–541.

O'Donoghue, T., Rabin, M., 1999. Doing it now or later. American Economic Review 89, 103–124.

O'Donoghue, T., Rabin, M., 2000. The economics of immediate gratification. Journal of Behavioral Decision-Making 13, 233–250.

O'Donoghue, T., Rabin, M., 2001. Choice and procrastination. Quarterly Journal of Economics 116, 121–160.

Parsons, T., 1967. Sociological Theory and Modern Society. Free Press, New York.

Smith, A., 2000[1759]. The Theory of Moral Sentiments. Prometheus, New York.

Zajonc, R.B., 1980. Feeling and thinking: preferences need no inferences. American Psychologist 35, 151–175.

JOURNAL OF
Economic Behavior
& Organization

Discussion

Culture, genes and cooperation: comment on Henrich

Louis Putterman[*]

Department of Economics, Brown University, Box B, Providence, RI 02912, USA

Henrich's paper lays out the issues very nicely. Humans are capable of greater degrees of cooperation with those with whom they are not in repeated contact than are members of other species. But whereas reciprocal cooperation with specific others and kin altruism serve the inclusive fitness of the individual's genes, offering cooperation or altruism less restrictively must by definition reduce individual fitness. A partial way out exists if noncooperation is regularly punished, but this just shifts the problem to another level: in the presence of punishers, fitness-seeking genes have incentives to dictate cooperation on the first-order task, but shirking on costly punishing.

Henrich and others propose to solve this problem by according a role to group selection. Within groups, individuals who free-ride on the services of other punishers will have higher payoffs; but on average, members of groups with more punishers will have higher payoffs. Can group advantage outweigh individual advantage in the course of evolutionary selection? The answer is yes, if differences between groups are sufficiently sustained while differences within groups are kept sufficiently small, so that the advantages of groups with more cooperators translate into overall populations with more cooperators. The problem with positing group selection in a model of human genetic evolution is that in known hunter-gatherer societies, migration and intermarriage are constantly diluting genetic differences among groups. Instead of genes, Henrich proposes culture as the mechanism reducing intra-group differences and maintaining inter-group differences, thanks to in-built psychological biases including the bias to conform with group norms. If the evolution of strong human sociality is attributable to humans' cultural capacities, this also helps to account for the greater pro-sociality of human than of other species (apart from the eusocial insects).

There is much that is appealing in this way of laying out the problem. Recent experiments nicely illustrate Henrich's point that "if . . . altruists can preferentially group with

[*] Tel.: +1-401-863-3837; fax: +1-401-863-1970.

E-mail address: louis_putterman@brown.edu (L. Putterman).

other altruists . . . altruism at least has a chance. If egoists can preferentially enter groups with altruists" this is not the case ($\beta_{x_j x_i}$ will be negative). Ehrhart and Keser (1999) and Page et al. (2002) conducted public goods experiments in which subjects were able to exercise some control over which other subjects they were grouped with. Ehrhart and Keser let subjects unilaterally choose what extant group to enter or whether to form a new group, so free-riders invaded groups of cooperators, the latter had to flee to avoid exploitation, and overall cooperation declined with repetition. Page et al., by contrast, grouped subjects together according to the mutual preferences of all subjects involved,[1] so that more cooperative subjects could exclude free-riders from their groups. The result was that cooperative types played together and earned more than less cooperative ones. Within-group differences declined and between group differences rose. And overall cooperation rose with repetition as subjects strove to be positioned to get into "better" groups when future preference rankings and regroupings occurred. Although the composition of the subject pool could not change during these experiments, if this had been an evolutionary setting in which types with higher payoffs become more numerous over time, selection would have favored cooperative behaviors (the highest contributors to the public good earned about 20% more than the lowest ones mainly because they played with other high contributors).

The experiments just mentioned do not try to distinguish the roles of environment and genes in accounting for subject behaviors. However, in the tradition of Boyd and Richerson (1985), Ben-Ner and I (1998) advocated a gene-culture co-evolutionary perspective on the origin of preferences as a key step towards understanding the two-way relationship between social organization and preferences. But what, exactly, are the respective contributions of culture and of genes to the evolution of human sociality? We inclined toward the view that the capacity to respond to moral socialization is a universal in *Homo sapiens*. Although I am reluctant to enter into a debate with someone from whom I have much to learn about evolutionary theory, I will reluctantly express my concern that Henrich's paper seems unclear about where to place the advent of human sociality.

Henrich emphasizes that cultural evolution can take place on scales of a few hundred to a few thousand years, far faster than biological evolution. And, citing Henrich et al. (2001), he argues that societies are less alike in their social norms than experimentalists who had previously studied only modern cosmopolitan populations had assumed. But where do the biological underpinnings of human pro-sociality figure in the account? As Bergstrom (2002, p. 83) writes after discussing Henrich and Boyd's (2001) solution to the problem of punishment, "There is room to question whether the visceral, seemingly irrational anger that people feel when they are cheated or otherwise violated can be explained by cultural transmission rather than as genetic hard-wiring."

Henrich's paper allows for culturally transmitted preferences to have been deepened by the biological selection that pro-social cultural forces make possible. But when did such biological selection take place? Did cultural differentiation lay the groundwork for selection of more cooperative genotypes in West Africa than in Amazonia, in Hungary than in New Guinea, and so on? Are some people, because of their genetic heritage, less easily socialized into certain cooperative norms than others?

---

[1] Subject *i* was more likely to be grouped with subject *j*, that is, when *i* ranked *j* highly and *j* ranked *i* highly.

The existing evidence seems consistent with the idea that all populations have more or less the same genetic receptivities to normative socialization, although there are differences among individuals in this respect just as there are differences among individuals in temperament, intellectual potential, and other attributes. If this is so, then the basic underpinnings of human pro-sociality must have been in place tens of thousands of years ago when the modern human sub-populations of Australia, Asia, Europe, etc. began their long term branching from a common ancestral population. Forces for homogeneity within groups—probably themselves embedded in genetic tendencies—could well have provided the crucial environmental soil in which the genetic fixation of these social propensities could occur, as Henrich argues. But it seems important to know to what extent the relevant cultural forces operated before the dispersal of human populations and the growth of large-scale societies that followed the agricultural revolution, to what extent only after them.

Perhaps it is not human pro-sociality in general that Henrich is trying to explain, but only certain forms of ultra-sociality—the readiness to put one's life on the line for the group in time of war, to rush into burning buildings to rescue those trapped inside, and, more prosaically, to do one's civic duty by studying the issues and voting. I do not doubt that cultural adaptation is central to making these things possible, and that relevant adaptations have occurred within historical time. But my views perhaps differ from Henrich's as to (a) the importance of these manifestations of human sociality compared to others and (b) the autonomy of these manifestations with respect to deep-seated and long-standing conceptual and emotional repertoires. I suspect that such aspects of human psychology as the desire to look good in one's own eyes, the concern with how one is viewed by others, and the willingness to join a common venture and to feel indignation at and incur costs to punish any who cheat when one is oneself sacrificing for it, long pre-date large scale social organization. Large scale societies undoubtedly built on these elements by defining how a self-respecting individual will behave, when he ought to feel shame, and so forth, but not by selecting for these basic elements themselves.

Despite these concerns, I find Henrich's suggestion of the joint cultural and genetic co-evolution of sociality to be attractive if pushed back to a sufficiently early date. More broadly, Henrich's research on the cultural specificity of cooperative norms (including Henrich, 2000) contributes enormously to the project of understanding how social organization affects human values and norms. In my view, too much of the project of the social sciences, and especially economics, has taken human preferences as givens and asked what social arrangements can best advance well-being in view of that constraint. It is important to recognize that preferences are also molded by our social environments, so that, when tinkering with our social institutions, we take stock of how we might be tinkering with ourselves (Ben-Ner and Putterman, 1998; Bowles, 1998).

References

Ben-Ner, A., Putterman, L., 1998. Values and institutions in economic analysis. In: Ben-Ner, A., Putterman, L. (Eds.), Economics, Values and Organization. Cambridge University Press, New York, pp. 3–29.

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

Bowles, S., 1998. Endogenous preferences: the cultural consequences of markets and other social institutions. Journal of Economic Literature 36, 75–111.

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

Ehrhart, K.-M., Keser, C., 1999. Mobility and Cooperation: On the Run. Working Paper 99s-24, CIRANO, Montreal.

Henrich, J., 2000. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. American Economic Review 90, 973–979.

Henrich, J., Boyd, R., 2001. Why people punish defectors. Journal of Theoretical Biology 209, 79–89.

Henrich, J., Bowles, S., Boyd, R., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review 91, 73–78.

Page, T., Putterman, L., Unel, B., 2002. Voluntary association in public goods experiments: reciprocity, mimicry and efficiency. Department of Economics, Brown University, Working Paper 2002-19.

Discussion

What can we learn from cultural group selection and co-evolutionary models?[☆]

Rajiv Sethi [a,*], E. Somanathan [b]

[a] *Department of Economics, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, USA*
[b] *Planning Unit, Indian Statistical Institute, 7 SJS Marg, New Delhi 110016, India*

The main thesis of Joseph Henrich's paper is that an understanding of cultural group selection and the co-evolution of genes and culture will promote our understanding of human psychology and preferences. This is an important and, in our view, correct assertion. While we agree with the basic thrust of the argument, however, we have a different perspective on some of its particulars. First, we believe it is useful to distinguish between unconditional altruism and reciprocity, both because the empirical implications of the two behaviors are quite different, and because reciprocity can evolve in a much larger class of environments than altruism. Second, we argue that Henrich's criticisms of some of the evolutionary models of "pro-social" behavior in the literature, particularly signaling models, are misplaced. Third, we disagree with the notion that pro-social preferences form the basis for the large-scale cooperation among non-kin which characterizes human societies.

Economists have long operated on the assumption that the hypothesis of material self-interest is an adequate approximation to human motivation in economic decision-making. This hypothesis has recently come under fierce challenge from experimentalists, who have documented robust and systematic departures from self-interest in controlled laboratory settings (see, for instance, Fehr and Gächter, 2000). Rejection of self-interest as a governing hypothesis does not, however, necessitate an embrace of altruism. Experimental subjects tend to care about fairness and efficiency as well as their own self-interest, but also reveal a strong propensity to reward generosity and punish opportunism on the part of others. Such preferences for reciprocity can indeed result in greater efficiency in some environments, but can also reduce efficiency in others. For instance, reciprocity tends to raise aggregate payoffs

in public goods games with punishment, while it lowers aggregate payoffs in ultimatum games. It is not pro-sociality with which humans are endowed, but rather a predilection for reciprocity which, given the right context, can have pro-social effects. It is the predilection itself, rather than its consequences in particular environments, that evolutionary models should account for.

