THE QUARTERLY REVIEW
of BIOLOGY

POWER IN CULTURAL EVOLUTION AND
THE SPREAD OF PROSOCIAL NORMS

NATHAN COFNAS

Balliol College, University of Oxford

Oxford OX1 3BJ United Kingdom

E-MAIL: NATHAN.COFNAS@BALLIOL.OX.AC.UK

ABSTRACT

According to cultural evolutionary theory in the tradition of Boyd and Richerson, cultural evolution is driven by individuals' learning biases, natural selection, and random forces. Learning biases lead people to preferentially acquire cultural variants with certain contents or in certain contexts. Natural selection favors individuals or groups with fitness-promoting variants. Durham (1991) argued that Boyd and Richerson's approach is based on a "radical individualism" that fails to recognize that cultural variants are often "imposed" on people regardless of their individual decisions. Fracchia and Lewontin (2005) raised a similar challenge, suggesting that the success of a variant is often determined by the degree of power *backing it. With power, a ruler can impose beliefs or practices on a whole population by diktat, rendering all of the forces represented in cultural evolutionary models irrelevant. It is argued here, based on work by Boehm (1999, 2012), that, from at least the time of the early Middle Paleolithic, human bands were controlled by powerful* coalitions of the majority *that deliberately guided the development of moral norms to promote the common good. Cultural evolutionary models of the evolution of morality have been based on false premises. However, Durham (1991) and Fracchia and Lewontin's (2005) challenge does not undermine cultural evolutionary modeling in* nonmoral *domains.*

Introduction

ACCORDING to mainstream cultural evolutionary theory, individuals possess "*cultural variants*"—"*idea*[s], *skill*[s], *belief*[s], *attitude*[s], and *value*[s]" (Richerson and Boyd 2005:63) mostly "stored in human brains" (Richerson and Boyd 2005:61). Through interacting with each other, people may adopt and transmit these variants. Cultural evolutionists seek to identify the dispositions that underlie the transmission process, and explain why those dispositions were favored by natural selection under ancestral conditions (Henrich and McElreath 2003; Lewens 2015:17). Having (purportedly) identified how people acquire and transmit culture, mathematical modelers can, given the fitness values of cultural variants and their distribution in a population, predict the future evolution of a culture, or explain retroactively why some variants proliferated and others disappeared.

Fracchia and Lewontin (2005:21–22) raise a fundamental objection to this research program. They note that the success or failure of a cultural variant can depend "crucially on the amount of power behind it" (Fracchia and Lewontin 2005:22). Cultural evolutionary models assume that people more readily acquire beliefs and values with certain intrinsic properties, or preferentially learn from certain kinds of people in certain contexts. The models do not seem to account for the possibility that cultural variants can be imposed on people coercively, regardless of their "learning rules" or "transmission biases" (Richerson and Boyd 2005). Indeed, in a "list of cultural evolutionary forces discussed" in *Not by Genes Alone*, Richerson and Boyd (2005: Table 3.1) say nothing about power or coercion. All of the forces mentioned involve voluntary imitation (see also Henrich and McElreath 2003:Box 3). This potentially leaves out a big factor that operates in real life. Whether "punishment," which Richerson and Boyd do not include on their list of cultural evolutionary forces but which they do discuss in other parts of the book and in other works, or "prestige bias" captures the phenomenon of power as it is exercised in real life will be discussed presently.

Some years before Fracchia and Lewontin (2005), Durham (1991) criticized Boyd and Richerson for what he described as their "radical individualism" that ignores "structured asymmetries or power relations, and coercion" (Durham 1991:182). When Boyd and Richerson do mention power or coercion at all (e.g., Boyd and Richerson 1985: 229–230), they treat it as "individually delivered 'punishment of noncooperators'" (Durham 1991:182).

Mainstream cultural evolutionists have not changed their approach in response to Durham or Fracchia and Lewontin. To illustrate, consider a classic empirical study in the Boyd and Richerson tradition in which Gürerk et al. sought to investigate "institutional selection" (Gürerk et al. 2006:108). In their experiment, *anonymous* subjects played a 30-round public goods game, with each round consisting of three stages: an "institution choice," a "voluntary contribution," and a "sanctioning" stage. In the first stage, subjects "simultaneously and independently" chose to adopt either a "sanctioning institution" (SI) or a "sanction-free institution" (SFI; Gürerk et al. 2006:108). In the second stage, subjects (anonymously) played a public goods game with all of the others who had chosen the same institution, and they were informed about how much each of the other players in *both* groups had contributed to the public good (thus the SI group members knew how well they were doing compared with SFI group members and vice versa). In the third stage, subjects in the SI group were allowed to pay to punish or reward each other.

In the first couple rounds, most subjects chose the SFI. However, freeriding in the SFI group led contributions to fall precipitously. Punishment in the SI group quickly led contributions to rise to the near-maximum level. The vast majority of subjects who initially chose the SFI ended up switching to the SI and adopting a prosocial strategy. Those who participated in the SI received a significantly higher payoff than those who stuck with the SFI. Henrich describes this study as an "experimental demonstration of cultural group selection in action." A key lesson, he says, is that "the players' degree of rationality did

not permit them to foresee the final outcome and select the higher payoff institution on the first interaction" (Henrich 2006:61).

Can we really draw conclusions about "cultural group selection" from Gürerk et al.'s (2006) experiment? Or were there "power relations" and "coercion" that were decisive in the real-life cultural evolution of cooperation and punishment but which were not captured in the experimental setup? It is true that subjects in the SI group were able to "punish" each other, and in that sense they wielded a kind of power. But there are some essential differences between power, punishment, and coercion in real life and that among Gürerk et al.'s (2006) subjects in the SI group. First, subjects were able to choose beforehand whether they wished to be subject to punishment or not (i.e., they chose whether to participate in the SI or the SFI). Henrich (2006) imagines that this is analogous to hunter-gatherers in the ancestral environment choosing which band to migrate to. However, the option to avoid being subjected to punishment by preemptively migrating was probably only an option for our hunter-gatherer ancestors in very limited circumstances (because migration opportunities are limited, hunter-gatherers often receive the threat of expulsion as a death threat. See the section below titled The Reverse Dominance Hierarchy and Deliberate Guidance of Cultural Evolution for a real-life illustration. When hunter-gatherers do have the chance to migrate, it is typically to culturally similar bands). Second, power in the experiment was symmetrical not, as it typically is in real life, asymmetrical (see Singh et al. 2017:461–462)—in the experiment anyone could punish anyone for any reason (that is to say, in real life, those with less power cannot punish individuals—or groups—with more power. And a noncooperative individual who violates group rules cannot go around punishing cooperative group members). Third, and perhaps most important, the experimental setup totally eliminated the possibility of collective decision-making. Again, *anonymous* subjects *"simultaneously and independently"* chose whether to participate in the SI or the SFI. Henrich points out that most subjects failed

to "foresee" that higher payoffs would ultimately be reaped by participants in the SI. He argues that this shows that successful institutions are not generally designed with rational foresight—people just cannot predict the consequences of different institutional arrangements. But if instead of simultaneously and independently choosing their "institution" subjects had been allowed to meet and coordinate a strategy, it is quite likely that most everyone would have been able to "foresee" that the best strategy would be to adopt the SI and to collectively punish free-loaders. The experimental setup *artificially* prevented this from happening because it was predicated on radical individualism that sees cultural evolution as driven by the decisions of uncoordinated, independently acting individuals.

The results of one of Ostrom et al.'s (1992: 412–413) experiments seem to support the assumption that, if people are allowed to communicate and make a collective decision, they *can* anticipate the benefits of punishment. Groups of eight subjects played a common pool resource game for 20 rounds. After the 10th round, subjects met face-to-face for 10 minutes and were allowed to decide (by a majority vote) whether to institute a "sanctioning mechanism" with a fee-to-fine ratio of 1/2 for their future interactions. Out of four groups in this experimental condition, two instituted a sanctioning mechanism (with fee-to-fine ratios of 10¢/20¢ and 20¢/40¢, respectively). However, Ostrom et al. (1992) note that the subjects in this experiment had previously participated in another experiment where they played a common pool resource game *with a sanctioning mechanism but no chance for communication*. Out of the 14 subjects who voted against the sanctioning mechanism, 11 had previously participated in a design with a relatively high fee-to-fine ratio of 20¢/80¢. Out of the 18 who voted in favor of the mechanism, just *three* had been in the 20¢/80¢ design. They "infer from this result that the high level of sanctioning activity in the 20¢/80¢ design, the lack of overall efficiency gains and the presence of blind revenge combined to impede the willingness of participants to choose a sanctioning mech-

anism" (Ostrom et al. 1992:413). They also suggest that "the experience of the first 10 rounds of the . . . game had an effect on mechanism choice" (Ostrom et al. 1992:413). To test this possibility, they conducted two more experiments where subjects could communicate and adopt a sanctioning mechanism *before playing any rounds*. The result: "In both of these experiments, the subjects quickly agreed to an investment strategy and a sanctioning mechanism to punish defectors. Across the two experiments, net yields averaged 95%–94% with fees and fines included" (Ostrom et al. 1992:413).

What Is "Power"?

