ELSEVIER

Evolution and Human Behavior 26 (2005) 47–63

Evolution and Human Behavior

The better angels of our nature: group stability and the evolution of moral tension☆

David C. Lahti[a,*], Bret S. Weinstein[b]

[a]*Program in Organismic and Evolutionary Biology, Morrill Science Center, University of Massachusetts, Amherst, MA 01003, United States*

[b]*The Evergreen State College, Lab II, 2700 Evergreen Parkway NW, Olympia, WA 98505, United States*

Initial receipt 3 November 2003; final revision received 2 September 2004

Abstract

Moral systems require individuals to act in service to their social groups. Despite the human tendency to view moral norms as invariant and constantly deserving of adherence, we vary not only in the moral norms that we espouse but also in the degree to which our behavior reflects those norms. Nevertheless, moral systems exhibit patterns and complexity that suggest the action of natural selection. We propose that much observed variation in commitment to the group can be explained by a rule of *stability-dependent cooperation*, where the adaptive level of individual commitment varies inversely with the stability of the social group. This hypothesis is rooted in the understanding that humans are caught in an evolutionary trade-off between two methods of increasing reproductive success: competing with fellow group members, and increasing the stability of the group relative to other groups. If cooperation is stability dependent, however, human groups in times of high stability and low cooperation may be susceptible to fast-acting extrinsic threats, as well as self-destructive competitive races to the bottom. In light of this, we hypothesize that the absolutism and unchangeableness commonly attributed to moral norms serve a *group stability insurance* function and we present predictions from this hypothesis.

Keywords: Morality; Evolution; Intergroup competition; Cooperation; Altruism

☆ The authors contributed equally to this paper.

* Corresponding author.

E-mail addresses: lahti@bio.umass.edu (D.C. Lahti), bretw@evergreen.edu (B.S. Weinstein).

1090-5138/04/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2004.09.004

1. Introduction

> We are not enemies, but friends. We must not be enemies. Though passion may have strained it must not break our bonds of affection. The mystic chords of memory, stretching from every battlefield and patriot grave to every living heart and hearthstone all over this broad land, will yet swell the chorus of the Union, when again touched, as surely they will be, by the better angels of our nature.
>
> Abraham Lincoln. Second Inaugural Address (1861)

The history of evolutionary approaches to morality has been characterized by debate between those who claim that "evolution has, as a matter of fact, constructed human beings to act for the community good" (Richards, 1987, pp. 623–624; see also Ruse, 1986; Sober & Wilson, 1998) and a diverse opposition maintaining either that morality is selectively neutral or that it requires us to combat the evolutionary process and "rebel against the tyranny of the selfish replicators" (Dawkins, 1976; see also Huxley, 1894; Williams, 1988). Despite the productivity of this discussion (Maienschein & Ruse, 1999), this paper argues that neither general perspective is sufficiently explanatory; moral action is not always adaptive, but neither is it neutral or maladaptive. We argue instead that *the propensity for moral deliberation* is the fitness enhancing characteristic, any given moral action shifting between adaptive and maladaptive, depending on context. This paper draws attention to the significant variation that humans exhibit in individual commitment to moral norms and proposes that the variation represents generally adaptive responses to dynamic social environments.

Although psychologists are intensely aware of the importance of social influences on morality, they are less aware of whether and how individual commitment to moral responsibility varies with societal variables (Hartup & van Lieshout, 1995). This gap provides an opening for hypotheses that predict empirical trends based on an evolutionary consideration of the nature of morality. This paper presents two principles that predict variability in individual commitment to moral norms as a function of one's perceived social context.

A background assumption for this discussion is that a moral rule tends to be manifest in consciousness as absolute, in two senses. First, when people promote one alternative as morally correct, they imply that it is superior to all others in some general way. Regardless of our actions or desires, we tend to treat moral rules as deserving of absolute adherence. Second, humans tend to consider moral rules absolute in the sense that they carry an implication of permanence across time and space. Despite moral variation within and between individuals, humans tend to operate under the assumption of an underlying truth to moral rules that does not change (Mackie, 1981, chap. 1). This idea that moral absolutism is widespread and general is a refutable psychological and sociological hypotheses. If morals are deemed absolute in the first sense (deserving of absolute adherence), people should consistently endorse alternatives that they deem morally right, even if their behavior or desires conflict with these moral precepts; and if morals are deemed absolute in the second sense (invariable in time and space), people should tend to evaluate the attitudes and behaviors of others according to their own moral belief systems, without regard for cultural or historical differences.

2. Commitment to moral rules varies

The central problem to be addressed in this paper is that, despite moral absolutism, people's lives exhibit variation in commitment and adherence to the moral rules that they recognize. Humans deliberate, calculate, and often struggle—sometimes adhering to the rules, sometimes not. Perhaps, the central paradox of morality is the fact that behavior does not always match the moral rules espoused by the agent. Moral rules are considered absolute, but adherence is facultative.

