THE EVOLUTION OF MORAL PROGRESS

A BIOCULTURAL THEORY

ALLEN BUCHANAN
AND
RUSSELL POWELL

OXFORD
UNIVERSITY PRESS

CHAPTER 4

Is Evolved Human Nature an Obstacle to Moral Progress?

The Introduction showed why it is important to resurrect the all-but-buried topic of moral progress and to restore its pride of place in contemporary liberal political theory. Part I took the first step toward reviving the victim of premature burial. This chapter confronts a potentially powerful conservative challenge to one especially important type of moral progress that is the central focus of this book—the growth of inclusivist morality. As we noted earlier, inclusivist moralities are those that reject group-based (e.g., race-, ethnicity-, nationality-, or species-based) restrictions on moral standing and moral status, as well as the notion that moral standing is to be attributed to "outsiders" only in virtue of self-serving strategic considerations.

The conservative challenge to the liberal faith in inclusivist moral progress that we discuss in this chapter rests on four pillars: the first is that human nature shapes the possibility-space of moral progress; the second is that human nature, if it makes sense to talk of such a thing at all, is not a fixed, timeless essence but rather a product of evolution; the third is that our evolved nature, at least so far as it includes our capacity for morality, heavily favors exclusivist (or "tribalistic") moralities over inclusivist ones; the fourth is that this evolved disposition toward exclusivist morality is highly recalcitrant to cultural modification.

The first pillar is not new. For centuries, and long before the Darwinian revolution in biology, conservative thinkers have held that our capacity for being moral and hence for moral progress is shaped, indeed seriously constrained, by human nature. And they have taken a rather pessimistic if not unflattering view of what our nature is and of our prospects for acting in morally progressive ways, typically emphasizing the dominance of passions over reason and selfishness over concern for the common good.

The second and third pillars, in contrast, are something new: they represent an attempt to enlist modern evolutionary science in the service of conservative thinking about the prospects for moral progress, at least so far as inclusiveness is concerned. This is not to say that all or even most thinkers who believe that our evolved morality is strongly anti-inclusivist are conservatives. So far as evolutionary moral psychologists are operating as scientists, they are merely characterizing the way they think human moral capacities are configured, without purporting to draw any moral or political philosophical lessons. But as will become clear in a moment, other thinkers have attempted to draw normative conservative conclusions from scientists' characterization of evolved human moral nature.

The lack of scientific backing for their rather dark characterization of human nature has always been the Achilles heel of traditional conservatism. What traditional conservatives have had to say about human nature and about the nature of society has often been a matter of empirically under-informed speculation or cherry-picking from the annals of human history, rather than the fruit of empirically informed scientific reasoning. This is not surprising, of course, given that through most of the history of conservative thought, indeed until very recently, little genuinely scientific knowledge of human nature and society was available. Conservatives can now at last tout a scientific basis for their view of human nature, one that can provide a more solid grounding for their pessimistic conclusions about the possibilities of moral progress. Evolutionary theory, the contemporary

conservative can proudly say, tells us that evolved human nature is a serious obstacle to moral progress, especially in inclusivist form. We call those who hold this modernized conservative view "evoconservatives" to distinguish them from traditional conservatives. Our attempt to revive thinking about moral progress has thus taken an unexpected turn: we must now consider whether the advent of evolutionary moral psychology can revitalize a conservative tradition according to which the scope of any plausible theory of moral progress must inevitably be quite modest.

The Adaptive Function of Morality: The Received View

What is the purported evolutionary function of morality, and on what evidential foundation does it rest? Before considering the received adaptationist explanation of morality, it is important to be clear about the *explanandum* (the phenomenon to be explained). "Morality" in the relevant evolutionary literature includes both social and individual dimensions of normative thought and behavior: it refers, broadly, to a social commitment to preference-independent norms, modulated by other-directed and inward-directed moral emotions and judgments and typically enforced through institutionalized sanctions. Painting a finer-grained picture would involve filling in specific moral content, such as a sense of fairness, prohibitions against particular behaviors, conceptions of virtues, specific punitive reactions to norm violations, and so on.