In order to properly articulate Durham and Fracchia and Lewontin's (DFL) challenge, we should have a clear idea of what power is. Fracchia and Lewontin (2005; Lewontin 2005) refer simply to "power," which they illustrate with examples of rulers imposing cultural variants by decree. They do not explicitly describe the mechanisms that make such decrees effective. Durham usually refers just to "power," and at one point to "structured asymmetries or power relations, and coercion" (Durham 1991:182)—phraseology that seems to imply that "power relations" and "coercion" are different things. "Power" is clearly a multifarious concept. Which type(s) of power might pose a challenge to traditional cultural evolutionary modeling?

In his influential analysis, Dahl (1957) suggests that the "intuitive idea of power . . . is something like" the following: "*A* has power over *B* to the extent that he can get *B* to do something that *B* would not otherwise do" (Dahl 1957:202–203; cf. Lukes 2005:30). Whether or not this definition describes our intuitive idea of power, it is clearly too broad for the present purposes. On Dahl's definition, virtually all cultural transmission would have to be conceived as a manifestation of power. Any time *B* copies *A*, *A* exercises power over *B*, even if *B* is guided by one of the learning biases represented in standard cultural evolutionary models.

Bachrach and Baratz (1970) offer a useful taxonomy of power (see discussion in Lukes 2005:21–22), which may allow us to pick out those forms that present a challenge for cultural evolutionary modeling. Their taxonomy includes influence, authority, force, and manipulation. *A* influences *B* when *A*, "without resorting to either a tacit or an overt threat of severe deprivations, causes [*B*] . . . to change his course of action" (Bachrach and Baratz 1970:30). *A* exercises authority when *B* "complies because he recognizes that the command is reasonable in terms of his own values" (Bachrach and Baratz 1970:34). *A* uses force when *A* obtains compliance by depriving *B* of choice. *A* manipulates *B* when *B*'s "compliance is forthcoming in the absence of recognition on the complier's part either of the source or the exact nature of the demand upon him" (Bachrach and Baratz 1970:28).

According to Bachrach and Baratz's (1970) jargon, "power" should refer only to cases where *B* complies with *A* because *B* fears that *A* "will deprive him of a value or values which he regards more highly than those which would have been achieved by noncompliance" (Bachrach and Baratz 1970:24). "Force," they say, should refer only to cases of physical manipulation. However, following Lukes (2005)—and common usage—we can regard influence, authority, force, and manipulation as species of power. When *A* deprives *B* of the option of following a course of action, this is an exercise of force.

Influence, as Bachrach and Baratz define it, is a form of power that can, in general, be easily accommodated by cultural evolutionary models. When people freely copy each other, this is a way of being influenced. Standard cultural evolutionary models simply reflect the rules, or the *learning biases*, that determine how people freely copy each other.

Force is the form of power that poses the most obvious (apparent) challenge to the models. If cultural variants are imposed on people because the option to reject them is taken away by force, then it seems that standard cultural evolutionary models cannot explain what happened.

Authority is established when individuals freely recognize the right of a leader to make certain demands on them. If cultural variants spread on a large scale because of the exercise of authority, this would seem to under-

mine cultural evolutionary modeling, at least to some extent. If people accept the authority of a leader to tell them what cultural variants to adopt, to the extent that the leader exercises their authority, the learning biases of cultural evolutionary models will cease to play a decisive role in (further) cultural evolution. Learning biases might explain why people choose to accept the authority of a particular leader in the first place, but they would not determine which variants subsequently spread. When religious leaders promulgate opinions about values or doctrine, their followers may adopt those opinions because they accept the authority of the leader. To reiterate, cultural evolutionary models might help explain how authority itself comes to be *established* (e.g., how people become Catholic and thereby accept the Pope's authority). But, just as when cultural variants are spread through force, when they are spread through the exercise of authority the learning biases of cultural evolutionary models play no role. When Durham (1991) distinguished between "power relations" and "coercion," perhaps he had in mind the distinction between "authority" and "force."

Authority and force can interact in interesting ways. For example, one reason that we accept authority might be that it is backed by force (would Prime Minister Theresa May retain her authority if the military and the police disbanded? Probably not for long). Conversely, the power to employ force is often made possible by widespread acceptance of the legitimacy of the authority (people tolerate Theresa May giving orders to the military and the police because they accept her right to be Prime Minister). Sorting out how this works in detail is not important for the present discussion. The point is that when cultural variants are spread through the exercise of either force or authority, the learning biases featuring in cultural evolutionary models are *to that extent* irrelevant.

Manipulation—surreptitiously generating desires in people to make certain choices—generally works by taking advantage of learning biases. For example, advertisers use celebrity endorsements to take advantage of prestige bias. The Sara Lee Corporation paid former basketball star Michael Jordan a great

deal of money to endorse Hanes underwear because many people are motivated to copy successful athletes. Richerson and Boyd (2005: 124) note that some men also started shaving their heads because this was Michael Jordan's practice. They cite imitating Jordan's choice of underwear and imitating his hairstyle as illustrations of "prestige bias." But these two cases of imitation illustrate different kinds of *power*. When men shave their heads because of Jordan, they are being *influenced*. When they buy Hanes because of him, they are (most likely) being *manipulated*.

In principle, cultural evolution driven by manipulation can be modeled in the same way as that driven by *influence*. Of course, if manipulation is a decisive factor, an explanation of cultural evolution that leaves this fact out and appeals only to the learning biases of the people who were manipulated would be impoverished. But there is no reason why standard cultural evolutionary models cannot be used to track how cultural variants spread in such circumstances. *Prestige*, however, has a superficial resemblance to *authority*. The following section considers whether the latter can be reduced to the former.

The fact that which cultural variants people adopt may be determined by their position in networks of power—specifically networks of force and authority—cannot be captured, at least in any obvious way, by models that are predicated on a strong form of methodological individualism. Models that are committed to a highly individualistic picture of decision-making and action will also miss the fact that, when it comes to force and authority, the *locus of power* is very often a *group* of collectively acting individuals.

Can Power Be Reduced to "Coordinated Punishment" or "Prestige Bias"?

Coordinated punishment and prestige bias are two phenomena that already feature in mainstream cultural evolutionary models. If either one captures the phenomenon of power—namely, the exercise of *force* or *authority*—in a sufficiently realistic way, then cultural evolutionary modeling would not be threatened by DFL's challenge. This section

considers coordinated punishment and prestige bias in turn.

Boyd et al. acknowledge that previous models of the evolution of punishment have been unrealistic in their assumption that "punishment is an unconditional and uncoordinated individual action automatically triggered by defection" (Boyd et al. 2010:617). To rectify this shortcoming, they propose a model (also defended in Bowles and Gintis 2011:Chapter 9) of what they call "coordinated punishment of defectors." This sounds like it may be a step toward incorporating collective social phenomena (including collectively exercised *force*) into cultural evolutionary models. However, closer inspection suggests that the kind of coordination captured by Boyd et al.'s (2010) model is still unrealistically individualistic. In their model, "coordinated punishment" means that, when individuals decide whether to punish defectors, they take into account whether other individuals have signaled *their* intention to punish defectors. But, as shall be argued, the decision to punish—as it is represented in this model—is made without the sort of collective plan that supports cooperative behavior in real life, and that, according to anthropological evidence, played an essential role in the evolution of prosocial norms.

Boyd et al. say that, in their model, "punishment is coordinated among group members so that it is contingent on the number of others predisposed to participate in the punishment" (Boyd et al. 2010:617). But there is much more to the coordinated punishment described by anthropologists, or even by Boyd et al. in their verbal description of anthropological observations. As Boyd et al. themselves say:

> ethnographic evidence indicates that punishment is coordinated by means of gossip and other communication among punishers, is contingent on the expected effectiveness of punishment in inducing cooperation, and it is not undertaken unless it is judged as legitimate by most group members (Boyd et al. 2010:617).

The collective agreement and action that emerges from this kind of cooperation can not be reduced to punishing in response to signs that others will punish.

In Boyd et al.'s (2010) model, individuals are drawn from a population of "punishers" and "nonpunishers" to form groups. Members of each group engage in an initial three-stage interaction. In the first stage, punishers exhibit a "signal." In the second stage, if a threshold number of other individuals in the group gave the signal, punishers cooperate (with probability $1 - e$)—otherwise they defect. Nonpunishers always defect at this stage. In the third stage, if the threshold number of punishers revealed themselves in the first stage, the punishers impose a cost on those who did not cooperate. Group members then engage in further interactions consisting only of cooperation/defection and punishment stages.

In administering a punishment, each punisher incurs a cost. The more punishers there are in a group, the less the *total* cost of inflicting punishment on a defector, since the defector's ability to retaliate at all can be assumed to diminish exponentially with increasing numbers of punishers.

Because punishment imposes a cost on defectors that is, on average, greater than the cost of cooperating, in general, cooperation maximizes expected payoff if defectors face punishment. In the model, defectors who are punished in the first round subsequently behave like punishers, cooperating with probability $1 - e$. They defect with probability e either because of error or because, due to random differences in circumstances, cooperation may be costlier than being punished.