When social and natural scientists have asked why this apparent incongruity exists, their answers often fall into two broad categories. One general solution is to view moral rules as contrary to self-interest, such that the two are continually in opposition. The other, perhaps more common, solution is to see morality as always consistent with globally calculated self-interest, and our moral struggles and deliberation as internal conflict between short- and long-term self-interest. Both these alternatives interpret variation in commitment to moral norms as a maladaptive by-product of a weakness or inconsistency in human psyche. Either the difficulty of self-sacrifice or the difficulty of foregoing short-term benefits for long-term ones is proposed as the psychological constraint that limits compliance with moral rules.

An evolutionary perspective raises doubts about the explanatory power of both these solutions. If moral behavior were simply a hindrance to the competitive ascendancy of the individual, one would expect it to dwindle, and if it were simply adaptive, then absolute compliance would evolve and presumably sweep to "fixation." Given this situation, Peters (2003) recently drew attention to the fact that evolutionary studies of morality have still failed to produce an effective explanation of the fact that humans appear to be disposed both toward and against prosocial or group-serving behavior.

What, therefore, needs to be explained? What is the hypothesized locus of adaptation? An attempt to explain the adaptive significance of perfectly moral behavior, although a common goal, is misguided because such a behavior is never observed. We contend that a successful theory must address the adaptive significance of the *facultative* adherence to moral absolutes. From this perspective, the traits that appear to be adaptations are the capacity for moral behavior and the tendency toward moral deliberation, as distinct from the execution of any particular behaviors all the time. This is similar to the suggestion of Nesse (2001) that natural selection may have shaped "commitment strategies" for effective use in society. If this is correct, the key to understanding morality from an evolutionary perspective lies in discovering the extrinsic factors that govern moral deliberation and moral commitment. The merit of an evolutionary explanation for moral behavior can be judged on its ability to predict what conditions will produce compliance versus defection. Can variability in human environments explain our plasticity in following the rules?

3. Morality and intergroup competition

Richard Alexander has shown that two related facts are key elements in an evolutionary understanding of morality. First, humans "evolved to live in groups, within which they both

cooperate and compete and outside of which they presumably failed consistently;" secondly, "some acts of costly beneficence enable the survival of the entire group, when that outcome is essential for our own survival" (Alexander, 2004; see also Alexander, 1987, 1992). Social grouping evolved in humans in an unprecedented way, with low within-group relatedness (relative to eusocial animals) and multiple breeding males within groups. Alexander built on earlier writers such as Darwin (1871) and Keith (1949) in explaining the evolution of this phenomenon. At first, our grouping was probably maintained by selection for predator avoidance, and later, for cooperative group hunting (Alexander, 1989), but eventually, as humans reduced their susceptibility to "hostile forces of nature," the main threat to an individual's reproductive success became other people, and competition among human groups became the primary function of group living. Human cooperation within groups, then, probably evolved as a way to compete between groups, as individual reproductive success was served increasingly by maintaining solidarity with one's fellows (Alexander, 1990). The most important mechanisms for this cooperation appear to have been (a) extensive and differential nepotism and, arising in this context, (b) social reciprocity of two different sorts. The direct sort is the process of "indefinitely continuing interactions between intelligent beings in which each can benefit from cooperating with the other, and. . .defection. . .will in the long run represent net losses to the defector" (Alexander, 1992; see also Trivers, 1971). The indirect sort of reciprocity arises when multiple parties interact in the same way that two do in direct reciprocity. In a species with powers of observation, memory, and communication of individual reputations within a social group, rewards for cooperating (and punishments for defecting) can be administered by "society at large, or from other than the actual recipient of beneficence" (Alexander, 1979).

If humans have tended throughout their history to fail outside social groups, and if threats from other groups have rendered the suppression of competition within a group necessary for individual reproductive success, it is not surprising that individuals should often subjugate their own interests to those of their groups. Insofar as morality requires attention to group causes, such as the welfare of others, morality functions as social cement and thus tends to contribute to long-term individual interests, that is, fitness. Moreover, as indirect reciprocity became important for insuring service to the group, individuals perceived as morally upstanding would gain additional benefits through the approbation of others. These two interactive processes are broadly similar to the two kinds of games (public goods and image scoring) that recent experimental studies have employed to illustrate the dynamics of human cooperation (e.g., Milinski, Semmann, & Krambeck, 2002).

Indirect reciprocity may be able to account for part of the facultative or inconstant nature of human commitment to moral norms. Indirect reciprocity is a system in which one's reputation, built on others' observations of one's behavior in the past, affects one's prospects. It is not that a particular moral or immoral action inherently enhances or degrades the actor's fitness. The net effect is dependent on (a) the direct costs and benefits of the action, (b) the likelihood of being observed or of reports being believed, (c) the reputational shift that will result, and (d) the expected return on that shift in future interactions. Sensitivity to cues of these parameters and their net effect on fitness would account for a degree of nonrandom variation in adherence to moral norms.