An important feature that distinguishes preferences for reciprocity from altruistic preferences is the fact that under rather general conditions, the presence of people with such preferences, if observable, can induce opportunistic individuals to act pro-socially. They do so at some cost to themselves, but avoid the greater cost of being punished. The fact that the presence of reciprocators in a group induces all group members to act pro-socially can allow preferences for reciprocity to survive and spread in competition with self-interested preferences *even in the absence of assortative interaction*. This occurs because reciprocators do better when matched with opportunists in mixed groups than opportunists do when matched with each other. The presence of reciprocators in a group induces opportunists to take group-beneficial actions, leading to higher payoffs for all group members, while groups consisting solely of opportunists are subject to the free-rider problem. When the population share of reciprocators is small, most opportunists find themselves matched with other opportunists and hence their mean payoff, averaged across the entire population, is lower than that obtained by reciprocators (Gintis, 2000; Sethi and Somanathan, 2003).

This argument relies on preference observability and is therefore vulnerable to the 'green beard' critique which Henrich raises. What prevents opportunists who manage to be perceived as reciprocators from entering and dominating the population? The answer is that, in order to do so, the entrant must be able to mimic reciprocators perfectly. When imitation is imperfect, individuals are forced to make probabilistic assessments of the nature of those with whom they are interacting, based on the noisy signals that they receive. The conditions for the survival and spread of reciprocity are accordingly more stringent. The results obtained under complete preference observability can be recovered, however, if signals convey enough information (Sethi and Somanathan, 2001). Furthermore, any informative signal, no matter how weak, can result in the survival of reciprocators if interactions are repeated sufficiently many times (Guttman, 2003). Nothing in this argument requires that signals be costly. Models that require signals to be differentially costly for different types of agents do so only because they endow all types with the ability to produce exactly the same signal at all times.

The green beard criticism is that one cannot simply assume away a mutant self-interested type who copies the signal exactly. Whether or not such a perfect mimic arose in the course of human evolution is, of course, an empirical question. The fact that humans are skilled at forming strong (and sometimes mistaken) impressions of others based on the most fleeting and superficial interactions suggests to us that people do carry informative signals. The experimental literature is replete with evidence that face-to-face communication does indeed change behavior, and it is impossible to escape the conclusion that people's appearance and demeanor convey information that is put to use by others. Moreover, the fact that reciprocity is observed even in anonymous experimental settings with no scope for signaling does not in any sense weaken the case for a signaling approach to the evolution of reciprocity. Laboratory subjects are aware that there are reciprocators in the population of players facing them even if they cannot see their signals. How they got there, whether by

evolution through informative signals, or by cultural group selection, is irrelevant to their decision problem.

If signaling explains cooperation in humans, then why, asks Henrich, do not we see more cooperation in chimpanzees, elephants, and dolphins? We maintain that most large-scale human cooperation depends not on pro-social preferences, but rather on incentive structures in organizations which animals with binding cognitive limitations have been unable to develop. The idea that large-scale cooperation, at the level of the nation-state or the conglomerate firm, should be explained by pro-social preferences will strike most economists as quaint. Clearly, it is the differences in organization and incentive structures between say, the United States and Somalia, not differences in pro-social preferences, which explain the differences in the economic performance of these two societies. It is true, of course, that preferences for reciprocity can contribute to the efficient functioning of organizations. Markets with asymmetric information, and labor markets in particular, are affected by the presence of people motivated by reciprocity. Such preferences can allow equilibrium contracts that are Pareto-superior to the contracts that would be observed if all agents were self-interested (Fehr and Schmidt, 2000). But note that reciprocal preferences do not have to be directed towards large groups to play this role. It is quite enough that reciprocity plays a role in horizontal interactions among small groups of workers or vertical interactions between workers and their immediate supervisors. Economic and political organization, in fact, often seems to take advantage of the pro-social behavior that emerges in small group settings (Richerson and Boyd, 1999). Pro-social behavior directed towards society at large is, in our view, a less economically and socially important phenomenon than pro-social behavior directed towards small groups.

None of these considerations should detract from Henrich's basic point, that cultural group selection is an underexplored and potentially valuable component of any evolutionary account of human motivation. The models of cultural group selection and gene-culture co-evolution that he exposits are persuasive because they fit with some robust features of human behavior. Perhaps the most important of these is the tendency to acquire group identities and to be xenophobic. Cultural evolution is required to explain this since the actual identities that individuals take on vary enormously. Genetic evolution in response to cultural variation seems plausible since the behavior appears so universally. The application of such ideas to economic problems is just beginning (Somanathan and Rubin, in press).

Henrich's account of conformist transmission and prestige-biased transmission draws our attention to how much of human behavior is drawn from cues that are given by other people. This deserves further investigation in the realm of preferences for consumption, leisure, and other goods. How do these evolve with the economy? What does this imply for welfare? These are important questions, the answers to which may very well lie in a more complete understanding of cultural evolution.

References

Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. Journal of Economic Perspectives 14, 159–181.

Fehr, E., Schmidt, K., 2000. Fairness, incentives and contractual choices. European Economic Review 44, 1057–1068.

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

Guttman, J.M., 2003. Repeated interaction and the evolution of preferences for reciprocity. Economic Journal 489, 631–656.

Richerson, P.J., Boyd, R., 1999. Complex societies: the evolutionary origins of a crude superorganism. Human Nature 10 (3), 253–289.

Sethi, R., Somanathan, E., 2001. Preference evolution and reciprocity. Journal of Economic Theory 97, 273–297.

Sethi, R., Somanathan, E., 2003. Norm compliance and strong reciprocity. In: Gintis, H., Bowles, S., Boyd, R., Fehr, E. (Eds.), Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life. MIT Press, Cambridge.

Somanathan, E., Rubin, P.H. The evolution of honesty. Journal of Economic Behaviour and Organization, in press.

Discussion

Cultural group selection, coevolutionary processes and large-scale cooperation, by J. Henrich

James F. Crow[*]

Genetics Department, University of Wisconsin, 445 Henry Mall, Madison, WI 53706-1574, USA

Received 13 May 2002; accepted 12 September 2002

I greatly enjoyed the target article and found it both thoughtful and stimulating. Henrich points out, as many have, that cultural evolution is much faster than biological evolution and not constrained by the Mendelian rules of inheritance. In particular the wide extent of horizontal transmission and such social factors as punishment of non-conformers appear to be characteristically human. To me, however, the most interesting part of the article is Henrich's adapting to cultural evolution the Price (1970, 1972) equations for biological evolution, in which genetic change is allocated to within- and between-group components. It is that on which I wish to comment.

Such equations are not new to geneticists. I presented one such a half century ago (Crow, 1955, p. 58), and I doubt that I was the first. I regarded it as a multi-level extension of Fisher's Fundamental Theorem of Natural Selection (Fisher, 1930/1999). The first person to use covariances rather than variances and thereby extend Fisher's Theorem to traits other than fitness was Robertson (1966). Put these two ideas together and you have Price's equations. Nevertheless, Price derived his results from scratch and learned of the earlier work only after his paper had been written (see Price, 1972, p. 490).

Price and Hamilton were the creative giants in this field, taking up where Fisher left off. Both were interesting personalities. Each had a curious mixture of excessive modesty and self-effacement along with full confidence in his ideas. Both were eccentric, although Price was far more. Both were indifferent to their own personal safety. Both died young, Price by suicide and Hamilton from malaria contracted, perhaps recklessly, on a hypothesis-testing tropical expedition. Hamilton's collected papers (1995/2001) include not only his major articles, but also an introduction to each one in which he describes the circumstances under which it was written. He also gives considerable biographical information about Price and himself and describes their scientific relationship.

[*] Tel.: +1-608-233-4438; fax: +1-608-262-2976.
E-mail address: jfcrow@wisc.edu (J.F. Crow).

I have used a Price-type equation to discuss the measurement of the degree of population subdivision (Crow and Aoki, 1984). In discrete-generation form, this is

$$\Delta C = \frac{(B_w V_w + B_b V_b)}{W} \tag{1}$$

in which ΔC is the single-generation change in mean value of the trait in the whole population, B_w and B_b the within- and between-group genetic regressions of fitness on the value of the trait, V_w and V_b the within- and between-group genetic variances, and W is the population mean fitness. The model assumes that the trait is polygenic, without dominance or epistasis and with loose linkage. You will note that this is exactly the same as Henrich's Eq. (8), although the models are different. Henrich is tracing the change of frequency of a single allele, whereas Aoki and I considered the change of frequency of a multigenic trait.

As Henrich notes, within-group selection has usually been regarded as more important than between group for two reasons. First, means are less variable than individual items and the numbers are smaller. Second, migration tends to reduce between-group variance. If dominance (within-locus interactions) and epistasis (between-locus interactions) are included, the effect is to reduce the first term in parentheses of (1) and increase the second. The reason is that within-group selection depends on the (least squares) additive component of the genetic variance (the genic variance), whereas the between-group component is determined by the total genetic variance. This means that gene interactions increase group selection relative to individual selection.

Henrich employs the same equation, but with new symbols, for within- and between-group cultural selection. He then argues that there are four different cultural mechanisms that reduce the within-group variance. These are (1) conformist transmission, (2) prestige-based transmission, (3) punishment, and (4) normative conformity. If not unique to the human species, these are at least far more important than in other animals. These mechanisms decrease the variance within groups; in many cases, he says, this is reduced to essentially zero. Therefore, the between-group component becomes much more important for humans than for other species.

Homogenization within groups has been discussed by others (e.g. Sober and Wilson, 1998), but Henrich has provided a more detailed argument. Thus, the effect of his cultural transmission mechanisms is in the same direction as dominance and epistasis in a purely genetic model, although undoubtedly cultural effects happen much faster and the between-group component is likely to be larger.

The mechanisms of cultural inheritance are much more varied than biological inheritance, which is constrained by Mendel's rules. In particular, horizontal transmission is far more significant for cultural than for biological evolution. There is abundant evidence for the importance of between-family influences relative to those within the family in determining many behavioral traits. Parents seem to be less important culturally than peers or other extra-family influences. In contrast to biological evolution, a great deal of cultural inheritance can occur with no change of allele frequencies.

The Price–Hamilton equations have led to countless words and occasional quantitative discussions, but mostly to plausibility arguments. They provide a beautiful framework for extending the results of kin selection to related problems, such as selection based on group

recognition. There has also been considerable mathematical discussion of different models of population structure. But the general framework, exemplified by Eq. (1) above persists.