Why think that morality at any level of description might be amenable to evolutionary explanation? One reason is that moral systems are spatiotemporally ubiquitous in human societies. Moral rules structure the behavior of all known hunter–gatherer bands, nomadic tribes, sedentary agricultural populations, and modern, post-industrial people on all habitable continents and across all ecological niches and modes of subsistence. Another reason is that morality as a functional kind is likely very old: moral

systems are presumed to have been in place at least since the origins of behaviorally modern humans in the upper Paleolithic and possibly much earlier as evidenced by high levels of cooperative foraging and coordinated warfare in the paleoanthropological record; such phenomena are hard to explain without postulating norms that underpin social cooperation and coordination. Third, although moral systems vary considerably, they exhibit significant commonalities in form and content.[1] Taken together, such patterns cry out for a selectionist explanation. This is because in any system with variation and heredity, including biological and cultural systems, the spatiotemporal ubiquity of some complex set of co-occurring features is indicative of adaptation or some other stabilizing constraint.

Furthermore, moral systems present as "adaptively configured" so as to foster cooperative social arrangements, producing a functional match to coordination problems that is incredibly unlikely to arise through chance processes alone—that is, in the absence of selection (or, more technically, the non-random sampling of competing variants). Just as it is unlikely that the length of a pollinating moth's proboscis just happens to match the size of the trumpet-shaped spur of the orchid from which it typically extracts nectar, so too is it unlikely that moral systems just happen to solve complex social coordination problems without having been through the filter of natural selection.

The basic logic of selectionist explanation is simple. If a population varies in some heritable trait and if such trait variations have differential effects on the probability of the survival and

[1] Donald Brown, *Human Universals* (McGraw Hill, 1991); Richard Joyce, *The Evolution of Morality* (MIT Press, 2006); but see Jesse Prinz, "Is Morality Innate?" in W. Sinnott-Armstrong (ed.), *Moral Psychology*, v. 1 (MIT Press, 2008, pp. 367–406). Cross-cultural universality should not be taken, in itself, to imply "innateness." Cooking, for example, is a ubiquitous human trait that appears to have dramatically shaped human morphological and social evolution, even though it is culturally acquired. See Richard Wrangham, *Catching Fire: How Cooking Made Us Human* (Basic Books, 2009).

reproduction of organisms that possess them, then there will be evolution by natural selection—"descent with modification," to use Darwin's phrase. Traits that are produced in this manner are known as "adaptations," and the mechanistic process that produces them is known as "adaptation."[2] In essence, selection pressures generated by the interaction of organismic traits and the fitness-relevant features of their environment act as a filter: traits (and their associated developmental generators, which are often, but not always, genes) that reduce biological "fitness" (expected reproductive success) will tend to not get passed on in sufficient numbers to determine the character of future populations. This is because the individuals who carry these relatively less fit variants die before they can reproduce, have fewer offspring, or have offspring that do not survive long enough to reproduce.

It is vital here to emphasize that to say that something is an "adaptation" is a strictly backward-looking statement—it is a claim about the selective etiology of a trait, not about its present utility or current contribution to survival and reproduction. Thus, to the extent that morality is associated with reproductive costs in the modern environment, this does little to undermine the selective-etiological claim that significant aspects of morality are adaptations.

However, even if some trait clearly presents as an adaptation, this does not mean that we can easily identify what the trait is an adaptation for. For example, the array of dorsal plates on the iconic dinosaur *Stegosaurus* looks like an adaptation, but there is little agreement as to its particular functions. Did the stegosaur's bony plates serve as a defensive bulwark against carnivorous dinosaurs, as a mechanism of thermoregulation, or as a mode of signaling to conspecifics and mates? We may never know the proper function of stegosaur plates because the crucial etiological information may be forever lost to the depths of geological time.