Such, perhaps, is the adaptive value of the refined moral systems characteristic of extant human groups, where the proceeds from indirect reciprocity arguably have grown to be more relevant to decision making than did the proceeds from continued group persistence. For an average member of a modern group, the likelihood of suffering a significant fitness cost from the damage to reputation that an immoral choice can produce is much higher than the likelihood of suffering a significant fitness cost from the loss of group unity that might arise from that choice. This is so even if Darwin and Alexander are correct in arguing that the need for solidarity against threats from other groups is precisely what drove the evolution of human cooperation, including those actions maintained by indirect reciprocity, in the first place. Thus, the system of moral reputation may have greater and more immediate fitness effects today than do the selective pressures that favored that system's origin.

To elaborate more fully the relationship between indirect reciprocity and human social structure, one can demonstrate that indirect reciprocity, today, depends on a concept of group service, but the existence of group service does not depend on indirect reciprocity. Indirect reciprocity is a means that needs an end; it requires a criterion for a "good reputation." If indirect reciprocity functioned independently of its evolutionary origin in intergroup competition, one's reputation would merely reflect shrewdness rather than group service, but service to others is a cardinal value fostered by indirect reciprocity. Humans expend significant effort debating the validity of claims of selflessness and bestowing praise for actions deemed selfless. Shrewdness does persist, because it can sometimes be an effective way to exploit the system, but it is discouraged by moral norms and thus suppressed by indirect reciprocity. The centrality of the concept of service or selflessness in moral norms (Ridley, 1996; Roes & Raymond, 2003) suggests that within-group cooperation in the face of intergroup competition still underlies indirect reciprocity today. Otherwise, entertaining the idea that the actions of others can have group-serving motivations would be maladaptive. Moreover, for the earliest form of indirect reciprocity to produce fitness benefits, a belief in group service must already have existed, implying an independent evolutionary origin. The human social situation of within-group cooperation as a form of between-group competition explains why the concept of selflessness, rather than shrewdness, is the value encouraged by indirect reciprocity.

Following this reasoning, we propose that there was a period in the evolution of morality when group service was adaptive due to rising intergroup competition, but before indirect reciprocity became dominant. We are not aware of any previous proposal of such a period. Furthermore, the dynamics that drove cooperation at this intermediate stage in the evolution of morality may still be important today. Because moral norms are still pervaded by a strong group-service element, something about human social group dynamics is probably still providing the values for indirect reciprocity, that is, determining what behaviors will be productive of what kinds of reputation.

4. Group stability

Competition between groups implies that groups, like individuals, vary in how well they are doing. In the absence of a governing structure, such that groups are autonomous or nearly

so, competition between groups for limited resources will function like competition between individuals, with variation in the groups' likelihood of persistence analogous to the concept of fitness that evolutionary biologists use to compare individuals. (Because "fitness," in a biological sense, generally refers to contribution to succeeding generations via reproduction, and because human groups do not reproduce as a whole or have discernable generations, we avoid the term fitness and use the term "group stability" to reflect this relative likelihood of group persistence).

4.1. Principle of stability-dependent cooperation

If humans are facultative in adherence to moral norms (Section 2), and moral norms arose and are probably still maintained in the context of intergroup competition (Section 3), then variation in individual commitment to moral norms over time and space may reflect variation in the degree to which groups require service. Thus, the first of two principles we introduce to describe the dynamical function of morality in human history is *stability-dependent cooperation*. We propose that people vary in the relative importance that they place on the individual versus the group in their working value systems or decision rules because individual sacrifice in service to the group at a given time and place is adaptive in inverse proportion to the stability of the group relative to its competitors.

The general idea of adaptive variation in moral psychology has some precedent. Even proponents of a relatively strict developmental structure to morality have allowed for and found indirect evidence of apparently adaptive differences among cultures in the way morality is used to guide individual decisions (Edwards, 1975; Nisan & Kohlberg, 1982). In addition, evolutionary psychologists and anthropologists have shown that rules of social exchange can vary in ways that are predictable from environmental conditions (review in Cohen & Vandello, 2001; Cosmides & Tooby, 1992). The principle of stability-dependent cooperation offered here predicts adaptive variation specifically in *commitment* to moral rules, as distinct from the rules themselves or the development of their recognition. The foundation for this principle is the dynamic of natural selection in situations where individuals with divergent interests exist in collectives on which their persistence depends. A parallel dynamic best explains the overarching cooperation of genetic elements temporarily united in a genome. The genome works together and subsets only rarely seek their own interests at the expense of other elements (Buss, 1987) because the persistence of a gene or chromosome depends on the survival and reproduction of the individual housing it. Cooperation to increase individual fitness is therefore usually the best strategy for a genomic element. When genetic elements behave competitively within a genome, as in T-haplotype mice (Lyon, 2003) this tends to produce negative fitness consequences for the individual and, thus, for all other elements within it.

Likewise, individual humans depend on their social groups. Service to group causes fosters unity and can decrease the effects of resource limitation (e.g., restrict hoarding and squandering, mitigate distribution of wealth effects, and otherwise increase efficiency of resource use), thereby decreasing within-group competition and increasing the group's stability and competitive prowess (Alexander, 1979; Frank, 2003). Conversely, within-group competition

arising from individual self-interest can be self-defeating: Resource utilization becomes less efficient, and group unity erodes, increasing susceptibility to intergroup competition.