Can these equations be made more predictive? One suggestion came from Aoki and me (Crow and Aoki, 1984) at the time when the great power of molecular genetics was first becoming evident. Eq. (1) can easily be used to assess the importance of Hamilton's well-known inequality for an altruistic trait. Identifying B_w as the fitness cost, c, to the individual and B_b as the benefit, $b - c$, to the group, the condition for the altruistic trait to increase is Hamilton's condition:

$$\frac{c}{b} < r$$

where r is Wright's coefficient of relationship (Crow and Aoki, 1982). Henrich reaches a similar conclusion, although without considering genetic niceties.

Population subdivision produces effects within local populations that are like those of inbreeding. We can therefore use molecular measures of population subdivision to estimate r and assess how much group selection is inherent in the population structure. This is best estimated by Nei's G_{st} which can be determined from the average probability of identical alleles from within the group and from the whole population (Crow and Aoki, 1984). Then $r = 2G_{st}/(1 + G_{st})$. At the time the paper was written, Aoki had some measurements of G_{st} based on blood groups in Japanese macaques. The average value of G_{st} was about 0.09, corresponding to $r \approx 1/6$, and values for other primates are similar. Thus, provided the requisite genes exist in this monkey population, we would expect that an altruistic trait for which $c/b < 1/6$ would increase.

In our discussion, we assumed that the monkeys need not recognize kin, but rather need only identify group members. Each monkey, including migrants, was assumed to behave unselfishly with members of the group in which it resides. Thus, a standard measure of gene identity within and between groups, G_{st}, seemed appropriate. We thought, from certain human studies at the time, that altruism toward members of a group or tribe was more likely than recognition of kin, except for very close relatives.

Molecular techniques are much better now. It should be possible to make such measurements on a wide variety of populations. Then, if behavioral traits can be appropriately quantified, it should be possible to make comparisons between these traits and the genetic structure of the population. We would not expect the values to be the same; molecular markers are usually chosen to be neutral whereas behavioral traits are presumably under strong social selection. But the extent of the departure from molecular predictions would be an indication of the strength of social selection. I do not know whether this amalgamation of psychological and molecular measurements, and of social and genetic theory, can be fruitful, but perhaps it is worth a try.

Henrich points out that interaction of cultural and biological evolution may be significant. For example, it is reasonable that agricultural practices developed mainly from intra-group cultural selection. This need not have involved any change in gene frequencies. But some evidence suggests that agriculture spread through Europe by differential proliferation (migration). Thus, if this view is correct, within-group cultural selection interacted with between-group biological selection. Further studies of the interaction of biological and cultural selection should be of great interest.

I should make clear that the Crow and Aoki prescription applies to behavioral traits that affect biological fitness. The well-developed theory of the evolution of fitness and characters correlated with fitness is directly applicable. Cultural evolution, in contrast, does not have any mechanistic basis comparable to Mendelism nor are there any measurement techniques with power and precision comparable to DNA analysis. The haphazard rules of cultural inheritance—especially the preponderance of extra-family influences—means that such elegant formulations as Hamilton's rule are not available. On the other hand there is much less constraint: mating is not required, reproduction can occur at all ages, and transmission to a large number of persons is possible. No wonder the rate of change can be high. But, if this subject is to be quantified, other rules, even if less neat and less precise, will have to be developed.

References

Crow, J.F., General theory of population genetics: synthesis, Cold Spring Harbor Symposium on Quantitative Biology, 20, 54–59.

Crow, J.F., Aoki, K.I., 1982. Group selection for a polygenic behavioral trait: a differential proliferation model. Proceedings of the National Academy of Sciences U.S.A. 79, 2628–2631.

Crow, J.F., Aoki, K.I., 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proceedings of the National Academy of Sciences U.S.A. 81, 6073–6077.

Fisher, R.A., 1930/1999. The Genetical Theory of Natural Selection. Oxford University Press, Oxford (Variorum edition, with introduction and notes by H. Bennett).

Hamilton, W.D., 1995/2001. Narrow Roads of Gene Land. I. Evolution of Social Behavior. II. Evolution of Sex. Oxford University Press, Oxford.

Price, G., 1970. Selection and covariance. Nature 227, 520–521.

Price, G., 1972. Extensions of covariance selection mathematics. Annals of Human Genetics 35, 485–490.

Robertson, A., 1966. A mathematical model of the culling process in dairy cattle. Animal Production 8, 95–108.

Sober, E., Wilson, D.S., 1998. Unto Others. The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge.

Discussion

Cultural group selection, co evolutionary processes and large-scale cooperation (by Joseph Henrich)

Daniel Houser, Kevin McCabe*, Vernon Smith

*Interdisciplinary Center for Economic Science, George Mason University,
4400 University Drive, Mail Stop 1B2, Fairfax, VA 22030, USA*

Before addressing Joseph Henrich's evolutionary explanation for altruism, it is important to ask whether the existing evidence taken in its entirety actually suggests that preferences are non-selfish in the standard game-theoretic (myopic) sense. Our own research corroborates Henrich's claim that the amount offered in ultimatum and dictator games is greater than predicted by standard economic theory based on payoff dominance. We also find, however, that this fact alone is subject to over interpretation, and that its robustness should be tested in less restrictive interactions. Specifically, the interpretation of the results in these decision environments confound the following motives: (1) self-interest based on dominance; (2) positive reciprocity; (3) negative reciprocity; (4) various forms of other-regarding utility, altruism representing only one such form; and (5) the constant sum form of some games precludes effects based on gains from exchange. Even within this class of games, the altruistic utilitarian explanation for giving does not accord well with the fact that experimental instructions and procedures typically have a dramatic affect on outcomes (Hoffman et al., 1994, 1996). Thus, in the ultimatum game offers by the first mover vary from 44.4 percent of the available sum down to only 27.8 percent for subjects from the same American undergraduate culture. Moreover, similar large increases in ultimatum offers are observed by simply adding to the instructions two short sentences prompting the subjects to think about what they expect the other player to choose before making their choice (Hoffman et al., 2000). This "innocent" attempt by the experimenters to get the subjects to think about the strategic aspect of the interaction apparently had the effect of focusing the proposer's attention on the possibility that his/her offer might be rejected by the responder. Hence, the critical importance that the experimenter not attempt *any instructional or procedural variations* without performing controlled comparisons. This maxim creates a nightmare of complication for important, but hazard-laden, programs designed to make cross-cultural comparisons.

---

* Corresponding author. Tel.: +1-703-993-4856; fax: +1-703-993-4851.
E-mail address: kmccabe@gmu.edu (K. McCabe).

The evidence we will cite is consistent with self-interested reciprocity as the primary motive for departing from dominant strategy equilibrium in experimental environments. Thus, in the investment trust game, which is formally equivalent to the dictator game, giving is greatly increased over that in the dictator game by the prospect of gains from exchange through reciprocity (Berg et al., 1995). We also have studied two versions of a trust game (McCabe et al., 1996; McCabe and Smith, 2000) that provide a direct test of altruism. Player 1 chooses between altruism and passing to player 2, who can then choose between the narrowly self-interested dominant strategy equilibrium and a cooperative outcome based on reciprocity. In both versions—one of which allowed player 2 to punish defection if player 1 did not reciprocate—we observed not even a single case in which altruism was chosen. Thus, "altruism" does not survive when the choice set is expanded to include competing alternatives.

Moreover, in McCabe et al. (2002) we report comparisons that provide a direct test of reciprocity against either self- or other-regarding utility. We find that the amount of cooperation by player 1 in a standard trust game depends crucially on the inferred "intentions" of the first mover, not the payoff preferences of player 2. In the case that player 1 was forced to "trust," by passing to player 2, the latter is much less likely to play cooperatively (reciprocate) than when player 1's trusting play is associated with a large opportunity cost.

In his article, Henrich dismisses reciprocity as a plausible evolutionary explanation for cooperation for reasons including (a) the theory has not yet been worked out; and (b) reciprocity (direct or indirect) seems implausible, even impossible, in large-scale societies where intertemporal transactions must be carried out with hundreds or thousands of different individuals. We hope for rapid progress in the former (see, e.g. Heiner, 2002), but believe the view that large-scale cooperation results from prosocial preferences is misconceived, implausible and ignores market institutions. Although a case can be made that impersonal market exchange grew out of personal exchange (Smith, 1998), the former has long dominated economic theory in accounting for the wealth of nations. Our reading of the evidence, including hundreds of market experiments, suggests that large-scale cooperation results from impersonal, selfish exchange that is mediated by institutions. The dramatic success of the online auction house eBay, where essentially anonymous buyers and sellers complete a very large number of successful transactions, provides a stark example of how well self-interested, but institution-mediated large-scale exchange can adapt quickly to a new technology. This example also illustrates the capacity for new market institutions to emerge from trial-and-error experimentation.

Henrich puts a new face on previously discredited evolutionary models of group selection in order to provide an answer to what some view as the "altruism dilemma." An important part of his argument is that conformist transmission and success copying can work both to enhance within-group homogeneity and to maintain between-group heterogeneity. This might be true if the variation is measured over very short time periods and with respect to individual decisions, but it seems less plausible that there could be actual strategy copying leading to longer-term within-group homogeneity, as Henrich suggests. The reason is that, while decisions are observable, strategies typically are not. Differences in strategies might tend to remain stable across individuals within a group even in the event that their decisions became quite similar.

Social dilemma experiments with the voluntary contribution mechanism provide evidence that this is, in fact, the case. Broadly speaking, it is usually found in these environments that contributions to the public good start at around 50 percent of subjects' aggregate endowment, and decline to around 10 percent or so of the aggregate if the game is repeated about 10 times (see, e.g. Houser and Kurzban, 2002). It is also routinely found that there is substantial heterogeneity in propensities to contribute to the public good, with some subjects initially contributing most of their endowment while others contribute none (see, e.g. Gunnthorsdottir et al., 2002). By the tenth round, most subjects contribute little of their endowment, yet this does not necessarily imply that they have adopted the free riding strategy. In fact, if the game is stopped and restarted, then we have found that those who were acting cooperatively initially tend to behave cooperatively again (Houser and Kurzban, 2003). This suggests that while decisions across individuals became similar over time, underlying individual strategies displayed no such tendency.

We are not aware of any evidence, experimental or otherwise, that strategies, or beliefs and preferences (strategies' primitives) are as malleable as Henrich suggests. Opportunistic, self-interested decision copying might occasionally arise, but if preferences are generally stable, and if there are a few free-riders in the population, then it does not seem likely that decision copying could lead to sustained cooperation. In particular, we are aware of no systematic experimental evidence that non-cooperators tend to copy cooperators' decisions in a way that generates sustained cooperation, even when cooperators are in the majority. Quite the opposite: experimental evidence suggests strongly that it is typically cooperators who make the greatest movements toward the decisions of dogmatic non-cooperators (Andreoni, 1995; Gunnthorsdottir et al., 2002), with only the speed of this adjustment perhaps depending on a group's type composition (Houser and Kurzban, 2003).