[2] See Robert Brandon, *Adaptation and Environment* (Princeton University Press, 1990).

In contrast, investigations of the adaptive function of morality are at once more promising and more challenging than the study of stegosaur plates. It is more promising because morality has a comparably recent origin (geologically speaking), and unlike stegosaur plates, it can be studied in living human beings at various stages of development in a wide range of societal contexts, some of which approximate the ancestral state of human societies in which morality first evolved. It is more challenging because morality is a social-psychological trait that is much harder to delineate than simple morphological features and must be inferred (rather than directly observed) in the fossil record.

Despite these epistemic challenges, a plausible empirical case for the specific adaptive functions of morality, on a certain coarse-grained description of the trait, has begun to emerge. The received view among evolutionary theorists who believe that human morality can be given a specific selectionist explanation goes roughly like this. Morality developed and spread among small, scattered hunter–gatherer groups in the middle to late Pleistocene, where it was selected for coordinating social behavior and managing patterns of interaction that resulted in costly intragroup conflicts. In particular, morality helped solve collective action problems by reducing free-riding, enabling individuals to resist temptations to act selfishly, and preventing dominant individuals from monopolizing the fruits of cooperation—thereby generating an evolutionary return that was greater for each individual than would have been possible if each had acted alone or as part of a group that did not cooperate effectively.[3] The fruits of increased

[3] Michael Tomasello, *A Natural History of Human Morality* (Harvard University Press, 2016); Chris Boehm, *Moral Origins: The Evolution of Virtue, Altruism, and Shame* (Basic Books, 2012); Chris Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard University Press, 2001); Jonathan Haidt, *The Righteous Mind* (Pantheon, 2012); Kim Sterelny, *The Evolved Apprentice: How Evolution Made Humans Unique* (MIT Press, 2012); Philip Kitcher, *The Ethical Project* (Harvard University Press, 2011); Joyce, *The Evolution of Morality*, supra note 1; Robert Wright, *Nonzero: The Logic of Human Destiny* (Pantheon, 2000); Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Harvard University Press, 1992).

social coordination and cooperation included (inter alia) higher foraging yields, enhanced warfare capabilities, territorial acquisition, the efficient management of common resources, and the resolution of internal disputes. Ethnographic research has established that the morality of hunter–gatherer societies, which is widely regarded as the ancestral state of human morality, is ubiquitously anti-hierarchical and that violations of so-called egalitarian norms—especially attempts to monopolize resources or to exercise authority over fellow group members (except very temporarily, as when one individual is selected to lead a war party)—are met with forceful sanctions, ranging from social ridicule to ostracism to execution.[4]

What explains the evolutionary shift from a distinctively chimp-like social life dominated by hierarchy and self-interest to a distinctively human society sustained by stable altruism and robust egalitarian moral norms? Although chimpanzees do engage in minimally cooperative behaviors, such as in monkey hunts, in raids on other chimp groups, and in internal struggles for dominance, this cooperation is generally fragile, easily disrupted by temptation, and for the most part instrumentally driven.[5] Why are human hunter–gatherer bands far more cooperative and egalitarian than chimp groups, and what role did this novel social structure play in human ecology?

A number of contemporary evolutionary theorists have converged on the hypothesis that cooperative foraging was the key "ecological design problem" that prompted the evolution of the egalitarian ethos in humans.[6] Though somewhat speculative, the empirically constrained hypothesis is as follows: early in

[4] Boehm, *Hierarchy in the Forest*, supra note 3, pp. 81–82.

[5] Felix Warneken and Michael Tomasello (2006), "Altruistic helping in human infants and young chimpanzees," *Science* 311: 1301–1303.