Some writers, often extrapolating from economic models, have hypothesized a general tendency of cooperation either to decay or to fluctuate in regular boom and bust cycles (Nowak & Sigmund, 1998). Some have gone on to suggest that, given these proposed tendencies, for cooperation to be maintained and group persistence to be assured over time, a certain specialized trait must have evolved and persisted at some threshold level in the population, such as "strong reciprocity" involving costly punishment of the selfish (Gintis, 2000) or "phenotypic defection" involving unintentional lack of service (Lotem, Fishman, & Stone, 1999). Although punishment of various sorts of noncooperators are certainly features of human culture (Axelrod, 1984), our hypothesis of stability-dependent cooperation is an alternative explanation for the persistence of groups. We propose that the reason why cooperation does not automatically collapse or cycle in the way suggested by economic models is because such models have not yet taken into consideration the general human tendency towards *facultative* adherence to moral norms and the resulting negative feedback on booms and busts of cooperation.

One's reproductive success can be advanced either by serving one's group (thereby slightly increasing the reproductive success of all group members relative to others) or by more immediately serving oneself (thereby increasing one's reproductive success relative to that of other group members). In many situations, these two options imply a continuum of behavioral options, or even two mutually exclusive options. This is the point at which many evolutionary studies of human behavior apply the terms altruism and selfishness, a misleading dichotomy that begs the question of which course of action is, in fact, adaptive for an individual in a given situation. A more precise set of terms would reflect the fact that certain behaviors are adaptive because they increase the between-group component of reproductive success (enhance group stability), whereas others are adaptive because they increase the within-group component (increase individual fitness relative to other group members). Most behavioral options probably affect both components of fitness, which necessitates a continuum rather than a dichotomy. For convenience, we will use the terms group service and self-service to refer to the two possible directions of movement on such a continuum.

The principle of stability-dependent cooperation implies that an individual's assessment of group stability is a major determinant of behavior in situations where group and self-service prescribe different courses of action. The more that group stability is threatened, the more that group service is likely to be adaptive for an individual, while the more successful or stable the group, the more adaptive is the relaxation of individual commitment to the group and an increase in self-serving behavior because the detriment to group stability of such relaxation is small. Security in the persistence of the group renders restraint from self-serving strategies less critical. If this hypothesis is correct, no single level of cooperation is adaptive all the time. Thus, we propose that the adaptation is not cooperation per se, but the propensity to evaluate the optimal level of cooperation in a given situation.

One prediction from this hypothesis is that studies of moral judgments in the social context (e.g., Carpendale & Krebs, 1992, 1995) will find the perception of group stability (a variable hitherto untested) to be a significant determinant of the outcome of moral deliberation. In

addition, the large proportion of unexplained variation in cross-cultural studies of moral intuitions (e.g., 39–44% in O'Neill & Petrinovich, 1998) may be reduced by taking a measure of perceived group stability into account. First, behavior should more closely approach the ideal of golden rules of general beneficence in times and places where group stability is threatened, whereas when group stability is more assured, "Do unto others. . ." may still be a mantra, but the evidence should indicate a more competitive edge to intragroup interactions and a tolerance of such competition in the community. Second, patriotism is a value that should be most emphasized and displayed when the nation is threatened, and actions that undermine national unity should be better tolerated in a time of peace than in time of war. Third, generosity and magnanimity should be higher within less stable groups and lower within more stable groups, because when one's group is doing poorly, group members are "all in it together" and should be more disposed to share and affiliate, and self-service should be viewed dimly. A sense of communal struggle will lose impact in times and places of group success, however, and individuals may get away with, and tolerate in others, more "materialistic" and competitive behavior, such as hoarding and extravagance. Indirect evidence for such trends already exists. For example, adversity tends to increase individual commitment to causes that are central to a person's values (Lydon, 1990; Brickman, 1987).

4.2. Group stability and indirect reciprocity

Indirect reciprocity provides the means for individuals to gain information about the dedication of others to the group, act to minimize parasitism by freeloaders or cheats, and allocate benefits to individuals in proportion to their level of commitment (Alexander, 1987). If, as proposed in Section 3, the dynamic of within-group cooperation in the context of intergroup competition maintains indirect reciprocity, determining the bases upon which reputations are made and broken, then the workings of indirect reciprocity may be expected to covary with group stability by a two-part mechanism.

First, one's assessment of group stability will affect the attitude that one has towards the behaviors of others. When group stability is under threat, people will be especially assiduous in assuring that others serve the group both because that service can aid group stability, which is a high priority, and because one's own restraint from self-serving action raises one's vulnerability to competitive exploitation by any group member who fails to exhibit similar restraint. Thus, unilateral group service is unlikely to be adaptive for individuals with the means to behave self-servingly; rather, group service will tend to be adaptive only when other members are also practicing it, and each individual has a stake in ensuring the cooperation of others.