Although we are sympathetic with the important objectives of his article, in our view it is implausible and experimentally unsubstantiated that the cultural transmission mechanisms Henrich identifies could lead to homogeneous beliefs, strategies or preferences. It is possible that decision copying, reinforced with the punishment mechanisms Henrich discusses, could generate somewhat homogeneous decisions over very short time periods. However, persistent and fundamental within-group differences in preferences, strategies and beliefs, particularly when combined with the stochastic forces Henrich mentions (plagues, environmental disaster), are quite likely to disrupt this homogeneity. Consequently, the effect and importance of co-evolution seem open to question.

An alternative, perhaps simpler, explanation for cooperation is that people have a propensity to reciprocate because it is in their self-interest to do so. Nature has sorted this out over the last 2–3 million years or more, and, in small interdependent social groups, provided humans with the ability to delay immediate gratification in the pursuit of greater mutual gains. Nature also gave humans the ability to reflect on their behavior and to evolve social tools (i.e. institutions) which allow ever-greater gains from impersonal trade. These emergent market institutions yield high fitness because they produce substantial personal and social benefits over large groups of strangers at relatively low cost, and require only non-cooperative individual strategies. Variation in the nature and implementation of institutions gives rise to the diverse "types" of cooperation that we observe across social groups.

We agree with Henrich that cultural co-evolution likely did occur. Certainly, it is not controversial that genes may respond to changes in their setting. We know that they respond

to changes in the natural environment (how did the Giraffe come to be 20 ft tall?), and expect that the same would be true of the cultural environment. The challenge is to determine the appropriate mechanisms and behavior to be understood.

References

Andreoni, J., 1995. Cooperation in public goods experiments: kindness or confusion? American Economic Review 85, 891–904.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10, 122–142.

Gunnthorsdottir, A., Houser, D., McCabe, K., Ameden, H., 2002. Disposition, history and contributions in a public goods experiment. Working paper, George Mason University.

Heiner, R.A., 2002. Robust evolution of contingent-cooperation in pure one-shot prisoners' dilemmas. Working paper, George Mason University.

Hoffman, E., McCabe, K., Shachat, K., Smith, V., 1994. Preferences, property rights, and anonymity in bargaining games. Games and Economic Behavior 7, 346–380.

Hoffman, E., McCabe, K., Smith, V., 1996. Social distance and other-regarding behavior. American Economic Review 86, 653–666.

Hoffman, E., McCabe, K., Smith, V., 2000. The impact of exchange context on the activation of equity in ultimatum games. Experimental Economics 3, 5–9.

Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. American Economic Review 4, 1062–1069.

Houser, D., Kurzban, R., 2003. Conditional cooperation and group dynamics: experimental evidence from a sequential public goods game. Working paper, George Mason University.

McCabe, K., Smith, V., 2000. A comparison of naïve and sophisticated subject behavior with game theoretic predictions. Proceedings of the National Academy of Sciences 97, 3777–3781.

McCabe, K., Rassenti, S., Smith, V., 1996. Game theory and reciprocity in some extensive form experimental games. Proceedings National Academy of Science 93, 13421–13428.

McCabe, K., Rigdon, M., Smith, V., 2002. Positive reciprocity and intentions in trust games. Journal of Economic Behavior and Organization 52, 267–275.

Smith, V., 1998. The two faces of Adam Smith. Southern economic association guest lecture. Southern Economic Journal 65, 1–19.

Discussion

Reply

Joseph Henrich*

*Department of Anthropology, Emory University, Geosciences Building,
Atlanta, GA 30322-1720, USA*

I am delighted by the engaging and constructive nature of these commentaries and would like to heartily thank all the contributors. In many positive responses, several authors built on the ideas contained in my article or developed parallel theoretical lines. Thus, rather than responding to each commentary individually, I have organized my reply around six recurrent themes. After discussing these themes, I will clean up a few loose ends from specific commentaries.

1. Point of departure: one cannot assume 'institutions' into existence

Several of the commentators suggested that 'institutions,' 'markets' or 'organizations' in one form or another might make an important contribution to the emergence of large-scale cooperation (e.g. Gintis; Hauser, McCabe and Smith; Sethi and Somanathan). However, assuming 'institutions,' 'organizations', or some undefined ability to build these, grants too much of the solution, if one (such as I) is interested in evolutionary origins. Such assumptions merely push the question back to 'what evolutionary processes explain the emergence of these institutions and who is paying the evolutionary costs for maintaining them?' My goal, which differs from most economists, is to establish the processes that led from the common ancestor of chimpanzees and humans to modern humans in all their diversity. As far as we know, non-human primate groups do not build institutions and organizations that vary across social groups and through historical time; consequently, I cannot begin with 'institutions' in place: cooperative institutions (or institutions of any kind) must have evolved from primordial origins. Of course, it seems perfectly acceptable for theorists interested in different questions to provisionally assume institutions, as long as the institutions themselves are understood to be in need of further explanation, and culturally

* Tel.: +1-404-727-5248; fax: +1-404-727-2860.
E-mail address: jhenric@emory.edu (J. Henrich).

and historically contingent. As I will discuss in more detail below, imprecise claims that envision 'institutions' as the products or byproducts of undefined 'cognitive abilities' (Sethi and Somanathan), 'reflection' (Hauser, McCabe and Smith), and language and foresight (Cohen, Axelrod and Riolo) seem unlikely to solve the critical problems (common intuitions tend to fail in this regard), and generally have not been precisely specified and formally modeled.

Gintis's model nicely illustrates the importance of removing such background assumptions. I focus on Gintis' work because he develops a formal model in his commentary that illustrates my point. In building his model, Gintis assumes the existence of institutions that socialize individuals for altruism and sets the strength of this socialization process to γ. To this learning process, Gintis adds vertical transmission and a replicator dynamic (which, in my terminology could formalize prestige-biased transmission). With this setup, he derives the conditions for a stable cooperative equilibrium. His derivation shows that if $\gamma = 0$—which it certainly must have at some point in human evolutionary history since other primates do not have institutions that socialize for group-level altruism—altruism will not evolve. Gintis's solution only works if one assumes into existence socializing institutions that promote altruistic behavior. It is not that I doubt the existence of such institutions (religions, schools, rituals, age rites) or that they have important effects on social behavior. Rather, I want to know how cooperative nstitutions, such as those related to religion, arose in our primordial primate ancestors. If institutions can be thought of as alternative culturally-evolved equilibria (Young, 1998), then cultural group selection provides a mechanism for selecting among institutions in favor of those that promote group-beneficial behaviors. Wilson (2002), Kelly (1985) and Stark (1997) provide three well-researched examples of cultural group selection operating in historical time.

Scholars focusing on the contemporary world in industrialized societies sometimes forget that throughout most of human history there were no market institutions, nation-states or corporations (e.g. Hauser, McCabe and Smith). Yet, by inference from the ethnographic record of small-scale societies and the archaeological record, Paleolithic peoples likely cooperated in n-person dilemmas in hunting, food sharing, territorial defense, fishing and warfare. So, while I have little doubt that cultural group selection favors institutional forms such as markets and corporate hierarchies that make effective use of self-interested incentives and our social psychology (Richerson and Boyd, 1998), the problem of large-scale cooperation was likely solved long before these institutions arose (in relatively recent historical time). As I note in the article, once the number of players in an n-person cooperative dilemma is greater than about 10, the canonical models of cooperation based on kinship and repeated interaction cannot explain very much cooperation, and the problem becomes exponentially worse as n increases. Furthermore, detailed empirical data from extant foraging groups indicates that the canonical evolutionary models (e.g. reciprocal altruism) *cannot* even account for the patterns of cooperative behavior in the simplest human societies (Hill, 2002). Thus, the roots of the solution to large-scale cooperation lie in the Paleolithic, not with markets, institutions, governments and police forces of some societies in more recent millennia.

My focus on evolutionary origins also relates to Cohen et al.'s questions about 'teachability.' In general, I think their emphasis on 'teachability' corresponds closely to my focus on social learning, but merely places more weight on the transmitter rather than the receiver. It

seems to me that the evolutionary onus will be on the learner for several reasons. First, those not familiar with the ethnographic record tend to drastically overestimate the amount of active teaching that occurs in human societies. In most small-scale human societies there is very little active teaching. Children learn by imitation and practice (Fiske, 1998; Hill, 2002; Lancy, 1996). This suggests that teaching by parents and schools as we know it in our society is the product of a particular trajectory of cultural evolution and tells us little about humanity or human nature. Second, for teaching to evolve to any significant degree (either genetically or culturally), children must have already evolved some substantial social learning abilities. Third, 'teaching' is its own cooperative dilemma. Genetic relatives (parents and siblings) have some fitness incentives to teach (kin selection), but they do not have incentives to teach 'pro-group' values that would be individually costly to the family (in fact, they have incentives to preach against such values), so the basic problem of teaching and altruism remains unsolved. The commentators might be interested in recent work by Castro and Toro (2002) and my own work (Henrich and Gil-White, 2001) in this regard.[1]

Norgaard asks why one needs to consider genes and culture together. In my view, the question of whether one needs to consider culture-gene coevolution or merely cultural evolution depends on the time-scale of the phenomena under investigation. On shorter time-scales (historical time-scales), genes can be assumed constant, and cultural evolution can be analyzed alone apart from genes (Henrich, 2001), as long as one realizes that the psychological mechanisms that give rise to and direct cultural change are founded on genetically-evolved capacities for social learning (and thus, understanding the evolutionary origins of these mechanisms might be useful). However, if one is interested in explaining the genetic and cultural evolutionary processes that transformed the common ancestor of chimpanzees and humans into the diversity of contemporary humanity, then, I would argue, one must consider culture-gene coevolution.