In the model, after engaging in a (randomly varying) number of interactions, individuals produce offspring according to the payoff they received, and new groups are formed from a mix of the offspring. Offspring inherit the strategy of "punishing" or "nonpunishing" from their parent.

Boyd et al.'s basic conclusion is that "the initial proliferation of punishment occurs under plausible levels of group genetic differences and results in persistent and high levels of cooperation" (Boyd et al. 2010:620). Punishers are able to proliferate in their model because—unlike in other models—they refrain from inflicting costly punishment un-

less there are enough other punishers in their group to reduce the cost, and they are able to do this because punishers "coordinate" by signaling their intentions. This paper will argue that prosocial-norm-enforcing punishment did not evolve in this way, with individual contingent, signaling punishers gradually infiltrating groups of nonpunishers in the ancestral environment. Rather, individuals within groups of nonpunishers formed coalitions to collectively impose prosocial norms on all group members. And, through a process of gene-culture coevolution, we became genetically adapted to be receptive to such norms, including norms to punish noncooperators. The evidence will show that Boyd et al.'s (2010) model gives an incorrect historical account of how punishment arose in human evolution, and of how we evolved a disposition to punish norm violators.

The discussion above suggests that power cannot be reduced to coordinated punishment. Can it be reduced to "prestige bias"? Prestige bias refers to our (well-established) tendency to preferentially copy individuals who are successful, or to whom *others* are paying attention or deferring (see, e.g., Chudek et al. 2012). This is, of course, one of the central learning biases taken into account by cultural evolutionary models.

It is true that prestigious and prominent people have (ceteris paribus) more influence over culture than the less prestigious, and this influence is a form of "power." When celebrity Kylie Jenner tweeted that she did not use the app Snapchat anymore, the company's stock lost 1.3 billion dollars in value. That is because when Jenner says she uses or does not use something, many people will be inclined to imitate her.

But the "power" of Kylie Jenner is fundamentally different from the power of a king, a hunter-gatherer coalition enforcing rules, or even the Pope. Jenner's influence derives from the fact that many people preferentially imitate her. When she exhibits behavior, or expresses an opinion, we can explain why it spreads by appealing to the learning biases (namely, the prestige bias) of her audience. The kind of power that threatens cultural evolutionary modeling is that which is backed by *force* or *authority*, and which thereby neutral-

izes all of the learning biases featuring in mathematical models. Neither force nor authority can be reduced to prestige bias, since force and authority neutralize all learning biases. When, in 1492, Isabella I and Ferdinand II issued the Alhambra Decree ordering the Jews of Spain to convert to Catholicism, leave, or be executed (as discussed in more detail below), it is theoretically possible that a small number of Jews were impressed by the religious passion of the prestigious queen and king and converted for that reason. But most of the Jews who converted probably did so because their freedom to choose had been coercively restricted. Living in Spain as a practicing Jew was literally no longer a *possible choice*. In hunter-gatherer groups, individuals did not have the *choice* to refrain from sharing meat or contributing to group defense—if they did not do these things, and were recalcitrant enough in the face of reproach, they would be subject to execution or (what was often effectively the same thing) expulsion.

APPROACHES TO ACCOMMODATING POWER

Lewontin (2005) illustrates the role of power in cultural evolution with an example. Bavarians are mostly Catholic, Westphalians Protestant. This state of affairs was not brought about (according to Lewontin) by any mechanism recognized by cultural evolutionists—not by biased transmission, the relative appeal of the content of the two religions, drift, Darwinian selection, or any other (see the list in Richerson and Boyd 2005: Table 3.1). Rather, Lewontin says, in 1555 the German princes and the Holy Roman Emperor adopted "the rule of *cuius regio, eius religio*, which allowed rulers to enforce their own religion in their own dominions and to expel those who were recalcitrant" (Lewontin 2005). If cultural change is usually driven in this way—by sweeping diktats issued by powerful individuals or small groups—then Boyd and Richerson-type modeling, which tracks "the aggregated effects of small-scale events" (Lewens 2015:17), will have a very limited range of application. Most of history—and the Paleolithic culture to which our minds are adapted—would better be explained by

a study of how individuals take power and impose their favored cultural variants on the masses, and cultural evolutionary models would not shed much light on the selective forces that shaped our social learning dispositions.

As Lewens notes, it would surely be a huge overestimation of the power of the typical ruler to suppose that they can change the religious beliefs of a population by mere decree. In regard to the specific example of *cuius regio, eius religio*, it is not clear that the German princes even tried to change people's religious beliefs on a large scale. They tended to "enforce" Catholicism or Lutheranism on their subjects according to what was already the prevailing religion, and even subjects in the minority did not necessarily convert in response to the official order (Lewens 2015:133).

But that is just one example. It seems clear that sometimes cultural variants, even religious beliefs, can spread as a result of being backed by power. As discussed, the forms of power that seem to undermine cultural evolutionary modeling are *force* and *authority*. Sometimes power makes all of the difference in which cultural variants prevail in a group. The Alhambra Decree, which expelled practicing Jews from Spain in 1492, caused more than two-thirds of the not already-converted Jewish population to become Catholic, at least nominally. The decree, initiated by one person (Isabella I) and signed by two (the aforementioned and Ferdinand II) radically changed the distribution (and attractiveness) of certain cultural variants in Spain, and its consequences reverberated for centuries. In the 4th century, Christianity surely got a boost from being supported by Constantine, although whether this was responsible for its ultimate triumph is difficult to say.

Lewens (2015:138–139) proposes three ways in which cultural evolutionary research programs could accommodate the phenomenon of power. First, restriction: we could simply restrict the application of models to times and places where cultural evolution is not driven by Lewontin scenarios—"where human societies have been free of interest groups, governments, unions, and perhaps gross asymmetries of individual power" (Lewens 2015:138). Second, presumption: we

could take systems of power as "fixed background conditions against which individuals interact" (Lewens 2015:138), using cultural evolutionary theory to track cultural change resulting from those interactions. And, finally, reduction: we could try to reduce systems of power—"governmental structures, judicial procedures, and so forth" (Lewens 2015:139)—to small-scale interactions dealt with by standard cultural evolutionary theory.

By restricting cultural evolutionary modeling to times and places "free of interest groups, governments, unions, and . . . gross asymmetries of individual power" (Lewens 2015:138), *restriction* undermines any modeling applied to modern societies. Lewens suggests that we could still use cultural evolutionary models to investigate early hominin history, since "hunter-gatherers were largely egalitarian with respect to individuals' access to economic resources" (Lewens 2015:138, citing Boehm 1999 and Knauft 1991). But even this is not certain. Despite their lack of formal leaders, great differences in individual power among adult males, or great differences in individual access to economic resources, hunter-gatherers do have systems of rules, coercion, and punishment. In fact, as shall be argued below, the egalitarianism of hunter-gatherers is itself maintained by the exercise of power wielded *by* the rank and file over their potential dominators (or alphas), creating what Boehm (1993, 1997, 1999) calls a "reverse dominance hierarchy." This means that even hunter-gatherers possess a power apparatus that could theoretically be used to impose cultural variants on individuals willy-nilly and force deviants into line. Whether hunter-gatherers actually do use this power to influence the spread of cultural variants in a way that undermines Boyd and Richerson-type modeling is the central question addressed in the present paper.

The problem with *presumption*—taking power relations as "fixed background conditions against which individuals interact" (Lewens 2015:138)—is that it seems to ignore DFL's challenge. Suppose we want to explain the spread of different sects of Christianity in mid-16th-century Germany. An explanation emphasizing the forces listed in Richerson and Boyd (2005:Table 3.1) and relegating power relations to "fixed background condi-

tions" would clearly be impoverished and misleading, in that it would be taking for granted the factor that ought to be a primary target of a scientific explanation (even if, as Lewens argues, it was not the whole story). How could we construct an explanation of the spread of Catholicism and Lutheranism taking the power of the princes as "fixed background conditions"? Perhaps we could treat the positive and negative incentives offered for adopting different beliefs in different places as being features of the environment like the weather. The fact that people prefer, ceteris paribus, denominations that do not cause them to be exiled from their home could be considered a "content bias." Although it may be possible to model the decision-making process of German peasants along these lines, this approach would ignore the fact that a major driving force of cultural evolution is the asymmetrical power relations of the princes and the peasants. Fracchia and Lewontin's (2005) point is that explaining historical trends—trends about how cultural variants spread—requires us to recognize power as a major factor. In Durham's words, "to ignore [power] . . . is to ignore what may often be the leading cause of transformation" (Durham 1991:182). The fact that we can treat the influence of power as a given and focus on individual decision-making does nothing to counter the claim that power often plays a decisive role in the fate of cultural variants. Lewens raises a similar problem for presumption, saying that as a strategy it "simply omits to explain many features that are admitted as important in determining cultural change" (Lewens 2015:138), although he says presumption can be useful for "understanding . . . individual interactions and their long-term effects" (Lewens 2015:138).

This leaves *reduction*—reducing the influence of power to individual-level interactions that can be handled by standard cultural evolutionary models. In Lewens's view, the justification for this approach is that power relations are all grounded in individual interactions, which should in principle be amenable to populational modeling. This does seem to be the approach advocated by cultural evolutionary modelers themselves (e.g., Boyd and Richerson 2002; Richerson and Henrich 2012), although in practice their

models do not generally reflect power relations at all.