When group stability is relatively assured, the principle of stability-dependent cooperation implies that people will be more permissive of intragroup competition, the balance thus shifting away from group service. Of course, one would prefer that everyone, except oneself, maintains a high level of group service regardless of group stability; but because apparent hypocrisy is disproportionately damaging to one's reputation, one cannot pursue self-interest alone, except by deception (which is risky) or despotism (which requires rare power). Absent these options, the best way for one's competition to be tolerated by the group is to foster tolerance of competition in general. On the group level, this dynamic pattern of individual

attitudes means that community enforcement of group service will be variably strict, depending on how members of the community perceive its stability.

The second reason to expect covariation between indirect reciprocity and group stability is that the assessment of the actions and attitudes of others involved in indirect reciprocity will be important in determining the optimal range of individual commitment to the group. People look to others not only to enforce their commitment but also to determine what level of commitment is required. As group stability changes, the tendency of the community to enforce service to the group will also change, producing positive feedback, as each assessor is also a participant; individuals must track all such changes, behaving differently as the community shifts in moral emphasis. Erring on the side of self-service will incur a reputational cost; erring on the side of group service will incur a sucker's cost.

In short, the dynamics of indirect reciprocity are expected to covary with the dynamics of group stability because (1) individuals benefit by assessing others in different ways depending on group stability, and (2) as these assessments change, individuals will benefit by accommodating their own behavior to those changes.

4.3. Dynamic moral tension: an illustrative model

The dynamic moral tension hypothesized here can be illustrated as follows. Consider a boat race in which a number of multirower craft race against each other over a predetermined course. Prize money is divided such that the first boat receives 50% of the total, the second boat receives 50% of the remainder, and so on down the standings. Within each boat, the position of each rower dictates what fraction of the boat's total winnings are his, such that the person in the first seat gets 50% of the boat's total, the person in the second seat gets 50% of what remains, and so on down the boat. (Astute readers will recognize that the prize allotment scheme leaves a small sum unawarded, which we would argue is best spent on imaginary beer for all participants). The rules permit individuals, alone or in collaboration, to dislodge boat mates from superior seats, but, as a practical matter, this cannot be accomplished while rowing.

In such a race, one can easily imagine that, as boats fall behind, they will be overcome by a sense of shared fate and the need to cooperate intensely so that they will have something substantial to divide. As any boat pulls ahead, individuals in the back of that boat (who stand to gain little from the win) will reasonably conclude that bettering their own standing within the boat is the best use of their efforts. Cooperation between competitors in the back of the boat is likely to arise, and breakdown of such alliances is increasingly probable as they move closer to the front of the boat and more prize money is at stake.

4.4. When groups collapse

One exception to the trend of stability-dependent cooperation follows from the fact that humans, although highly dependent on their groups, are not absolutely so. When stability is so low that the group might be doomed to dissolution, group members may consider the benefits of leaving the group to be greater than the benefits of serving it. Moreover, as

individuals cease striving for the group when they believe that the cause is lost, they will be accelerating the collapse of the group, both by their withdrawal of aid and by the effects of that withdrawal on the assessments of others. This consideration indicates a threshold effect, with a sharp decline in cooperation and, thus, self-fulfilling group dissolution, once hopelessness of group persistence begins to spread. The existence of this tendency, however, depends on a perceived probability of successfully integrating into new groups following past group failures. Where there is no such hope, individuals would be expected to go down with the ship, continuing to employ the only strategy with any apparent chance of success.

4.5. Zero-sum vs. nonzero-sum dynamics

We propose that this model generates behavior readily recognizable to observers of human behavior. However, it only represents that portion of human endeavor that is characterized by conditions that are zero-sum or nearly so. Zero-sum conditions are those in which resource utilization is complete, thus, one individual can only benefit at a cost to others. Because humans, like all organisms, tend toward carrying capacity, near zero-sum dynamics as the typical state may be a plausible hypothesis. Nevertheless, the dynamics of moments in human history are worth considering when resource limitation temporarily decreases, to the point where a rapidly growing population can nevertheless become dominated by strategies that ramp up the extraction of resources from the environment. In such situations (e.g., dispersal onto a newly discovered landmass), intragroup competition, although still a functional strategy for individuals, may bear disproportionate opportunity cost because effort spent on infighting could be spent instead on resource extraction (thereby increasing the overall size of the pie), a strategy that, in these situations, is more profitable and likely to increase the future size of the group, thereby reducing future susceptibility of the group to intergroup threats. Despite the potential importance of nonzero-sum dynamics to human morality (Wright, 2000), the remainder of this paper will continue to focus on the more conventional zero-sum situation.

5. Morality as group stability insurance

5.1. Variation in commitment can endanger group stability

The principle of stability-dependent cooperation predicts an inverse correlation between the stability of groups and the tendency of their members to adhere to moral rules. However, if the dynamics of human cooperation were this simple, there are at least two potential problems that, when significant, could cause group stability to deteriorate too quickly for individual behavior in service to the group to increase to counteract it.

First, factors influencing group stability can be extrinsic and therefore not under the control of the group. When factors like resource limitation and intergroup competition act quickly, group destabilization may be difficult to anticipate and prevent. If this is true, high group stability and its concomitant low levels of service to the group will lead to an increased vulnerability of the group to fast-acting extrinsic sources of group instability.