2. Culture is part of human biology

Comments made by Norgaard, Gintis and Putternman indicate that I must briefly clarify the relationship between culture and biology. In my view 'culture' or 'cultural capacities' should not be considered as separate from biology. Cultural information (learned preferences, practices, knowledge) is acquired by some organisms because they have genes that give rise to brains that can acquire and store information gleaned by observing and interacting with other organisms. For example, some rat species learn food tastes from their mothers and some birds learn songs from other members of their species. We would not say that a rat's particular dietary choices are not part of her biology because they were learned from her mother, or that bird song is a purely cultural phenomenon, divorced from the species' biology. Thus, the essential distinction is not 'culture versus biology,' but rests on where the behavioral information is stored and how it is transmitted across generations. At least in part, the rat's information about what to eat or the bird's information about what

---

[1] These commentators also pointed out that I did not cite Simon or Dewey. Along these lines, I also did not cite Darwin (1981), Baldwin (Baldwin, 1896a,b), Tarde (1903), Bandura and Walters (1963), Miller and Dollard (1941), Campbell (Campbell, 1965), Cavailli-Sforza (1971) and Pulliam and Dunford (1980), just to name a few.

to sing was not stored in their genes. Instead, some of this information was 'stored' in the behavior of other members of the species and transmitted by observational learning. From this perspective, culture is part of an animal's biology, but is not stored in its genes (and, as is often confused, 'biology' $\neq$ 'genetics').

Along these lines, Gintis divorces "socialization" (vertical and oblique transmission) from "biologically adaptive processes" (replicator dynamics). In my view, however, this is a mistake because our heavy reliance on high fidelity imitation and other forms of sophisticated social learning likely evolved as part of an adaptive learning complex that integrates a constellation of mechanisms. This combination of learning mechanisms was favored by natural selection because it generated lower cost, quicker, and more efficient adaptation—on-average—to the rapidly changing environments of human evolutionary history (Boyd and Richerson, 1985; Henrich and McElreath, 2003).[2] The presence of vertical transmission, for example, is likely part of a more extensive social learning adaptation that integrates a variety of mechanisms (e.g. prestige-biased, conformist and vertical transmission) over the developmental cycle of humans. Children, for example, may acquire an initial set of skills, preferences and social understandings by imitating their parents (vertical transmission) and older siblings; then, later as their circle of social contact expands, they may benefit from conformist and prestige-bias transmission (replicator dynamics).[3] Gintis's distinction seems to suggest that vertical transmission is not part of our evolved adaptation for imitative learning, yet evolutionary thinking suggests that under conditions of costly model-search (finding a good model to imitate) and costly access, some reliance on vertical transmission can be favored. Furthermore, there is no empirical reason to believe that what one learns from one's parents, siblings or peers is 'learned' or 'stored' in a psychologically different manner from what one learns from skilled, successful and prestigious individuals.[4]

3. Proximate psychological details

Some confusion was apparently created by my discussion of cultural learning and imitation because I avoided exploring the psychological details behind imitative learning processes such as conformist transmission and prestige-bias transmission. My views on the psychology of imitation are largely consistent with those of Tomasello (2000). Based

---

[2] It has been argued that cultural learning mechanisms are specialized adaptations to variable environments of the last couple million years and that humans merely anchor the tail of a distribution of learning adaptations that characterize a more general trend in several mammalian lineages (Henrich and McElreath, 2003; Richerson and Boyd, 2000).

[3] It is worth noting that extensive data from behavioral genetics indicates that vertical transmission plays little role in adult behavior (Harris, 1998). This does not mean that children do not learn from their parents, but it is evidence that whatever they do learn is swamped by non-vertical cultural transmission, perhaps during adolescence.

[4] I am not clear on what Cohen et al. refer to when they claim to disagree with my "minimization of the differences between how genes and culture are transmitted." One of the major points of the paper is rooted in the difference between genetic and cultural transmission, and the sections on conformist and prestige-biased transmission describe some of how genetic and cultural transmission differ. Interestingly, I am usually on the other side of this criticism, as I have spent substantial effort in specifying and formalizing the difference (Henrich and Boyd, 2002; Henrich et al., 2003c).

on his comparative work with human children and chimpanzees, Tomasello argues that at around nine-months human infants show clear evidence of developing a 'Theory of Mind'. Not found in other species, this cognitive ability allows individuals to represent others as possessing underlying goals and beliefs that causally generate actions. This ability seems to be achieved, perhaps, by allowing individuals to imagine themselves in the shoes of the individuals they are trying to imitate. The capacity leads children (and adults) to imitate both the motor patterns and underlying goals, strategies and affective responses of other individuals. Recent experimental work shows that this 'goal imitation' is in place by 14 months of age and that children will imitate adult motor patterns and infer their underlying goals, even when the adult fails to achieve the goal (Gergely et al., 2002). That is, 14 month-olds imitate 'goals' they never see achieved.

This line of research also addresses Possagennikov's concern about imitating rare punishment. Children and adults can imitate preferences and goals inferred from behavioral cues without actually seeing punishment carried out or the goal achieved.[5] Tomasello suggests that 'Theory of Mind' (which may have initially evolved to improve imitation) allows for the development of empathy (feeling the world from another's point of view) and thus seems linked to moral development.

Addressing Putterman (and Bergstrom, 2002) in this light, cultural learning makes connection between emotions and situations/circumstances (Ekman, 1999a,b; Fessler, 1999). Even among monkeys, individuals learn to connect 'fear' and 'snakes' by watching other members of their social group. Monkeys that have not observed a conspecific show fear towards a snake will not acquire a fear of snakes (presumably, until they are bitten). Humans, with their more powerful 'Theory of Mind' based imitation, learn to attach emotions to all kinds of things real and imaginary (more below). The cultural variation in what people are disgusted by, afraid of, angered by, and ashamed by vary enormously—much more than most non-anthropologists usually realize. Cultural learning allows individuals to adapt to the specifics of their local circumstances by connecting evolved 'basic emotions' and affective reactions to different contexts (norms, for example). Having said that, there is little doubt that human societies show some broad patterns in what affective responses get attached to what 'things in the world' (Brown, 1991). But, this does not change the fact that growing up in a particular place means that you experience different emotions in different circumstances (see Nisbett and Cohen (1996) below). Thus, this suggests that cultural learning likely has important effects on preferences.

We are just beginning to realize how much the human brain, and particularly the prefrontal cortex, continues developing and constructing itself over the first 20 years of life (Giedd et al., 1999; Quartz, 1999; Quartz, 2002). Perhaps illustrating this developmental process within a culturally-evolved social environment, recent experimental work with children in the US shows both our 'sense of fairness' and our 'taste for punishment' (as measured by the Ultimatum, Dictator and Public Goods Games) also continue developing over the first two

---

[5] Possagennikov is also concerned that the Henrich and Boyd (2001) model requires that conformist transmission be the same strength relative to prestige-biased transmission at higher punishment levels. My own recent work (unpublished) on this issue suggests that the relative strength of conformist transmission should have evolved to increase (not decrease) at higher punishment levels (*i*) because of the rising uncertainty in the payoff differences between alternative strategies (assuming noisy signals of payoffs)—although this conclusion is very preliminary.

decades of life (Harbaugh and Krause, 2000; Harbaugh et al., 2002). Finally, neuroimaging work shows that this culturally-constructed part of the brain, the prefrontal cortex, is the focus of activity while playing both 'trust' and 'reciprocity' games with adults (McCabe et al., 2001). This suggests that the 'preferences' measured by experimental games emerge during the first 20 years of life, and is likely affected by the culturally-evolved environments of ontogeny (Henrich, 2003).

4. What is the nature of the 'genetic part' of culture-gene coevolution for prosociality?

Several commentators pose an important question: if prosocial genes did follow on the heels of cultural group selection, what is the nature of these genes? I have no good answer to this question, other than to suggest some constraints and mention some possibilities. First, whatever these genes do, they would have to be integrated into the learning process, as they would have been genetically evolving in a diversity of social groups at different stable culturally-evolved equilibrium. Not only would these social groups have varied in the degree or scale of cooperation and punishment, their domains of cooperation would have varied from behaviors involving fishing, ritual, warfare, and group decision-making to hunting, food sharing, and public works (e.g. community house construction). Thus, we should not expect these 'prosocial genes' to lead to simple predispositions or 'tastes' in the same way that genes cause people to like fat, sugar, their own kids, and sex. Such genes must be intimately intertwined with the cultural learning process. Most likely, in my view, these genes would facilitate learning social norms (and some of these norms would involve cooperation dilemmas). For example, individuals who (because of their genes) can learn a norm from being punished only once will be favored over those who require two punishings (paralleling Simon's interest in docility). She who can infer the proper norm using observation, inference and conformist transmission,[6] and then internalize it (thereby avoiding all punishing from mistaken defections), may do the best.[7]

Linking with Gintis, if one internalizes a norm, she is probably less likely to mistakenly defect. As Gintis points out, these kinds of evolutionary processes may be the evolutionary origin of shame (Fessler, 1999), pride, guilt and anger at norm violations—which, with the possible exception of guilt, are universal human emotions that show tremendous cross-cultural variability in the kinds of situations to which they are applied to.

I hope this discussion clarifies that when I speak of 'prosocial behavior,' I not thinking of pure, simple-minded altruism (Sethi & Somanathan), but something like the 'Strong Reciprocity' proposed by Fehr, Gächter, Bowles and Gintis (Fehr et al., 2002a). Because many group-beneficial norms are fitness neutral within-groups, the bias may first emerge to pick more group-beneficial equilibria over less group-beneficial equilibria.

---

[6] Henrich and Boyd (2002) have shown how conformist transmission drastically reduces the inferential learning errors that are likely associated with some kinds of cultural transmission.

[7] In our model, Henrich and Boyd (2001) established the conditions under which a 'prosocial gene' that caused cooperators to mistakenly defect less often ('internalization') would be favored by natural selection. Interestingly, once such 'prosocial genes' have spread through the group, the maximum size (N) of the culturally-evolved stable cooperative group increases, illustrating one example of the culture-gene coevolutionary ratchet.

Coevolving with culture, Strong Reciprocity leads individuals to acquire, internalize and uphold norms (including cooperative ones), reward other norm upholders, and punish norm violators at a cost to themselves—even in one-shot situations. Contrary to Güth and Kliemt's assertion, a strong empirical case has been made for Strong Reciprocity that severs it from reciprocal altruism and the other canonical evolutionary models (i.e. from the 'Big Mistake Hypothesis'), and connects it to the kinds of theorizing discussed in my article (Fehr et al., 2002a). Furthermore, Fehr and I (2003) have argued, using evidence from foraging populations, paleo-archaeology and non-human primates, that the Big Mistake Hypothesis was never on firm ground in asserting that ancestral human societies lacked fitness-relevant low frequency (and one-shot) interactions. We argue that such interactions were fitness relevant and that the experimental data is consistent with a keen ability to distinguish situations involving substantial repeat interaction and reputation from situations without those.