But redescribing the exercise of power in terms of individual interactions does not negate the fact that power is a factor, and it does not neutralize DFL's challenge. Consider again the Alhambra Decree. Isabella and Ferdinand issued an order (Jews had to convert, leave, or be executed). Since those with the ability to enforce the order—the soldiers—accepted the authority of the monarchy, they conveyed the threat to the Jews, who chose whether to convert or leave (no one volunteered for execution). If we want to explain why Spanish soldiers accepted the monarchy, we would probably need to refer to all of the learning biases that figure in cultural evolutionary models. If we want to explain how Jews responded to the threat, we could say that the cost of being overtly committed to Judaism had increased, and since people have a content bias to acquire cultural variants with lower costs, Judaism became less attractive. So rather than saying that the Jews converted in response to the exercise of power by Isabella and Ferdinand, we can say that they decided which practices to adopt in light of the incentives. But we still cannot explain what happened without referring to the fact that, because the social structure gave power to the monarchs, two individuals drove cultural evolution in a specific direction. To explain why the powerful individuals acted as they did, we would need to engage in traditional historical, sociological, or psychological analysis.

CAN CULTURAL EVOLUTIONARY MODELING BE RESTRICTED TO THE ANCESTRAL ENVIRONMENT?

Modern societies are founded on extensive, complex systems of power. The information we are able to obtain, the opinions we are exposed to, and the options that are presented to us are constrained in all sorts of ways by the (often hidden) exercise of force and authority. DFL's challenge to cultural evolutionary modeling appears to have great force when it comes to understanding how cultural variants spread in extremely hierarchical societies where powerful individuals and groups exercise such far-reaching con-

trol. This is not to deny that the models can be used to explain some aspects of cultural evolution even in modern societies. When people's choices in a particular domain are largely unconstrained by power, the models can sometimes explain how and why culture develops as it does. For example, cultural evolutionary models can explain the demographic transition in the West (Richerson and Boyd 2005:Chapter 5), since people in Western countries in recent history have been largely free to choose how many children to have. Power is rarely the *only* factor influencing what people believe and how they act, so models might be used in conjunction with traditional historical/anthropological/sociological methods. But to explain phenomena such as the spread of communism and the concomitant values in the 20th century, a study of the influence of *power* will often be more fruitful than a study of the aggregate effect of the decisions of many individuals.

Cataloging all of the ways in which power can or does influence the spread of cultural variants, and in which it might undermine cultural evolutionary modeling, is not the aim of the present paper. This paper has a narrower focus, which is to consider whether or to what extent power undermines cultural evolutionary modeling applied to nomadic, foraging societies. It seeks to test the possibility raised by Lewens (2015) that we can *restrict* the application of Boyd and Richerson-style models to conditions that prevailed during the long period before the advent of sedentarism, food storage, and agriculture, when human society was reputedly egalitarian.

There are two reasons why it is worthwhile to investigate whether Boyd and Richerson-style models can be applied unproblematically to nomadic foragers. First, the models have been used to give accounts of the historical origins of specific cultural variants, such as certain ethical norms. If the models cannot be applied to nomadic foraging societies (where the variants in question originated), then we will have to reconsider these accounts. Second, the models—predicated on the idea that cultural variants spread in the ancestral environment due largely to content and context learning biases—have been used to make predictions about what psychological adaptations we might have. Given that variants

spread due to content and context learning biases, cultural evolutionary theorists ask what psychological dispositions we might have evolved in order to extract adaptive information from the cultural environment. But if *power* was a major factor influencing how variants spread in the ancestral environment, then we might make different predictions about how our psychology evolved. We might predict, for example, that we evolved to have different attitudes toward variants imposed on us by power and those that we adopted due to learning biases.

Chimpanzees (*Pan troglodytes*) have a strong dominance hierarchy based on physical force. The alpha individual rules because he (sometimes with the help of his coalition partner) can beat up every other chimp. Beta can beat up gamma, and so on down the line. Under natural conditions, all adult males, being physically stronger, are dominant to all females (Goodall 1986).

Dominant chimps use their social position primarily to secure access to food and mating opportunities. Dominance relations have little direct influence on cultural transmission. Chimps have different, culturally transmitted ways of fishing for termites, building nests, and performing a few other tasks. Which cultural variant an individual acquires depends on the practice of the most accessible models and features of the local environment (McGrew 2004:Chapter 7). Under experimental conditions, chimps preferentially attend to dominant individuals when learning how to solve foraging problems, and are more likely to copy their methods (Horner et al. 2010). But dominant individuals have no interest in—and may even lack awareness of—the cultural practices of their subordinates, hence they make no attempt to forcibly spread their favored practices. The common ancestor of chimps and humans probably had a similar hierarchical social structure (Boehm 1999).

At some point, hominins developed much greater awareness of cultural possibilities, and a much greater capacity for deliberately transmitting learned practices. It became feasible for dominant individuals to force subordinates to adopt their preferred cultural variants. Richerson and Henrich note that "[p]redatory elites and other self-interested subgroups with some form of coercive power" could

establish norms and institutions that "disproportionately benefit them. Ideologically motivated groups with coercive power may sustain equilibria at mad extremes, at least for brief periods of time" (Richerson and Henrich 2012:50; they do not comment on the implications this might have for cultural evolutionary modeling). But elites who impose cultural variants on others do not necessarily have to be motivated by narrow self-interest—they could be prompted by altruism, concern for the commonweal, or the belief that a well-functioning group will ultimately benefit themselves (i.e., enlightened self-interest). In any case, if dominant individuals or coalitions in the Paleolithic did exercise power to impose cultural variants on others, it could create a problem for cultural evolutionary modeling for the reasons discussed above.

There are two crucial empirical questions to answer. First, was there any locus of power in Paleolithic communities *capable* of imposing cultural variants on subordinates? Second, if the answer to the first question is yes, was this power routinely exercised in a way that supports DFL's critique of cultural evolutionary modeling? These questions are dealt with in turn in the following two sections. Based on work by Boehm, the first will be answered in the affirmative. The second will also be answered in the affirmative, but with a qualification. It will be argued that, in the ancestral environment, power was used primarily as a means of enforcing moral norms. Therefore, DFL's critique undermines standard cultural evolutionary modeling applied to moral norms that arose in the past 250,000 years, but it does not necessarily undermine modeling applied to nonmoral norms in the ancestral environment. It also does not necessarily undermine modeling applied to the evolution of moral norms and intuitions that might have occurred before 250,000 BP (the specific example of Baumard et al.'s 2013 model will be discussed).

THE EXISTENCE OF POWER IN THE PALEOLITHIC

Knauft notes that "male status differentiation in human evolution is U-shaped" (Knauft 1991:397). The social organization of the chimp-human common ancestor is presumed to have been as hierarchical as that of modern chimps. After hunter-gatherers settled down and developed the means to stockpile food, human societies became even more hierarchical than those of chimps. For some millions of years after the chimp-human lineages split and before sedentary, hierarchical societies were established, humans lived as nomadic hunter-gatherers. All nomadic hunter-gatherers that have ever been described, without exception, live by an ideology of extreme egalitarianism (among men). No nomadic hunter-gatherer society tolerates adult males issuing direct orders to other adult males. Some groups have a titular leader, but his role never goes beyond leading by example and helping to organize group discussions aimed at reaching a consensus on group actions (whether to go to war, where to migrate, and the like; Boehm 1999).

As noted earlier, Boehm (1993, 1997, 1999) argues that nomadic foraging societies are characterized by a "reverse dominance hierarchy": men who would otherwise be subordinate to an alpha band together to keep would-be alphas from assuming positions of dominance. In contrast, Erdal and Whiten (1994:177) argue that there is an *absence*, not a *reversal*, of hierarchy. According to them, would-be subordinates simply refuse to obey would-be leaders. A hunter-gatherer group does not (in their view) "start with a hierarchy and [then] reverse it" (Erdal and Whitten 1994:177). Rather, "counterdominant" behavior prevents leaders from arising in the first place.

The debate between Boehm and Erdal and Whiten is not just scholastic. If Erdal and Whiten are right that forager egalitarianism is based on a widespread refusal to obey orders, then there would have been essentially no locus of power among Paleolithic adult male foragers that could have potentially directed cultural evolution. If no one ever obeys authority, no individual or coalition can exercise power over anyone else. But if Boehm is right that egalitarianism is maintained by the asserted efforts of the rank and file to keep down ambitious, aggressive, and accomplished individuals, Paleolithic bands would have had groups—coalitions of the majority—with the power to potentially impose cultural variants on those whom they dominated.

There are compelling reasons to side with Boehm in this debate. Perhaps most significant is the fact that nomadic hunter-gatherers deliberately and actively *enforce* egalitarianism, employing a variety of *coercive* methods ranging from gossip to ridicule to ostracism to expulsion (often a de facto death sentence) to outright execution. They do not simply refuse to obey commands and rebuff would-be dominators. They espouse an explicit egalitarian ideology (Cashdan 1980) and take active measures to head off possible power grabs by individuals. As Boehm (1999) argues, our primate heritage makes us readily susceptible to developing orthodox hierarchies (i.e., with alpha-types dominating everyone else). To prevent orthodox hierarchies from emerging requires the group to continually keep down those who would take control if given the opportunity.