Second, positive and negative effects on group stability are asymmetrical. As with many organized structures, an individual has more power to affect group stability negatively than positively. Under circumstances favoring group stability, each cooperator restrains self-service for the sake of the group, but generally contributes to group stability only in a small way. However, if individuals jockey for position within the group, they can initiate a rapid decline in group stability, as the prospect of exploitation shifts everyone's adaptive strategy away from group service towards self-service. If moral rules are less important to people in times of group stability, and the usual restraints on within-group competition are relaxed, the opportunity would be created for individuals to compete to slightly exceed their neighbors' moral decay. An individual would attempt to gain the greatest possible benefits from the group's moral relaxation. The result would be a "race to the bottom," where the bottom is the breakdown of group-serving cooperation and the outright neglect of group stability.

Of course, the dependence of individuals on their group means that when the disastrous race began to threaten group stability, the interests of everyone would be served by reversing the trend and maintaining the group. However, in cases where everyone's interests are served by community action that is costly to each individual if unilaterally pursued, a tragedy of the commons results (Hardin, 1968). Everyone may continue to pursue actions that are beneficial to no one in the end, resulting in group destabilization. Indirect reciprocity is unlikely to be able to rescue a community from this situation. (Milinski et al., 2002, concluded otherwise, but the situation being described here is different from their experimental milieu. In actual societies, reputational costs and benefits may return too slowly to counteract the immediate benefits accruing to competitors in a race to the bottom).

Thus, in both classes of hazard—extrinsic threats as well as races to the bottom—the fast-acting nature of the changes is what is expected to jeopardize stability in human groups.

5.2. Morality buffers variation in commitment

Functional systems are often buffered against perturbations that threaten their integrity. For instance, in contrast to some proteins, such as MHC, that experience rapid evolution, histone proteins, which are critical to the stability of all eukaryotic genomes, have evolved an extremely low mutation rate (Kornberg & Lorch, 1999). If fast-acting or unpredictable forces can threaten the stability of human groups, and if susceptibility to these forces is affected by group members' adherence to moral rules, then a buffering system that lowers the risk from such threats to group stability may be in place in moral systems. Therefore, the second principle we propose is the group stability insurance function of morality, whereby certain features of human morality are adaptive primarily because they buffer group stability against fast-acting threats. The two major features of morality that may be serving this buffering function are absolutism and viscosity.

5.3. The function of moral absolutism

Individuals in groups where moral rules appear to be constant and always deserving of adherence should be less likely to discard or neglect those rules than will individuals in groups

where rules have no such absolutist qualities. We hypothesize, then, that the air of absolutism surrounding moral rules has been maintained in human culture because it buffers the changes in attitude and behavior that would be engendered by stability-dependent cooperation. In particular, absolutism works against the natural slippage of adherence to moral rules that occurs during times of group stability, decreasing the susceptibility of the group to sudden extrinsic threats and heading off the tendency for a rapid competitive race to the bottom. Moral groups, on this hypothesis, insure their stability by "consecrating" their rules in the minds of their members, just as political theorist Edmund Burke suggested political groups do (Burke, 1790, par. 159–164). Hence, the adaptive dynamics of social groups provide the basis for an explanation of how humans benefit by associating their moral rules with the most sacred and authoritative aspects of their culture, despite facultative adherence to these rules. We are not aware that any other hypothesis addresses this apparent paradox.

5.4. The function of moral viscosity

The second feature of the adaptive buffering system that we propose to be in place in human moral systems is viscosity with regard to moral rules. Viscosity in this sense is suggested by the old notion of moral *character*, the quality of individuals that makes them significantly influenced by habit and slow to change their attitudes and behavior patterns once developed (Kohlberg, 1964). If humans are resistant to change and susceptible to entrainment or habit formation in morality, they will be less likely to engage in rapid changes of commitment level that can compromise the efficacy of indirect reciprocity and ultimately threaten group stability. They will also be less likely to track drastic fluctuations in perceived group stability, decreasing susceptibility to sudden extrinsic threats.

Available psychological evidence does suggest that moral attitudes are viscous in this sense (Eisenberg et al., 2002; Eysenck & Eysenck, 1963; Kagan, 1989). In fact, this appears to be a relatively old feature of the human psyche that functions in a variety of other contexts and, hence, is not unique to the moral sphere. Nevertheless, if the above considerations are correct, moral character or viscosity, together with moral absolutism, can be explained biologically as a system providing insurance of group stability. The system buffers the impact of threats to group stability at the level of individual adherence to moral norms.

6. Status and power inequality as a modifier of within-group moral variation

Discussion to this point, while not assuming egalitarianism within groups, has not dealt with the variation in moral commitment that results from inequalities of power and status. At an authoritarian extreme, the effect of power disparity will swamp the effects of group stability, for the ability of most individuals to modify their level of service to the group will be very limited. In general, predictions like those at the end of Section 4.1 are more applicable the more freedom that individuals have to make behavioral decisions and are best tested on behaviors that are not legislated or coerced, except by community expectations. Even when individuals do have such freedom, status inequalities still probably modify the expected dynamics.