Cohen and coworkers felt strongly that I gave "insufficient weight to more general properties of human beings—such as language and foresight—that have enormous survival value for many other reasons." This criticism is both odd and premature. It is odd because the authors themselves have published mathematical models on the evolution of cooperation that have neither language nor foresight built into them (Axelrod, 1984; Riolo et al., 2001)—so they have also given 'insufficient weight' to those capacities. They are premature because the available theoretical work suggests that cognitive abilities, foresight and language may negatively impact the evolution of cooperation. For example, Bendor (1993) and Bendor et al. (1996, 1991) have shown that when information about an individual's partner's cooperative contribution is noisy, strategies with memory (book keeping skills), 'cognitive processing' and 'foresight' did worse than simpler strategies with little memory or foresight. These authors argue that cognitively powerful strategies did poorly because the errors in perception caused them to randomly walk through the space between zero and one. Such smart strategies over-fit their noisy observations and 'over-responded' to deviations. Of course, as I mentioned in the paper, the success of a strategy depends on the mixture of other strategies in the population, so these findings need qualification. Nevertheless, they do indicate that we should be skeptical of the intuition that more powerful cognitive abilities improve cooperation.

The idea that 'language' somehow facilitates the evolution of cooperation is a common intuition (Henrich et al., 2003a; Smith, 2003). Unfortunately, language creates a whole new set of cooperative dilemmas. In evolutionary models, telling someone 'sure, I will cooperate', is cheap talk and not likely to solve one-shot *n*-person cooperation.[8] As an information spreading device, 'gossip' only allows for cooperation via indirect reciprocity if it is very accurate (Panchanathan and Boyd, 2003) and accuracy likely diminishes as group size increases. Furthermore, gossipers often have an incentive to provide incomplete or inaccurate information. Thus, despite common intuitions, it is far from clear that 'language' facilitates cooperation (Richerson et al., 2003).

One other clarification related to the Cohen et al. commentary should be entered at this point. These commentators mistook my criticism of existing models, saying "he [me]

---

[8] Perhaps language could facilitate costly signals, and thereby provide another avenue to solve some types of cooperative problems. Interestingly, costly signals might evolve *culturally* in humans, and cultural group selection may select among alternative stable costly-signally equilibria.

tends to imply that any model that fails to explain everything explains nothing." This was certainly not my intention! I tried to emphasize that models of kinship, reciprocity, and costly signaling help us tremendously to understand human social behavior, but that while they are successful in addressing many aspects of cooperation, they fail to explain large-scale cooperation.

5. Terminological issues

Both Guth and Kliemt, and Hauser, McCabe and Smith seemed to react to my use of the term 'group selection'. Because I believe that arguing about terminological labels wastes enormous amounts of time in the social sciences, I will only state that in using the term 'group selection' I follow Price's (1972) and Hamilton's (1975) practice of calling the between-group component of a selective evolutionary process (based on a partition of variance) 'group selection'. As Price (1970) was attempting to develop a general account of selective mechanics with genes as a special case, applying his formulation and terminology to cultural systems merely maintains consistency with his original endeavor. Furthermore, in applying the term 'group selection' to selection among multiple stable equilibria, I am following a practice that is consistent with Hamilton's and Price's emphasis on the 'between-group' term in the Price equation and is justified by usage in economics (Bergstrom, 2002), biology (Sober and Wilson, 1998) and anthropology (Boyd and Richerson, 1990). Concerns and confusions related to Wynne-Edwards (1962) work of over 40 years ago should, in my view, be relegated to history books.

6. The old guard

Of all the excellent commentaries, including some from eminent population geneticists and biologists, only Hauser, McCabe and Smith (HMS) defend the 'old guard' position against anything that smacks of 'group selection' or 'culture'. Following the old guard's usual rhetorical tactic, these economists dismissed my argument with a hand wave, claiming that "Henrich puts a new face on previously discredited evolutionary models of group selection." Nevertheless, anyone who understands the theoretical and empirical reasons why biologists and anthropologists have been skeptical about the importance of *genetic* group selection for our species (a skepticism I share) will see that what I am proposing does not suffer from the same criticisms. The standard arguments against genetic group selection do not apply to cultural group selection or selection among multi-stable equilibria.

HMS miss one of the main points of my article when they express skepticism about the long-term effects of conformist transmission on within-group homogeneity. Putting aside the empirical question about conformist transmission, within-group homogeneity is not crucial and suggests a deeper misunderstanding. What matters is that within-group learning processes generate locally stable equilibria (these can be polymorphic equilibria), and that there are multiple stable equilibria that generate different group-level properties (such as having different mean group payoffs: Young, 1998). As long as local interaction and learning generate such multiple stable equilibria, cultural group selection can gradually

sort among these equilibria. This has never been discredited, as HMS apparently do not realize. And, if it had been, readers venturing into new territory should trust that one of our theoretical biologists—perhaps Crow or Aoki—would have mentioned something in their commentaries.

In the target article, I used the specific examples of prestige-biased transmission and conformist transmission to illustrate how the process might work for two reasons. First, a combination of field and experimental data from psychology, economics and anthropology (see target article) show that both types of cultural transmission are an important part of our learning repertoire, so these mechanisms are real, not theoretical speculation. Second, two kinds of theoretical foundations are already in place: (1) formal analyses show that both prestige-biased and conformist transmission will evolve genetically under a wide range of conditions (Henrich and Boyd, 1998; Henrich and Gil-White, 2001), and (2) Boyd and I have shown that prestige-biased transmission with just a bit of conformist bias will generate multiple-stable equilibria in *n*-person one-shot cooperation-punishment dilemmas (Henrich and Boyd, 2001).

However, a variety of cultural learning processes may create multiple stable equilibria, which cultural group selection can select among. Aoki's proposal is an interesting example. Interpreted as a cultural evolutionary process, his model could explain a wide range of cultural diversity (e.g. in mate preferences) and could generate cultural group selection among groups with different average mate preferences (we already know that mate preferences vary both cross-culturally and historically). Similarly, cultural evolution through the logic of costly signaling can generate an immense variety of stable signaling behaviors, some of which may be group-beneficial. Cultural group selection will favor groups that happen to land on signaling equilibria with group-beneficial properties.

HMS express skepticism about the effectiveness of cultural learning on social behavior and state that "we are not aware of any evidence, experimental or otherwise, that strategies, or beliefs and preferences are as malleable as Henrich suggests." In discussing evidence for their 'hardwired view' the commentators mention only data from experimental economics laboratories (and mostly from their own labs). There are several problems with their approach. First, they seem to ignore the possibility that cultural learning (of which conformity is one aspect) operates on our social behavior, preferences and strategies during our extended period of brain development (ages 0–20), which occurs principally in the prefrontal cortex (a center of moral decision-making and 'game play'). It is certainly plausible that, during a 10-period experiment, people might not throw away the strategies and preferences that were 'built-in' to their brain architecture during cultural ontogeny (Henrich, 2003). Thus, people's resilience to substantial preference change during a brief experiment reveals little on this issue.

Second, outside of economics laboratories, the social science literature reveals substantial amounts of evidence for culturally-malleable preferences. For example, Nisbett and Cohen (1996) have shown that culture runs deep, links to our emotions, and has important effects on social behavior. Combining a variety of field and experimental results, these psychologists show that 'white rural Southerners' have a more aggressive and violent reaction to actions that could be construed as insults to one's personal honor than Northerners—the researchers call this a 'Culture of Honor' and trace its historical origins. In one telling experiment, the researchers show the 'being bumped' in a hallway had the effect of dramatically raising the

Southerners cortisol (stress hormone) and testosterone (a hormone associated with 'readiness for aggressive interaction') relative to Northerners and control subjects. The authors further provide data that this cultural difference explains differences in the patterns of murder and military service rates between the North and South (controlling for things such as income and population). Their results strongly suggest that growing up in the rural South changes one's social behavior—right down to the hormones—vis-à-vis one who grows up in the Northern parts of the US. Consistent with the above discussion of cultural ontogeny, when Southerners attend Northern universities, they do not suddenly change their culturally learned preferences and emotional responses; they still behave like Southerners.

Third, even if we restrict ourselves to experimental games, the available database shows tremendous variability in preferences and is consistent with a wide range of work outside of Economics. Using experimental tools such as the Ultimatum Game and Public Goods Game, a wide range of cross-cultural data show substantial variation in 'tastes for punishment' and 'notions of fairness' (e.g. Henrich et al., 2003b; Kachelmeier and Shehata, 1997; Ockenfels and Weinmann, 1999; Smith, 2000). While HMS attempt to dismiss this substantial set of results by pointing out the effects of contextual manipulations and suggesting that differences in linguistic translation might account for the variation, a closer look at the fine details of each study shows that substantial cultural variation remains even after language and experimenter are taken into account.[9] *Within* the following studies both language and methodological protocols are constant, yet significant differences in behavior remain: Patton (in press), Henrich and Smith (in press), Smith (2003), Kachelmeier and Shehata (1997), Paciotti and Hadley (2003), Gurven (in press) and Ockenfels and Weinmann (1999).

HMS and I agree on the importance of context in understanding economic and experimental behavior. However, whereas HMS maintain that this context depends on the operation of a human-universal psychology built on reciprocal altruism (and thus invariant across the species), I expect the importance of specific contextual manipulations will vary across cultures, and depend on cultural learning within the context of daily life. Thus, where HMS predict that the same contextual manipulation in experimental games will have the same effects in different places, I predict that 'context-effects' in experiments will vary across societies in accordance with real life in those places. Along these lines, (Hayashi et al., 1999; Pingle, 1995) show that simple framing differences strongly affect rates of cooperation in a 2-person prisoner's dilemma. Contrary to HMS's view, the emergence of the context effects depends strongly on whether one is from Japan or the US. Similarly, Kachelmeier & Shehata in a Public Goods experiment comparing Canadian, mainland Chinese and Hong Kong students, showed that low anonymity conditions led Asian students, especially Mainlanders, to behave very cooperatively while having no effect on Canadians (also see Ensminger,

---

[9] HMS also exaggerate the effects on context vis-à-vis the Ultimatum Game (Camerer, 2003). The trio acts as if 'any old change' in the protocol would achieve large variation. When, clearly, they specifically chose to test protocol modifications that they thought would produce significant variation. For example, in the UG game with the mean of offer of 27.8 percent, the experimenters introduced a sense of entitlement on the proposer by allowing her to win the role of 'proposer' in a trivia contest—everyone knows, without hearing the results, that this is going to make a huge difference. These kinds of contextual modifications (adding a trivia contest) do not just 'slip into' one's translation (and testing) of written protocol instructions. What HMS do not reveal is that an enormous variety of protocol modifications in the Ultimatum Game produce statistically identical results among university students. This is one of the reasons that we chose this game for our cross-cultural tests (Henrich et al., 2003b).

in press). The available evidence suggests that 'contextual effects' vary, at least to some degree, across different societies.