What is more, the group actively controls the behavior of potential alphas, compelling them to procure meat for others and allowing them to act as *informal* leaders in hunting and warfare (so that the group benefits from their expertise without submitting to their domination; Boehm 1997:S104). Again, this goes beyond refusal to obey orders and reflects bona fide domination and control of potential alphas.

Everyone, not only potentially dominant individuals, is subject to group rules. People who are not suspected of political ambitions but who horde food or engage in deception can be targeted by some of the same sanctions as those who try to gain *individual* power. Boehm (1999) quotes Service's (1975:48–49) observation that in foraging societies social life is intensely regulated by "codes, rules, expectations, habits, and customs that are related to etiquette, ethic, and role. And because these are not [normally] explicit, nor revealed by frequent breaches, the society might give the impression of freedom and lack of conflict" (Boehm 1999:84). But the threat of collectively imposed sanctions of varying degrees of severity hangs over everyone's heads.

The foregoing suggests that, in the sort of hunter-gatherer bands that existed since at least the early Middle Paleolithic and to which our social psychology is presumed to be adapted, there did exist a locus of power that was *theoretically* capable of driving cultural evolution in the way DFL suggested (i.e., by fiat issued by a consensus of groups). Whether this threatens cultural evolutionary modeling turns on the answer to an empirical question: Did hunter-gatherer coalitions of the majority, as a matter of *fact*, exercise the power described above to promote their favored cultural variants, or did they use it only to suppress power grabs by would-be alphas? It will be argued that the empirical evidence supplied by Boehm (1999, 2012) strongly suggests that hunter-gatherers *did* use their collective power to impose cultural variants that they thought would promote group success and well-being—variants broadly related to morality (note that a great variety of behavior can be moralized, from food preferences to factual beliefs). They did *not* use power to impose variants related to practices that were not thought to directly concern the group, like individuals' specific spear-making techniques. Consequently, cultural evolutionary models can be applied more or less unproblematically to the *nonmoral* domain of culture. It is cultural evolutionary modeling of the transmission of group-concerning moral practices such as reciprocity and punishment (e.g., Henrich and Boyd 2001) that may be problematic.

THE REVERSE DOMINANCE HIERARCHY AND DELIBERATE GUIDANCE OF CULTURAL EVOLUTION

This section briefly outlines Boehm's (1999, 2012) account of how subordinate humans circa 250,000 BP overthrew their alphas and, around the same time, used their newly acquired collective power to impose a deliberately engineered moral code. In the light of Boehm's theory, the following section will critically analyze some well-known modeling work on the coevolution of prosocial norms and norm psychology.

From our primate ancestors we inherited three key dispositions: an enjoyment of dominating, a capacity for submission, and a dislike of being dominated. Among chimpanzees, these same dispositions lead to the development of a dominance hierarchy. Subor-

dinates—particularly males—wait for the opportunity to challenge higher-ups and ascend the hierarchy, if possible. Subordinate chimpanzees (*P. troglodytes*) often form coalitions to challenge the power of alphas (Boehm 2012:95–96) but, because of limits on their coordination abilities and the physical supremacy of the dominant individuals, they never succeed in doing away with the practice of alpha rule itself.

Chimps, bonobos, and gorillas have the ability to make generalizations about what sort of positive or negative behavior will elicit the wrath of their superior, and purposefully comply with the superior's demands (Boehm 2012:106; the behavior demanded by dominant apes mostly concerns feeding priority and mating; Boehm 2012:107–108). This was an important preadaptation for our ancestors to develop the ability to appreciate group-imposed *rules*. Taking advantage of our ability to understand and adopt regular patterns of behavior, coalitions of the majority in human groups instituted rules to enforce egalitarianism and promote the good of the collective. For hunter-gatherers, the coalition of the majority is the object of fear, rather than, as for chimps or gorillas, powerful dominant individuals. Boehm expounds: "People like Bushmen or Pygmies gossip incessantly and are highly judgmental, and group opinion is something to be feared because moral outrage can lead to ostracism, expulsion from the group, or even execution. This is true of all hunter-gatherers" (Boehm 2012:107).

Darwin speculated that, when language developed, "the wishes of the members of the same community could be distinctly expressed, the common opinion how each member ought to act for the public good, would naturally become to a large extent the guide to action" (Darwin 1871:72). The "power" of public opinion derived, he said, not from coercion but from our natural love of approbation and "horror of scorn and infamy," which itself is rooted in our "[i]nstinctive sympathy" inherited from our distant ancestors (Darwin 1871:86; see Lewens 2007:162–167 for an overview of Darwin's account; Darwin 1871 does not explain precisely how "[i]instinctual sympathy" gives rise to a desire for moral approval). Although Darwin was right about the new possibilities that language created for social organization, the ethnographic evidence discussed below suggests that hunter-gatherers employ coercion to enforce behavior approved by public opinion. There is good reason to think that coercive methods were employed by ancient humans as well. We do value the moral approval our fellows, and this is also an important source of motivation to conform to group rules (Mameli 2013: 911). But our desire for moral approval is most likely an evolved response to the widespread social practice of collectively punishing norm violators, and is not simply an outgrowth of "sympathy."

Language allows members of coalitions in a hunter-gatherer band to collectively agree on what Boehm terms a "blueprint" upon which to model their society. Echoing Darwin, he says that because "humans are able to communicate in great detail, . . . groups can develop precise notions about the kind of society in which they wish to live" (Boehm 1999:193). Whether, as a matter of historical fact, humans began collectively suppressing alphas *and then* started enforcing other moral behavior or vice versa (see Boehm 1999: 194), ultimately hunter-gatherers came to implement blueprints of egalitarian, moral societies. Boehm (2012:Table 1) reports the number of cases recorded by ethnologists where capital punishment was meted out for various offenses in 50 mobile foraging bands. There are records of capital punishment in 24 of the 50 bands, for a total of 45 instances. Boehm notes that this is likely a significant underestimate of the prevalence of capital punishment, because foragers have learned to conceal such practices from outside authorities (the count here also does not include expulsion, which, as noted, is often a de facto death sentence). The most common crimes eliciting a lethal response were various forms of intimidating the group with aggression, violence, or "malicious sorcery." But people were also executed for such moral violations as theft, failing to share meat, incest, adultery, and failing to respect taboos. To reiterate, execution is just the most extreme form of punishment. The vast majority of norm violators in hunter-gatherer bands reform in response to less serious sanctions

such as ridicule. People use the collective power of the group to make society conform to a desired blueprint.

Humans were hunting with wooden weapons at least 400,000 BP, and were taking down big game as a regular source of food 250,000 BP (Boehm 2012:146). Boehm speculates that the key impetus for reversing the dominance hierarchy came when meat from large animals became a crucial part of our diet. Hunting bands can only be effective if all hunters are reasonably well nourished and motivated, and this requires that the meat be shared rather than hoarded by an alpha and his coalition partners (Boehm 2012:151–152) or taken by "greedy thieves and cheaters" (Boehm 2012:155).

When coalitions of the majority started to collectively enforce group rules, this created a strong selection pressure for the development of a *conscience,* and concomitant emotions such as shame and moral pride. To have a conscience is to "*personally [identify] with community values,* which means internalizing your group's rules" so that you are emotionally invested in following them (Boehm 2012:113). The human capacity for language allows for an extreme degree of effective surveillance. No matter who witnesses a rule violation, word can spread throughout the whole group. People who "failed to control their predatory tendencies" (Boehm 2012:67)—who bullied or cheated or otherwise violated group norms—would have been at a great disadvantage. Those who were able to internalize norms as goals in themselves would have avoided trouble much better than those who, like chimps, observed the rules only when they thought an authority was watching (Mameli 2013). Given the prospect of group punishment, developing a good reputation became an important way to increase individual fitness.

Consider an illustrative example of group enforcement of norms, reported by Turnbull (1961:94–108) and cited by Boehm (2012: 37–43). The Mbuti have a hunting practice where the men set up long nets in the shape of a semicircle while the women and children make noise to drive animals (small- and medium-sized game) toward them. According to their rules, a man can kill any animal that falls into his own net and take the meat for his own family. Although there may be chance variation in how well a man fares on a given hunting expedition relative to other men, on average this is a fair way of distributing the spoils.

In the course of one such hunt, a man named Cephu felt he was getting short-changed, so he surreptitiously moved his net ahead of the others. The strategy worked as far as increasing his catch but, unfortunately for him, he was spotted by another hunter. Word quickly spread of his misdeed. Most of the other families arrived back at the camp before him, and a man named Kenge announced: "Cephu is an impotent old fool. No, he isn't, he is an impotent old animal—we have treated him like a man for long enough, now we should treat him like an animal. Animal!" (Boehm 2012:38–39). This triggered an outpouring of gossip and condemnation of Cephu as, in Boehm's words, "a group consensus materialized" (Boehm 2012:39).