One set of examples of such complexity relates to the adaptive strategies of people in positions of power. Like other people, the powerful will benefit when group stability is high, but they also have a special stake in promoting group service, both because they get disproportionate shares of the profits of collaboration and because competition is much more likely to move them down the intragroup hierarchy than to move them up. Two tactics that are therefore likely to be employed by powerful members of stable groups are misinforming the group by understating group stability (perhaps by manufacturing or exaggerating threats) and enforcing group service through penalties. In the terms of the boat race model presented in Section 4.3, those seated in the fronts of boats, and perhaps, especially in the front of leading boats, will tend to exaggerate the threat from competing boats because they have little to gain and everything to lose from intraboat competition. Moreover, in certain cases, such as in democratic groups, where the continued tenure of leaders depends on perceptions of their having done a good job, leaders may gain from preaching the high stability of the group, at least relative to when they were not leaders. These considerations illustrate the importance of distinguishing between actual and perceived group stability in predicting the optimal degree of group service. Such status-by-stability interactions also become important when stability is dangerously low. At some low threshold of group stability, group members may do better by leaving, but leaders will benefit by keeping others in the group, providing an incentive to misinform group members that the group is more stable than it is.

7. The multiplicity of social groups

This discussion has portrayed moral deliberation as unidimensional, from self- to group service. In fact, the moral landscape is more complicated because individuals belong simultaneously to different groups. Some may overlap, such as an ethnic group and a workers' union, and some are concentric, such as a neighborhood within a city within a state within a nation. The rule of stability-dependent cooperation can certainly be dealt with on the simple continuum of self- to group service, but the actual decisions faced by individuals may often be a matter of how large a group to align with in a particular situation (the self being one end of that continuum), or which of two group memberships to prioritize (Mason, 1996). Analysis focused on one group identity may misinterpret prioritization of another as defection towards self-service.

8. Relation to two other perspectives on the evolution of morality

8.1. Cultural evolution of memes

The hypotheses introduced above treat facultative adherence to morality as a trait that is adaptive. This contrasts with the view that the persistence of cultural elements, or "memes" (Dawkins, 1976), is unrelated to individual reproductive success. In the view of Dawkins and others, memes need not increase the bearer's inclusive fitness to persist; rather, they evolve and

adapt to each other in an autonomous system. This viewpoint surely has some utility in the short term, and individual cultural elements may subvert the interests of others and of the genes. However, the perspective that has led to the principles of stability-dependent cooperation and the group stability insurance function of morality is based on the assumption that radical separation of culture from individual fitness cannot commonly be the case in the long run.

Because the capacity for culture is genetic in evolutionary (and developmental) origin, to have become fixed in our species the system must have returned benefits to the genomes of its bearers throughout the period of its elaboration. The net average fitness effect of all memes, genetically speaking, must therefore have been positive. At whatever point the net effect of all memes on genes becomes negative, natural selection should disassemble the genetic capacity for memes. Recognizing this, most will acknowledge that memes must have been fitness enhancers early in their evolution, but some will claim that the system has more recently become autonomous, and memes need only be neutral in fitness impact (on average) to persist, independent of genetic interests. But because all behaviors take time and effort, and therefore have an opportunity cost, memes that are not beneficial are expected to be short lived. When cultural elements are widespread and persistent, they are likely to have become so because they tended to benefit their bearers.

8.2. Group selection

The hypotheses developed here clearly depend on the interests of groups of individuals. Nevertheless, we have not relied on the group selection perspective (where natural selection is discussed at various levels, including the social group, as if selection at multiple levels reflects multiple evolutionary mechanisms). In fact, even proponents of multilevel selection admit that selection at various levels can be reduced to a single mechanism (Sober & Wilson, 1998). An individual-level perspective in evolutionary discussions of human groups therefore has the advantage of discussing the complexity of natural selection, without an apparent proliferation of evolutionary mechanisms. Moreover, as mentioned in Section 7, humans are members of various social groups that are not always concentric, as the notion of a "level" of selection would imply. Finally, our individual-level perspective also refrains from making assumptions about the degree to which groups replicate or otherwise resemble genomes (Williams, 1966). Nevertheless, the hierarchical relationship between the competitive success of the group and that of the individual remains explicit in these hypotheses and the perspective underlying them. Individual fitness is considered overwhelmingly dependent on group persistence, and groups are assumed to vary in their capacity to persist. From this perspective, selection on the individual results in the individual's capacity to prioritize among several avenues of potential fitness return.

9. Conclusion

We have drawn implications from the understanding that humans, as social animals who are nevertheless genetically individualistic, must strike a balance between strategies for

competition within a group and strategies for increasing group stability. This assessment of the human situation follows from the evolutionary theory of human culture and morality developed by Alexander (1987, 1990, 1992, 2004).

In particular, we have proposed that much observed variation in commitment to moral norms is explained by a rule of stability-dependent cooperation, where the adaptive level of individual commitment is a function of the stability of the social group. If this were true (and the predictions from this hypothesis are numerous), variability in human moral commitment reflects our ability to track variation in the expected benefits of competition versus cooperation. However, groups in this situation would still be susceptible to fast-acting extrinsic threats, as well as self-destructive competitive races to the bottom. In light of this, we have proposed that the absolutism and unchangeableness that people attribute to moral norms, features that have bewildered moral philosophers for centuries (Williams, 1985), might function as group stability insurance against fast-acting threats.