Fourth, while the subjects playing experimental games should not be expected to drastically alter the preferences they have acquired over a lifetime, a wide range of experimental results from economics and elsewhere shows that experimental subjects do use mechanisms like prestige-biased transmission and conformist transmission in adapting to experimental situations (Ball et al., 2001; Eckel and Wilson, 2000; Gaechter and Falk, 2002; Kroll and Levy, 1992; Offerman and Sonnemans, 1998; Pingle, 1995; Pingle and Day, 1996; Wit, 1999).[10] For example, in a study of the effect of social influence on common-pool resource games, Smith and Bell (1994) used confederates to show that players often imitate the behavior of other players when they are uncertain of what to do, even when it is against their self-interest to do so.

Finally, HMS defend the Big Mistake Hypothesis when they propose "an alternative, perhaps simpler, explanation for cooperation is that people have a propensity to reciprocate because it is in their self-interest to do so." Besides not acknowledging the substantial experimental evidence to the contrary and the basic theoretical problems with their approach (Fehr et al., 2002b; Fehr and Henrich, 2003), the commentators claim belies a misunderstanding of one of my article's principle points. I proposed a line of evolutionary reasoning consistent with the logic of natural selection and the optimization of fitness under informational constraints. In this class of models, if prosocial behaviors or psychologies evolve, it is because they were favored by natural selection (i.e. they are fitness maximizing under informational constraints). And, there is no qualitative difference between an 'informational constraint' and the constraints on the mutational spectrum required in models of reciprocal altruism (actually, I think that the information constraints are more plausible than the mutational constraints required for reciprocity models).

7. Evolutionary approaches and rational choice approaches are complementary

Khalil's commentary ends by presenting us with a choice between "the rational theory of choice" and "Darwinian theory basically of no choice." In setting up this unnecessary opposition, Khalil misunderstands the research program of which this paper is a part. Standard economic approaches take preferences as exogenous. Darwinian approaches, which include models of both cultural and genetic evolution, provide a theoretical foundation for explaining the origin and evolution of human preferences (one can and should account for 'tastes'). Thus, the two sets of theoretical approaches can be quite complementary. Researchers like Jack Hirschleifer (1977, 1978) have long recognized the potential contributions of a Darwinian approach to standard economic theory, but only recently has the potential contribution of the cultural evolutionary component of the Darwinian paradigm been clearly recognized (e.g. Bowles, 1998).

The power of the synthesis is now emerging in behavioral economics with work on social preferences. Recent research has revealed two relevant patterns. First, behavioral

---

[10] In selecting these citations, I have restricted myself to economic experiments. If I were to bring evidence more broadly from social psychology and sociology, the list would expand dramatically.

economists have shown that adding various kinds of non-selfish preferences into standard models allows them to account for a substantial range of behavior among undergraduates in experimental settings (Bolton and Ockenfels, 1999; Charness and Rabin, 2000; Fehr and Schmidt, 1998)—although what exactly those preferences should be is a matter of some debate. Second, experimental work suggests that underlying preferences vary across human groups (Henrich et al., 2003b). These two findings lead to two questions that are well-suited to Darwinian Theory: (1) what are the genetically evolved foundations of human social preferences, and how can those be explained? (2) How can these genetic foundations be influenced and modified by individual and cultural learning to allow individuals to flexibly adapt to the range of social, economic, and institutional environments (i.e. multiple equilibria) that they might find themselves in? Addressing these questions not only provides a theoretical foundation for the origins of human preferences, it also feeds back into the construction of improved behavioral models and the positing of non-selfish preferences (e.g. Fehr et al., 2002b; Fehr and Henrich, 2003).[11]

8. Clean up

Cohen, Axelrod and Riolo criticize my assessment of the explanatory breadth of existing models (i.e. my "unsound rhetoric") by pointing to their own model of the evolution of one-shot cooperation based on "tags" (Riolo et al., 2001). In highlighting their model, they claim that it explains "precisely the kind of one-shot, anonymous situations for prosocial action that Henrich claims cannot be explained with simple models that do not rely on culture-gene coevolutionary processes." This claim combined with the Riolo et al. model demonstrates two standard flaws in thinking about cooperation that I discussed in my target article. First, Riolo et al.'s model is premised on a rather untenable constraint in the mutational spectrum: no pure defectors allowed! Richard McElreath and Roberts and Sherratt (2002) have shown that, as soon as defectors are allowed into the mix, cooperation completely collapses. While some kind of constraint is necessary, constraining unconditional defectors out of existence does not seem plausible since 'defector' is presumably the primordial state. Thus, Riolo et al. is your basic greenbeard model in disguise. McElreath has further shown (analytically) that this kind of mutational bias against defectors will stabilize one-shot cooperation in the Prisoner's Dilemma for a wide range of conditions *without* any tags—the tags are superfluous with that kind of mutational bias. In their reply to the Roberts and Sherratt's devastating critique in *Nature*, Riolo et al. state that a weaker mutation bias in favor of defectors (rather than a complete restriction) allows for some cooperation to evolve—but why should any bias against defectors be justified? They further mention that 'language' may provide a 'hard to fake' tag (a 'hard to grow' greenbeard). Interestingly, their suggestion has already been extensively (and unsuccessfully) explored (see Nettle and Dunbar, 1997), and further critiqued (McElreath et al., 2003). McElreath et al. show that culturally-transmitted 'tags' will substantially affect social behavior in coordination (but

---

[11] Incidentally, Khalil's idea that "Darwinian theorizing" leads to models in which individuals are immune to flexibility and incentives is bizarre. The entire field of Human Behavioral Ecology is premised on this opposite assumption—I recommend Cronk et al. (2000) as an introduction to the field.

not cooperation) dilemmas, and that this will lead to a genetic response (and they explicitly simulated the coevolutionary culture-gene interaction).

Second, the commentators' claim that their model addresses "precisely" the kind of cooperation I am interested in shows that they have not yet recognized that 2-person interaction models do not generalize to *n*-person cooperative situations (Boyd and Richerson, 1988). For example, as I pointed out in the target article, tit-for-tat does not produce much cooperation once *n* goes very far above two. Thus, given that their model involves two-player interactions, it cannot go very far in explaining *n*-person cooperation.[12]

Wilson is concerned that my claim that "if groups are randomly remixed every generation ... altruism cannot evolve" is not accurate.[13] My derivation is based on assuming a large (infinite) population (Hamilton, 1975). However, if groups are small, drift may be enough to maintain sufficient variation in the population for group selection to operate. However, in genetic systems and among species with human demographics, strong within-group selection and migration between groups will generally destroy the variation between groups produced by drift. Boyd et al. (2003) have recently examined the interaction between group size (drift), cooperation and punishment vis-à-vis the question of whether cultural or genetic group selection might favor cooperation. Given what little is known about ancestral human societies, they show that while pure genetic group selection is unlikely, even given that a combination of cooperation and punishment make it more likely, cultural group selection will favor group-beneficial combinations of cooperation and punishment, especially when 'punishing' is potent. Note, this is a 'pure' cultural group selection model, not selection among multiple stable equilibria.

Aoki pointed out (Aoki et al., 1996) that the spread of early farming in Europe may be closer to prestige-biased group selection than demographic swamping. Based on his analysis, I agree. And, from an ethnographic and historical perspective, this mechanism is quite plausible.

Finally, Crow and Wilson suggest that complex interactions between genes and between genes and behavior (or genes and psychology), can favor group selection. This nicely puts my argument about culture-gene interaction in a broader evolutionary context, and is provocative since the merely 1.3 percent difference in the DNA of humans and chimpanzee gives way to a 5.5-fold difference between the gene expression in the brains of chimpanzees and humans—suggesting there must be lots of complex interactions between genes, and between genes and environments (Enard et al., 2002).

References

Aoki, K., Shida, M., Shigesada, N., 1996. Traveling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. Theoretical Population Biology 50, 1–17.

[12] Riolo et al. also made the modeling mistake of using only one finite group in their simulation, making their results dependent on drift (McElreath personal communication). This same mistake was also made by Nowak and Sigmund (1998), and critiqued by Leimar and Hammerstein (2001).

[13] Wilson suggests that 'my error' results from the use of absolute instead of relative fitness. I am not sure how he arrived at this conclusion, given that I used the Price equation, which is based on relative fitnesses, and followed Hamilton (1975) closely. I think the text explains the real source of our disagreement.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.

Baldwin, J.M., 1896a. A new factor in evolution. American Naturalist 30, 441–457 and 536–554.

Baldwin, J.M., 1896b. Physical and social heredity. American Naturalist 30, 422–428.

Ball, S., Eckel, C., Grossman, P., Zame, W., 2001. Status in markets. Quarterly Journal of Econoomics 155, 161–181.

Bandura, A., Walters, R.H., 1963. Social learning and personality development. Holt, Reinhart and Winston, New York.

Bendor, J., 1993. Uncertainty and the evolution of cooperation. Journal of Conflict Resolution 37, 709–734.

Bendor, J., Kramer, R.M., Stout, S., 1991. When in doubt . . . cooperation in a noisy prisoners-dilemma. Journal of Conflict Resolution 35, 691–719.

Bendor, J., Kramer, R., Swistak, P., 1996. Cooperation under uncertainty: what is new, what is true, and what is important. American Sociological Review 61, 333–338.

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

Bolton, G., Ockenfels, A., 1999. A theory of equity, reciprocity and competition. American Economic Review 90, 166–194.

Bowles, S., 1998. Endogenous preferences: the cultural consequences of markets and other economic institutions. Journal Economic Literature XXXVI, 75–111.

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

Boyd, R., Richerson, P.J., 1988. The evolution of reciprocity in sizable groups. Journal of Theoretical Biology 132, 337–356.

Boyd, R., Richerson, P.J., 1990. Group selection among alternative evolutionarily stable strategies. Journal of Theoretical Biology 145, 331–342.

Boyd, R., Gintis, H., Bowles, S., Richerson, P.J., 2003. The evolution of altruistic punishment. PNAS 100, 3531–3535.

Brown, D.E., 1991. Human Universals. McGraw-Hill, New York.

Camerer, C., 2003. Behavior Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton.

Campbell, D.T., 1965. Variation and selective retention in socio-cultural evolution. In: Barringer, H.R., Glanksten, G.I., Mack, R.W. (Eds.), Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory. Schenkman Publishing Company, Cambridge, Massachusetts, pp. 19–49.

Castro, L., Toro, M.A., 2002. Cultural transmission and the capacity to approve or disapprove of offspring's behavior. Journal of Memetics, 6.