Cephu soon returned, and Kenge shouted at him that he was an animal. Everyone else was silent and for a short while just ignored him. Then, after the whole crowd confronted Cephu with vague accusations that he was selfish, someone finally made a direct accusation against him that he had stolen meat. One man said that "he hoped Cephu would fall on his spear and kill himself like the animal he was. Who but an animal would steal meat from others? There were cries of rage from everyone, and Cephu burst into tears" (Turnbull 1961; quoted in Boehm 2012:39–40).

At first Cephu argued that he had made an honest mistake. When no one accepted this excuse, he claimed that he deserved to place his net in a better position, considering that he was "an important man, a chief, in fact, of his own band" (referring to his extended family; Boehm 2012:40). Someone responded that the Mbuti do not have chiefs. If Cephu was a chief of his own band, "let him go with it and hunt elsewhere and be a chief elsewhere" (Boehm 2012:40)—a suggestion that, were it followed, would have meant starvation in the forest for Cephu and his family.

Faced with the threat of expulsion, Cephu began to apologize profusely. He agreed to turn over all of his meat from the day's catch to his accusers, and the other band members then took everything from him and his wife. Despite all the drama, a few hours later he participated with everyone else in the evening singing and all was set right again—and presumably he never repeated his misdeed.

Notice how this real-life example of enforcing a cooperative norm is fundamentally different from what was possible in the experimental setup of Gürerk et al. (2006; discussed in the introduction to this paper). All of the members of the group explicitly coordinated in advance, agreeing on what they *foresaw* would be a set of practices that would lead the group to collectively acquire as much meat as possible and to distribute it fairly. Members of the society did not have a realistic option to opt out of the system—they could not choose to participate in a "sanction free institution." The decision to adopt a cooperative institution and punish defectors/deviants was made collectively. In Cephu's case, we see that punishment was preceded by collective coordination. Kenge led the group by asserting that they should change the way they treat him: for his misdeed they should treat him as an animal. Before Cephu arrived, all of the members of the group (besides his immediate family members) had shared their sense of outrage with each other, and established that they were on the same page. When everyone confronted Cephu, they were not doing so as individuals expressing their personal anger, but as members of a group that had already come to a collective understanding.

The collective rebuke and punishment of Cephu illustrates how our moral sense plays an important role in undermining the radical individualistic approach of cultural evolutionary models. Chimp society is despotic and without morality. In humans, collectively imposed and accepted moral rules allow coordinated group action among all those who accept the rules. Hunter-gatherers use morality to enforce egalitarianism, whereas sedentary, resource-hording people use morality to enforce despotism. Whether morality is used to promote egalitarianism or despotism, it binds people into a collective decision-making body that cannot be legitimately atomized, as in cultural evolutionary models.

CULTURAL EVOLUTIONARY MODELS OF THE DEVELOPMENT OF PROSOCIAL NORMS

In cultural evolutionary models, cultural variants are distributed in a population, and individuals with content and context learning biases choose whom to imitate. Modeling work suggests that, in populations composed of social learners interacting with each other, stable behavioral patterns will emerge as a "by-product" of our learning biases (Chudek et al. 2013:442). These behavioral equilibria are more likely to be group *detrimental* than beneficial (Boyd and Richerson 1992; Henrich and Boyd 2001:86; Chudek and Henrich 2011:222). But, according to other models, cultural group selection favors the spread of those equilibria that happen to be beneficial (e.g., Boyd and Richerson 2002). Then, in a process of gene-culture coevolution, we become genetically disposed to be receptive to prosocial behavioral patterns, or "norms" (Henrich and Boyd 2001; Chudek and Henrich 2011; Chudek et al. 2013).

The aforementioned models assume that there is no organized enforcement of norms—such as the collective enforcement described in the previous section. Boyd and Richerson say that "[f]or most of human history, states were weak or non-existent, and norms"—such as "rules against murder"—"were not enforced by external sanctions" (Boyd and Richerson 2002:287). Each individual makes a decision whether or not to imitate the behavioral patterns represented in the group. Among these behavioral patterns are tendencies to *punish* those who fail to exhibit certain other behaviors. *Punishing norm violators* is just another cultural variant that can spread among individuals. A norm (whether it is beneficial or detrimental to the group) can be stabilized (in part) by punishment. Punishment is presumed to involve a cost, but it can be reinforced by the punishment of nonpunishers. Punishment of nonpunishers does not have

to go on forever (punishing those who fail to punish nonpunishers) because a combination of conformism and fear of punishment will drive cooperation at the first level to near fixation, which in turn lowers the cost of first-order punishment (Boyd and Richerson 1992).

As noted, Boyd and Richerson's models are based on the assumption that during the period of gene-culture coevolution in which norms first evolved and we became adapted to live in a norm-governed environment, "norms were not enforced by external sanctions" (Boyd and Richerson 2002:287). The ethnographic evidence discussed in connection with Boehm's theory suggests that this is false. Although there were no "states" or governments in the modern sense, there was a decision-making body—a coalition of the majority—that exercised coercive power, which amounts to what is effectively the same thing. The cultural variants to refrain from murder, to share meat, and so on were collectively adopted and formally *enforced*, not just punished by isolated individuals who possessed a variant to punish those who did not exhibit the relevant behaviors (or to punish in response to signals that others will join in their punishment, as in Boyd et al.'s 2010 model, discussed earlier).

Chudek et al. (2013) list three possible ways that different behavioral equilibria could be selected. The first is that "rational, forward-looking individuals" perceive the ultimate benefits of being in a cooperative equilibrium, "assume others are similarly sensible, and choose the prosocial state" (Chudek et al. 2013:439). They give three reasons why they think this was not a significant factor in real life. First, Chudek et al. say, people are not actually good at making the calculations required to determine what practices would be beneficial. Second, "group decisions are often heavily influenced by leaders and coalitions whose interests diverge from the overall group" (Chudek et al. 2013:439). Third, we see many examples of patently *non*prosocial institutions in societies throughout the world. Chudek et al. seem to recognize the possibility of "group decisions" (in the second reason), but they do not explore the implications of this phenomenon for cultural

evolutionary theory. Insofar as group decisions do play a role in cultural evolution, these authors think that they tend to reflect the narrow self-interest of powerful political factions. They do not consider the possibility that group decisions can be made by a coalition of the majority that is more or less interested in the success of the group as a whole—but if Boehm was right then this is in fact a big part of how groups in the ancestral environment selected among different possible behavioral equilibria to collectively enforce (cf. Singh et al. 2017:470–471).

The second possible way Chudek et al. (2013) mention to select among behavioral equilibria is "stochasticity." By chance, groups move from one equilibrium to another, and groups are likely to spend more time in equilibria with larger basins of attraction.

The third way is cultural group selection, which acts on the variation provided (in part) by stochasticity. Cultural evolutionists identify cultural group selection as by far the most important factor in the spread of group-beneficial norms. Groups randomly adopted more or less group-beneficial or detrimental norms, and those with beneficial norms survived, enjoyed more immigration, and their individuals were preferentially copied by individuals in other groups.

Boyd and Richerson (2002) provide a model showing that, in a population consisting of small groups with behavioral norms at different equilibria, *group-beneficial* norms can spread if people tend to copy more successful individuals in either their own or in neighboring groups. Individuals in groups with more group-beneficial norms will tend to be more successful and, therefore, their norms will be copied by the members of other groups. But this model fails to account for the empirical fact that moral norms are formally enforced (by collective action) within foraging bands. Individuals from one band cannot copy the moral norms of individuals from another (imagine if Cephu responded to the accusations against him by saying that he subscribed to the different meat-sharing practices of a neighboring people). However, a group can *collectively* decide to copy another, more successful group, and begin enforcing a new suite of norms.

Henrich describes "prestige-biased group transmission" (Henrich 2016:168) as a matter of "individuals" in one group copying "individuals" in more successful groups. But he also describes some striking cases where the *leadership* of one group decides to adopt the practices of another group. In the case of the Irakia Awa of New Guinea, "*senior men*" (Henrich 2016:173; emphasis added) decided to copy the pig-rearing practices of their more economically successful neighbors, the Fore. The process of transmission did not involve individuals copying individuals. In Henrich's words: "[T]his transmission between groups occurred rapidly because the Irakia already had a political institution in the village, which involved a council of the senior members of each clan, who were empowered by tradition (social norms) to make community-level decisions" (Henrich 2016: 174).

Power played a complicated role in the adoption of new communal practices among the Irakians. A few years after they copied the Fore, some young men did not want to continue raising pigs anymore. When they communicated their preference to the village elders, the elders "would not even discuss it [and] . . . disparaged the idea and criticized the younger people for being lazy and unwilling to lead proper lives. . . . The young men . . . admitted that it would be impossible to make such a change with the elders firmly against it" (D. J. Boyd 2001:270). However, another group of young adults converted to Seventh-Day Adventism—a religion that prohibits the consumption of pork. They outright refused to raise pigs, and no serious sanctions were imposed on them. Apparently the power (that probably took the form of *authority*) of the village elders was substantial but not absolute.