These hypotheses deserve testing for at least three reasons. First, they explain why morality has an air of absolutism despite the facultative nature of human commitment to moral rules. Second, they resolve the longstanding debate about the adaptive status of moral rules by placing the locus of adaptation not in particular kinds of acts, but in the moral agent's ability to weigh options and choose a commitment strategy based on the current social environment. Third, they connect moral attitudes to environmental variables and thus have the potential to explain hitherto perplexing moral variation within and between individuals and cultures.

References

Alexander, R. D. (1979). *Darwinism and human affairs*. Seattle, WA: University of Washington Press.

Alexander, R. D. (1987). *The biology of moral systems*. Hawthorne, NY: Aldine de Gruyter.

Alexander, R. D. (1989). Evolution of the human psyche. In P. Mellars, & C. Stringer (Eds.), *The human revolution* (pp. 455–513). Edinburgh: University of Edinburgh Press.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Darwin, C. (1871). *The descent of man and selection in relation to race*, (1874 ed). London: John Murray.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Edwards, C. P. (1975). Societal complexity and moral development: a Kenyan study. *Ethos*, *3–4*, 505–527.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q., & Carlo, G. (2002). Prosocial development in early adulthood: a longitudinal study. *Journal of Personality and Social Psychology*, *82*, 993–1006.

Eysenck, S. B. G., & Eysenck, H. J. (1963). The validity of questionnaire and rating assessments of extraversion and neuroticism, and their factorial stability. *British Journal of Psychology*, *54*, 51–62.

Frank, S. A. (2003). Perspective: repression of competition and the evolution of cooperation. *Evolution*, *57*, 693–705.

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

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*, 1243–1248.

Hartup, W. W., & van Lieshout, F. M. (1995). Personality development in social context. *Annual Review of Psychology*, *46*, 655–688.

Huxley, T. H. (1894). Evolution and ethics. *Evolution and ethics* (pp. 46–86). London: Macmillan.

Kagan, J. (1989). *Unstable ideas: temperament, cognition, and self*. Cambridge, MA: Harvard University Press.

Keith, A. (1949). *A new theory of human evolution* (pp. 383–432). New York: Philosophy Library.

Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman, & L. W. Hoffman (Eds.), *Review of Child Development Research*, *vol. 1*. (pp. 383–432). New York: Russell Sage Foundation.

Kornberg, R. D., & Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, *98*, 285–294.

Lotem, A., Fishman, M. A., & Stone, L. (1999). Evolution of cooperation between individuals. *Nature*, *400*, 226–227.

Lydon, J. E. (1990). Commitment in the face of adversity: a value affirmation approach. *Journal of Personality and Social Psychology*, *58*, 1040–1047.

Lyon, M. F. (2003). Transmission ratio distortion in mice. *Annual Review of Genetics*, *37*, 393–408.

Mackie, J. L. (1981). *Ethics: inventing right and wrong* ch. 1. Harmondsworth: Penguin.

Maienschein, J., & Ruse, M. (Eds.) (1999). *Biology and the foundations of ethics*. New York: Cambridge University Press.

Mason, H. E. (1996). *Moral dilemmas and moral theory*. New York: Oxford University Press.

Milinski, M., Semmann, D., & Krambeck, H. -J. (2002). Reputation helps solve the "tragedy of the commons". *Nature*, *415*, 424–426.

Nesse, R. (2001). Natural selection and the capacity for subjective commitment. In R. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 1–47). New York: Russell Sage Press.

Nisan, M., & Kohlberg, L. (1982). Universality and variation in moral judgment: a longitudinal and cross-sectional study in Turkey. *Child Development*, *53*, 865–876.

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

O'Neill, P., & Petrinovich, L. (1998). A preliminary cross-cultural study of moral intuitions. *Evolution and Human Behavior*, *19*, 349–367.

Peters, K. E. (2003). Pluralism and ambivalence in the evolution of morality. *Zygon*, *38*, 333–354.

Richards, R. J. (1987). *Darwin and the emergence of evolutionary theories of mind and behavior*. Chicago: Chicago University Press.

Ridley, M. (1996). *The origins of virtue*. London: Penguin.

Roes, F. L., & Raymond, M. (2003). Belief in moralizing gods. *Evolution and Human Behavior*, *24*, 126–135.

Ruse, M. (1986). *Taking Darwin seriously: a naturalistic approach to philosophy*. Oxford: Blackwell.

Sober, E., & Wilson, D. S. (1998). *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.
Williams, B. (1985). *Ethics and the limits of philosophy.* London: Fontana.
Williams, G. C. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.
Williams, G. C. (1988). Huxley's evolution and ethics in sociobiological perspective. In P. Thompson (Ed.), *Issues in evolutionary ethics* (pp. 317–349). Albany, NY: SUNY Press.
Wright, R. (2000). *Nonzero: the logic of human destiny.* New York: Vintage Books.