Cavailli-Sforza, L.L., 1971. Similarities and dissimilarities of sociocultural and biological evolution. In: Hodson, F.R., Kendall, D.G., Tauta, P. (Eds.), Mathematics in the Archaelogical and Historical Sciences. Edinburgh University Press, Edinburgh, pp. 535–541.

Charness, G., Rabin, M., 2000. Social Preferences: Some Simple Tests of a New Model.

Cronk, L., Chagnon, N., Irons, W. (Eds.), 2000. Adaptation and Human Behavior. Evolutionary Foundations of Human Behavior. Aldine De Gruyter, New York.

Darwin, C., 1981. The descent of man in relation to sex (J. Murray London). Princeton University Press, Princeton.

Eckel, C., Wilson, R., 2000. Social Learning in a Social Hierarchy: An Experimental Study.

Ekman, P., 1999a. Basic emotions. In: Dalgleish, T., Power, T. (Eds.), The Handbook of Cognition and Emotion. Wiley, Sussex, pp. 45–60.

Ekman, P., 1999b. Facial expressions. In: Dalgleish, T., Power, T. (Eds.), The Handbook of Cognition and Emotion. Wiley, Sussex, pp. 301–320.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E., Paabo, S., 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296, 340–343.

Ensminger, J., in press. Market integration and fairness: evidence from ultimatum, dictator and public goods experiments in East Africa. In: Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (Eds.), Cooperation, Punishment and Self-Interest: Ethnographic and Experimental Evidence from 15 Small-Scale Societies. Oxford University Press, New York.

Fehr, E., Henrich, J., 2003. Is strong reciprocity a maladaption? In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge.

Fehr, E., Schmidt, K., 1998. A Theory of Fairness, Competition, and Cooperation.

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

Fehr, E., Fischbacher, U., Gaechter, S., 2002b. Strong reciprocity, human cooperation and the enforcement of social norms. Human Nature 13, 1–25.

Fessler, D.M.T., 1999. Toward an understanding of the universality of second-order emotions. In: Hinton, A. (Ed.), Beyond Nature or Nurture: Biocultural Approaches to the Emotions. Cambridge University Press, Cambridge, pp. 75–116.

Fiske, A.P., 1998. Learning Culture the Way Informants Do: Observing, Imitating, and Participating. Mimeo.

Gaechter, S., Falk, A., 2002. Reputation and reciprocity: consequences for the labour relation. Scandinavian Journal of Economics 104, 1–27.

Gergely, G., Bekkering, H., Kiraly, I., 2002. Rational imitation in preverbal infants. Nature 415, 755.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience 2, 861–863.

Gurven, M., in press. Does market exposure affect economic game behavior? The ultimatum game and the public goods game among the Tsimane of Bolivia. In: Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (Eds.), Cooperation, Punishment and Self-Interest: Ethnographic and Experimental Evidence from 15 Small-Scale Societies. Oxford University Press, New York.

Hamilton, W.D., 1975. Innate social aptitudes of man: An approach from evolutionary genetics. In: Fox, R. (Ed.), Biosocial Anthropology. Malaby Press, London, pp. 133–156.

Harbaugh, W.T., Krause, K., 2000. Children's altruism in public goods and dictator experiments. Economic Inquiry 38, 95–109.

Harbaugh, W.T., Krause, K., Liday, S.G., 2002. Bargaining by Children. Mimeo.

Harris, J.R., 1998. The Nurture Assumption: Why Children Turn Out the Way They Do. Touchstone, New York.

Hayashi, N., Ostrom, E., Walker, J., Yamagishi, T., 1999. Reciprocity, Trust, and the Sense of Control. Rationality and Society 11 (1), 27–46.

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. American Anthropologist 103, 992–1013.

Henrich, J., 2003. A cultural species. In: Brown, M. (Ed.), Towards a Scientific Concept of Culture. Mimeo.

Henrich, J., Boyd, R., 1998. The evolution of conformist transmission and the emergence of between-group differences. Evolution and Human Behavior 19, 215–242.

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

Henrich, J., Gil-White, F., 2001. The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior 22, 165–196.

Henrich, J., Boyd, R., 2002. On modeling cultural evolution: why replicators are not necessary for cultural evolution. Journal of Cognition and Culture 2, 87–112.

Henrich, J., McElreath, R., 2003. The evolution of culture and of cultural evolution. Evolutionary Anthropology 12(3), 123–135.

Henrich, J., Smith, N., in press. Comparative experimental evidence from Peru, Chile and the US shows substantial variation among social groups. In: Camerer, C. (Ed.), Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from 15 Small-Scale Societies. Oxford University Press, New York, University of Michigan.

Henrich, J., Bowles, S., Smith, E.A., Young, P., Boyd, R., Karl, S., Richerson, P., Hopfensitz, A., 2003a. The cultural and genetic evolution of human cooperation. In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge, in press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (Eds.), 2003b. Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from 15 Small-Scale Societies. Oxford University Press, New York, in press.

Henrich, J., Boyd, R., Richerson, P., 2003c. Five common mistakes in cultural evolution. In: Sperber, D. (Ed.), Epidemiology of Ideas. Open Court Publishing, in press.

Hill, K., 2002. Altruistic cooperation during foraging by the Ache and the evolved human predispostion to cooperate. Human Nature 13, 105–128.

Hirschleifer, J., 1977. Economics from a biological viewpoint. Journal of Law and Economics XX, 1–52.

Hirschleifer, J., 1978. Natural economy versus political economy. Journal of Social and Biological Structures 1, 319–337.

Kachelmeier, S., Shehata, M., 1997. Internal auditing and voluntary cooperation in firms: a cross-cultural experiment. The Accounting Review 72, 407–431.

Kelly, R.C., 1985. The Nuer Conquest. University of Michigan Press, Ann Arbor.

Kroll, Y., Levy, H., 1992. Further tests of the separation theorem and the capital asset pricing model. American Economic Review 82, 664–670.

Lancy, D.F., 1996. Playing on Mother Ground: Cultural Routines for Children's Development. The Guilford Press, London.

Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. Proceedings: Biological Sciences 268, 745–753.

McCabe, K., Houser, D., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proceedings of the National Academy of Science 98, 11832–11835.

McElreath, R., Boyd, R., Richerson, P., 2003. Shared norms can lead to the evolution of ethnic markers. Current Anthropology 44, 123–129.

Miller, N.E., Dollard, J., 1941. Social Learning and Imitation. Yale University Press, New Haven.

Nettle, D., Dunbar, R., 1997. Social markers and the evolution of reciprocal exchange. Current Anthropology 38, 93–99.

Nisbett, R., Cohen, D., 1996. Culture of Honor. Westview Press, Boulder, CO.

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

Ockenfels, A., Weinmann, J., 1999. Types and patterns—an experimental east-west-German comparison of cooperation and solidarity. Journal of Public Economics 71, 275–287.

Offerman, T., Sonnemans, J., 1998. Learning by experience and learning by imitating others. Journal of Economic Behavior and Organization 34, 559–575.

Paciotti, B., Hadley, C., 2003. The Ultimatum Game in Southwestern Tanzania. Current Anthropology 44(3) 427–432.

Panchanathan, K., Boyd, R., 2003. A Tale of Two Defectors: The Importance of Standing for the Evolution of Indirect Reciprocity Journal of Theoretical Biology, 224, 115–126.

Patton, J.Q., in press. Coalitional effects on reciprocal fairness in the ultimatum game: a case from the Ecuadorian Amazon. In: Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (Eds.), Cooperation, Punishment and Self-Interest: Ethnographic and Experimental Evidence from 15 Small-Scale Societies. Oxford University Press, New York.

Pingle, M., 1995. Imitation versus rationality: an experimental perspective on decision-making. Journal of Socio-Economics 24, 281–315.

Pingle, M., Day, R.H., 1996. Modes of economizing behavior: experimental evidence. Journal of Economic Behavior & Organization 29, 191–209.

Price, G., 1970. Selection and covariance. Nature 227, 520–521.

Price, G., 1972. Extensions of covariance selection mathematics. Annals of Human Genetics 35, 485–490.

Pulliam, H.R., Dunford, C., 1980. Programmed to Learn: An Essay on the Evolution of Culture. Columbia University Press, New York.

Quartz, S.R., 1999. The constructivist brain. Trends in cognitive sciences 3, 48–57.

Quartz, S.R., 2002. Liars, Lovers and Heroes. William Morrow, New York.

Richerson, P., Boyd, R., 1998. The evolution of ultrasociality. In: Eibl-Eibesfeldt, I., Salter, F.K. (Eds.), Indoctrinability, Ideology and Warfare. Berghahn Books, New York, pp. 71–96.

Richerson, P., Boyd, R., 2000. Climate, culture and the evolution of cognition. In: Heyes, C.M. (Ed.), The Evolution of Cognition. MIT Press, Cambridge, pp. 329–345.

Richerson, P., Boyd, R., Henrich, J., 2003. The cultural evolution of cooperation. In: Hammerstein, P. (Ed.), Genetic and Culture Evolution of Cooperation. MIT Press, Cambridge.

Riolo, R.L., Cohen, M.D., Axelrod, R., 2001. Evolution of cooperation without reciprocity. Nature 414, 441–443.

Roberts, G., Sherratt, T., 2002. Does similarity breed cooperation? Nature 418, 499–500.

Smith, N., 2000. Ethnicity, reciprocity, reputation and punishment: an ethnoexperimental study of cooperation among the Chaldeans and Hmong of Detroit, Michigan. Dissertation thesis, University of California, Los Angeles.

Smith, N., 2003. Are Social Preferences Context Dependent? Mimeo.

Smith, E.A., 2003. Human cooperation: perspectives from behavioral ecology. In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge.

Smith, J.M., Bell, P.A., 1994. Conformity as a determinant of behavior in a resource dilemma. Journal of Social Psychology 134, 191–200.

Sober, E., Wilson, D.S., 1998. Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge.

Stark, R., 1997. The Rise of Christianity: How the Obscure, Marginal Jesus Movement Became the Dominant Religious Force in the Western World in a Few Centuries. Harper Collins, New York.

Tarde, G., 1903. The Laws of Imitation. University of Chicago Press, New York.

Tomasello, M., 2000. The Cultural Origins of Human Cognition. Harvard University Press, Cambridge.

Wilson, D.S., 2002. Darwin's Cathedral: Evolution, Religion, and the Nature of Society. University of Chicago, Chicago.

Wit, J., 1999. Social learning in a common interest voting game. Games and Economic Behavior 26, 131–156.

Wynne-Edwards, V.C., 1962. Animal Dispersion in Relation to Social Behaviour. Hafner Publishing Company, New York.

Young, H.P., 1998. Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press, Princeton, NJ.