Although the Irakians were not foragers, this story illustrates how the assumptions of cultural evolutionary models (that transmission is from individual to individual and that variants are not imposed on groups of people all at once by means of power) do not apply to many real-life scenarios that have been studied by cultural evolutionists. Henrich's *verbal* description of how the Irakians adopted the Fore's practices refers to forces that cultural evolutionary *models* do not accommodate.

Gene-Culture Coevolution: The Development of "Norm Psychology"

As discussed, cultural evolutionary theorists argue that the learning biases that lead us to acquire adaptive cultural information tend to lead, when many individuals interact with each other, to the development of norms—"stable group-wide patterns of behavior" (Chudek et al. 2013:442). Cultural evolutionary models of the development of transmitted behavioral patterns show that a variety of norms can be stable (Chudek et al. 2013:438–439). Some equilibria are group beneficial, such as those involving widespread cooperation and punishment of defection. Most equilibria are group detrimental, such as those involving widespread noncooperation, or widespread, enforced performance of costly rituals (Chudek and Henrich 2011: 222 give several examples, including disease-spreading endocannibalism). Although group-detrimental norms are *more* likely than beneficial norms to develop out of the interaction among individuals within a group (according to the modelers), cultural group selection will nevertheless tend to favor the proliferation of the latter.

Groups that happen to have cooperative norms tend to outsurvive other groups, receive more immigrants, and (as discussed above) their individuals tend to be selected as models to imitate by individuals in other groups. On Chudek et al.'s (2013) account, because norms inevitably arise as a "by-product" of cultural learning, natural selection would have favored innate dispositions for handling norms—a "norm psychology." They suggest that our innate norm psychology prepares us to recognize, and motivates us to observe, the social norms in our environment. Since cultural group selection caused cooperative norms (including norms to punish defectors and punish nonpunishers) to prevail in the ancestral environment, our norm psychology should make us especially disposed to acquire prosocial norms and punish violators (see also Henrich and Boyd 2001:87). Indeed, in accordance with Chudek et al.'s

predictions, evidence suggests that young children automatically infer behavioral rules by observing people (particularly adults), and that they internalize rule adherence as a personal goal and even enforce observance in other children (Chudek and Henrich 2011; Chudek et al. 2013).

The theories of both Boehm and Chudek et al. predict the existence of a "norm psychology." Both claim that Paleolithic hunter-gatherers lived in norm-governed societies and that individual success was tied to the ability to recognize and follow local norms, which tended to be prosocial. Laboratory experiments revealing our disposition to infer and follow norms will not adjudicate between the theories.

The theories, although making virtually the same predictions about how recent hominin evolution shaped our psychology, give different explanatory accounts of how prosocial norms initially spread, and they paint radically different pictures of the human capacity to *deliberately* guide our own evolution (both cultural and genetic). The empirical evidence reviewed in this paper seems to favor Boehm. Among nomadic hunter-gatherers around the world—all that have been studied by anthropologists—powerful coalitions enforce prosocial norms with a great deal of explicit awareness of the social benefits that observance of these norms will produce. These coalitions enforce the norms because individuals have an explicit blueprint in their minds for what kind of society they want to live in, and norm enforcement is a deliberate strategy for bringing that society about. In many cases humans are blind to the consequences of their socially transmitted practices—the evidence for that is undeniable (Henrich 2016). But our hunter-gatherer ancestors were not blind to the consequences of all of their practices. When coalitions of the majority seized power from alphas, they gained the ability to impose practices that had consequences that they favored.

AN ALTERNATIVE EXPLANATION FOR THE
DEVELOPMENT OF PROSOCIAL BEHAVIOR:
THE "MUTUALISTIC THEORY" OF MORALITY

Baumard et al. (2013) defend an alternative explanation for the evolution of moral judgment and prosocial behavior. They suggest that, in the ancestral environment, people could benefit by cooperating with each other in a variety of ways: they could hunt together, share food with the expectation of future reciprocation, and so on. Each person would bring a certain amount of resources, effort, and talent to cooperative ventures. This led people to *compete* to be chosen as partners for cooperation. If *A* will get a lower return by cooperating with *B* than *A* would get on average from cooperating with someone else, *A* will be better off not choosing *B* as a partner, and *B* will lose out. Our ancestors faced the adaptive challenge of seeking out good exchange partners while making themselves attractive exchange partners to others. Baumard et al. argue that, as a consequence, we evolved a moral disposition to value "fairness"—the notion that people are entitled to share in the product of cooperation in proportion to their contribution to creating that product. In hunter-gatherer tribes the moral order was not sustained by collective agreement about rules and collectively administered punishment, but by individuals exercising "partner choice."

Baumard et al.'s (2013) theory seems to explain a range of findings in experimental games. For example, in dictator games, where one person (the dictator) decides how to divide a pot of money with an anonymous partner, dictators tend to keep most of the pot for themselves. But suppose the dictator is asked to distribute money that was *earned* by the anonymous partner by performing well in a quiz contest or on an exam. In that case, dictators tend to become very generous, and give their partners more or less what they earned (Ruffle 1998; Oxoby and Spraggon 2008). We seem to intuitively feel that people are entitled to the product of their efforts and, according to Baumard et al. (2013), we object to inequality only when it is the consequence of *unfairness*.

The theories of Baumard et al. and Boehm can, to some extent, be reconciled. It could be that groups began collectively enforcing moral codes around 250,000 BP (as Boehm argues), but voluntary mutualistic cooperation contributed to the evolution of our moral sense (in the way described by Baumard et al.) *before* that time. Even if early human

groups were dominated by alphas who bullied group members and sometimes appropriated resources for themselves by force, not all interactions had to be based on bullying: group members could still have cooperated with each other in some contexts, and competed to be chosen as cooperation partners. Furthermore, even after hunter-gatherers established a reverse dominance hierarchy and egalitarian political norms, people were still allowed to obtain unequal rewards due to greater ability in some contexts. Although the rules among nomadic foragers for sharing meat from big game generally demand a more or less equal distribution, the rules for sharing other kinds of food are often much looser, and people who contribute more to obtaining such food may have a degree of freedom to distribute it as they wish (see Gurven 2004). Even if, since 250,000 BP, hunter-gatherer bands were politically egalitarian (among males) and big game hunting/distribution was a *largely* socialist enterprise, people still formed cooperative relationships with each other on a smaller scale, and it would have been advantageous for them to follow norms of fairness even if these norms were not always enforced by the collective effort of the group.

Although some of our behavior and moral intuitions can be explained by Baumard et al.'s (2013) account, we cannot discount the evidence that antisocial behavior is (and presumably was) often collectively punished in egalitarian bands, and that our moral psychology is to some extent tailored to living in such egalitarian bands. As noted, Baumard et al. suggest that antisocial hunter-gatherers are, in general, not formally "punished," but rather they suffer when other members of their group refuse to cooperate with them. But the anthropological evidence reviewed above suggests that hunter-gatherers do not merely withdraw cooperation from antisocial individuals. Groups collectively mete out positive punishments—the death penalty is widely administered for a range of offenses. It is true that, as Baumard et al. say, "[p]unishment . . . is uncommon in societies of foragers" (Baumard et al. 2013:66). But this can be explained by the fact that the threat of punishment preemptively stops most serious offenses (cf. Service 1975:48–49). The death

penalty rarely needs to be administered because it is preceded by many escalating warnings of what is to come should offenders fail to reform.

The hypothesis that our moral sense is calibrated for living under conditions of somewhat enforced egalitarianism also explains some results from experimental games, which Baumard et al. (2013) cannot so easily explain. They discuss the following experiment conducted by Dawes et al. (2007). Subjects were divided into groups containing four anonymous members for a one-shot interaction. Each subject was given a random amount of money by a computer, and was informed how much money had been granted to the other three members. The players then had the chance to give "positive" or "negative" tokens to each other. Giving a positive or a negative token cost the giver one monetary unit, and increased or decreased (respectively) the recipient's payoff by three monetary units. Groups were broken apart and formed with new anonymous members to play again for a total of five rounds.

The results of Dawes et al.'s study suggest that people are willing to pay a penalty to equalize outcomes. In the course of five rounds, 68% of subjects "reduced another player's income at least once, 28% did so five times or more, and 6% did so ten times or more. . . . 74% . . . increased another player's income at least once" (Dawes et al. 2007: 794). Those who received a very high initial endowment tended to receive many negative tokens, while those who received a very low initial endowment tended to receive many positive ones. Subjects' token-distributing behavior cannot be seen as a rational strategy because, again, the interactions were one shot. It cannot be interpreted as retaliation or punishment, since subjects knew that all of the payoffs had been determined randomly by a computer. Rather, subjects seemed to be bothered by inequality per se. Baumard et al. offer the following explanation: "Overall, the distribution of [tokens] . . . displays the logic of fairness: The more a participant received money, the more others would 'tax' her. Conversely, the less she received, the more she would get 'compensated'" (Baumard et al. 2013:75). However, the way subjects reacted to inequality

does *not* make them attractive partners for cooperation. Paying a personal cost in order to reduce the payoff of luckier individuals simply reduces the expected payoff from engaging in cooperation. Unlike a "tax," which *redistributes* payoffs and thereby benefits the recipient, the negative tokens administered by 68% of subjects reduced both their own and the recipients' payoffs *without benefiting anyone*. Thus, the disposition that motivates negative-token giving does *not* make us attractive partners in cooperation. But the behavior is easily explained if we assume that we are adapted to living in hunter-gatherer bands where a certain degree of egalitarianism was enforced, and people were prohibited from accumulating significantly more resources than their fellows.

Discussion

Mainstream cultural evolutionary theory in the Boyd and Richerson tradition assumes that cultural variants spread as a consequence of individuals' content and context learning biases, Darwinian selection, and random forces in a way that is amenable to mathematical modeling. Durham (1991), Fracchia and Lewontin (2005), and Lewontin (2005) raise the challenge that often cultural variants are spread through the exercise of power, implying that the target of explanation for cultural evolution should be the behavior of powerful individuals and groups. This paper argued that Durham and Fracchia and Lewontin's critique may be valid as far as the evolution of morality and prosociality goes (at least since 250,000 BP), but less so for the spread of non-moral cultural variants (perhaps until the advent of sedentarism, food storage, and agriculture).

Cultural evolutionists emphasize that their models are not intended to capture everything that happens in real life (Richerson and Boyd 1987; McElreath and Boyd 2007: 4–6). Like all models, they are meant to be simplifications that isolate some of the main forces at play. But if something like Boehm's (1999, 2012) account of the evolution of morality is correct, cultural evolutionary models of the evolution of morality and the coevolution of norms and norm psychology are *distortions*, not simplifications. A legitimate model may isolate one or a few forces among the multitude that exist in the real world. It cannot postulate forces that are not operative at all, or ignore those that have a decisive influence on the phenomena under investigation.

In light of the anthropological evidence reviewed in this paper, we can see how Boyd et al.'s (2010) model of the evolution of "coordinated punishment" distorts history. According to Boyd et al., there was a genetic mutation(s) associated with the behavior to *punish cheaters if τ other people in my group (honestly) signal that they are prepared to punish cheaters* (again, this mutation can survive better than the mutation associated with *always punish cheaters*, since always punishers are disadvantaged when they are a small minority). In a population containing some conditional punishers, by chance a few groups will have the threshold number (i.e., $\tau + 1$), and they will prosper and replace other groups, thereby increasing the number of conditional punishers.

The story that Boehm (1999, 2012) tells is very different. He says that coalitions within human groups formed and explicitly agreed to enforce certain kinds of prosocial behavior through punishment. The people involved did not have any preexisting disposition to punish. The practice of punishing people to enforce prosocial norms arose all at once. In contrast, on Boyd et al.'s (2010) account the genetically based disposition to punish arose *first*, then people developed a social norm to punish. On Boehm's (1999, 2012) account, because groups started practicing collective punishment, this created selection pressures for people with a genetic disposition to punish norm transgressors.

To be clear, the point of this paper is *not* to say that there are real-life complexities that cultural evolutionary models fail to capture. The point is that some of the fundamental forces that drive cultural evolution, and drove the coevolution of norms and norm psychology, cannot be accommodated by the sorts of models used in cultural evolutionary theory. Most of the mathematical models employed in cultural evolutionary theory are best adapted to cases where forces act in regular, iterative ways. Perhaps evolution in some domains of culture does work like that—cultural evolutionary modeling in these domains

is safe from Durham (1991) and Fracchia and Lewontin's (2005) challenge. In other domains—including the moral—a more traditional historical/anthropological/sociological approach may be more fruitful (and legitimate).

REFERENCES

Bachrach P., Baratz M. S. 1970. *Power and Poverty: Theory and Practice.* New York: Oxford University Press.

Baumard N., André J.-B., Sperber D. 2013. A mutualistic approach to morality: the evolution of fairness by partner choice. *Behavioral and Brain Sciences* 36:59–78.

Boehm C. 1993. Egalitarian behavior and reverse dominance hierarchy. *Current Anthropology* 34:227–240.

Boehm C. 1997. Impact of the human egalitarian syndrome on Darwinian selection mechanics. *American Naturalist* 150:S100–S121.

Boehm C. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior.* Cambridge (Massachusetts): Harvard University Press.

Boehm C. 2012. *Moral Origins: The Evolution of Virtue, Altruism, and Shame.* New York: Basic Books.

Bowles S., Gintis H. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution.* Princeton (New Jersey): Princeton University Press.

Boyd D. J. 2001. Life without pigs: recent subsistence changes among the Irakia Awa, Papua New Guinea. *Human Ecology* 29:259–282.

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

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

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

Boyd R., Gintis H., Bowles S. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328:617–620.

Cashdan E. A. 1980. Egalitarianism among hunters and gatherers. *American Anthropologist* 82:116–120.

Chudek M., Henrich J. 2011. Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences* 15:218–226.

Chudek M., Heller S., Birch S., Henrich J. 2012. Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior* 33:46–56.

Chudek M., Zhao W., Henrich J. 2013. Culture-gene coevolution, large-scale cooperation, and the shaping of human social psychology. Pages 425–457 in *Cooperation and Its Evolution*, edited by K. Sterelny, R. Joyce, B. Calcott, and B. Fraser. Cambridge (Massachusetts): MIT Press.

Dahl R. A. 1957. The concept of power. *Behavioural Science* 2:201–215.

Darwin C. 1871. *The Descent of Man, and Selection in Relation to Sex*, Volume 1. London (United Kingdom): John Murray.

Dawes C. T., Fowler J. H., Johnson T., McElreath R., Smirnov O. 2007. Egalitarian motives in humans. *Nature* 446:794–796.

Durham W. H. 1991. *Coevolution: Genes, Culture, and Human Diversity.* Stanford (California): Stanford University Press.

Erdal D., Whiten A. 1994. On human egalitarianism: an evolutionary product of Machiavellian status escalation? *Current Anthropology* 35:175–178.

Fracchia J., Lewontin R. C. 2005. The price of metaphor. *History and Theory* 44:14–29.

Goodall J. 1986. *The Chimpanzees of Gombe: Patterns of Behavior.* Cambridge (Massachusetts): Harvard University Press.

Gürerk Ö., Irlenbusch B., Rockenbach B. 2006. The competitive advantage of sanctioning institutions. *Science* 312:108–111.

Gurven M. 2004. To give and to give not: the behavioral ecology of human food transfers. *Behavioral and Brain Sciences* 27:543–559.

Henrich J. 2006. Cooperation, punishment, and the evolution of human institutions. *Science* 312:60–61.

Henrich J. 2016. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter.* Princeton (New Jersey): Princeton University Press.

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., McElreath R. 2003. The evolution of cultural evolution. *Evolutionary Anthropology* 12:123–135.

Horner V., Proctor D., Bonnie K. E., Whiten A., de Waal F. B. M. 2010. Prestige affects cultural learning in chimpanzees. *PLOS ONE* 5:e10625.

Knauft B. M. 1991. Violence and sociality in human evolution. *Current Anthropology* 32:391–409.

Lewens T. 2007. *Darwin.* London (United Kingdom): Routledge.

Lewens T. 2015. *Cultural Evolution: Conceptual Challenges.* Oxford (United Kingdom): Oxford University Press.

Lewontin R. C. 20 October 2005. The wars over evolution. *New York Review of Books.* http://www.nybooks.com/articles/2005/10/20/the-wars-over-evolution/.

Lukes S. 2005. *Power: A Radical View.* Second Edition. New York: Palgrave Macmillan.

Mameli M. 2013. Meat made us moral: a hypothesis on the nature and evolution of moral judgment. *Biology & Philosophy* 28:903–931.

McElreath R., Boyd R. 2007. *Mathematical Models of Social Evolution: A Guide for the Perplexed.* Chicago (Illinois): University of Chicago Press.

McGrew W. C. 2004. *The Cultured Chimpanzee: Reflections on Cultural Primatology.* Cambridge (United Kingdom): Cambridge University Press.

Ostrom E., Walker J., Gardner R. 1992. Covenants with and without a sword: self-governance is possible. *American Political Science Review* 86:404–417.

Oxoby R. J., Spraggon J. 2008. Mine and yours: property rights in dictator games. *Journal of Economic Behavior & Organization* 65:703–713.

Richerson P. J., Boyd R. 1987. Simple models of complex phenomena: the case of cultural evolution.

Pages 27–52 in *The Latest on the Best: Essays on Evolution and Optimality,* edited by J. Dupré. Cambridge (Massachusetts): MIT Press.

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

Richerson P. J., Henrich J. 2012. Tribal social instincts and the cultural evolution of institutions to solve collective action problems. *Cliodynamics: The Journal of Theoretical and Mathematical History* 3:38–80.

Ruffle B. J. 1998. More is better, but fair is fair: tipping in dictator and ultimatum games. *Games and Economic Behavior* 23:247–265.

Service E. R. 1975. *Origins of the State and Civilization: The Process of Cultural Evolution.* New York: Norton.

Singh M., Wrangham R., Glowacki L. 2017. Self-interest and the design of rules. *Human Nature* 28:457–480.

Turnbull C. M. 1961. *The Forest People: A Study of the Pygmies of the Congo.* New York: Simon and Schuster.

Associate Editor: Michael Ruse
Handling Editor: Daniel E. Dykhuizen