[6] See, e.g., Sterelny, *Evolved Apprentice*, supra note 3; Boehm, *Hierarchy in the Forest*, supra note 3.

human evolution (~400,000 years ago), there was a shift to hunting large dangerous quarry, particularly during frequent periods of glaciation when edible plants and small game animals were scarce. Such large game included extremely dangerous animals like mammoths, extinct giant buffaloes, extinct giant baboons, hippopotamuses, and the like. For 98 percent of human history, this intensively cooperative feat was accomplished with rudimentary stone-tipped wooden spears and other non-projectile weapons. This required not only meta-cognitive capacities such as shared intentionality (or "plural agency")[7] that were presumably lacking in the last common ancestor of humans and chimpanzees but also sophisticated normative mechanisms for underwriting the equitable distribution of the spoils once the fruits of cooperation were realized. If any single dominant individual were (in standard chimp style) to dominate the spoils of the hunt, others would refrain from cooperating in future hunts. The evolution of an egalitarian ethos (at least among hunters)—including the institutional enforcement of equitable distribution—ensured that the spoils of cooperation were divided evenly and that all who participated would benefit from the hunt. By reducing human tendencies to act selfishly and hierarchically, morality made ultra-cooperation in distantly related individuals possible.

Various evolutionary theoretical accounts have been offered to explain stable cooperation in moderate-sized non-kin groups, such as reciprocal altruism, indirect (reputation-based) reciprocity, and punishment-reinforced cooperation.[8] There is, however, reasonably broad agreement on the basic Darwinian logic: in a population of competing cultural groups subject to the climatic

[7] Tomasello, *A Natural History of Human Morality*, supra note 3.

[8] See, respectively, Robert Trivers (1971), "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46(1): 35–57; Richard Alexander, *The Biology of Moral Systems* (De Gruyter, 1987); and Robert Boyd, Herbert Gintis, Samuel Bowles, and Peter Richerson (2003), "The Evolution of Altruistic Punishment," *Proceedings of the National Academy of Sciences USA* 100(6): 3531–3535.

upheavals of the late Pleistocene, those that developed effective moralities, that is moralities that were capable of avoiding the costs associated with cooperation failures, were more likely to pump hominins into the next generation, to persist as groups, to sustain and transmit their social structures, and/or to give rise to offspring groups.[9] These ecological conditions, so the argument goes, conferred a reasonably high probability on the evolution of morality in broad strokes and go some way toward explaining its more specific contours, such as our evaluative attitudes toward kin, kith, strangers, patriots, non-reciprocators, gluttons, cheats, murderers, and the like.

[9] Samir Okasha and Peter Godfrey-Smith discuss several ways in which group-level selection might be cashed out. See Samir Okasha, *Evolution and the Units of Selection* (Oxford University Press, 2006); Peter Godfrey-Smith, *Darwinian Populations and Natural Selection* (Oxford University Press, 2009). There is continued controversy over the level at which selection must operate in order to stabilize cooperative interactions among non-kin. A growing chorus of biologists, anthropologists, and philosophers of science now argue that robust cooperation in moderate-sized groups of non-kin is only likely to evolve through a process of selection at the group level, given the costs of altruism and norm enforcement to individual fitness within groups. See Haidt, *Righteous Mind*, supra note 3; David Sloan Wilson and Edward O. Wilson (2007), "Rethinking the Theoretical Foundation of Sociobiology," *Quarterly Review of Biology* 82(4): 327–348; Boyd et al., "The Evolution of Altruistic Punishment," supra note 8; Samuel Bowles (2009), "Did Warfare Among Ancestral Hunter–Gatherers Affect the Evolution of Human Social Behaviors?" *Science* 324(5932): 1293–1298; Samuel Bowles (2008), "Conflict: Altruism's Midwife," *Nature* 456: 326–327; Boehm, *Hierarchy in the Forest*, supra note 3; Elliott Sober and David Sloan Wilson, *Unto Others* (Harvard University Press, 1999); for a partially dissenting view, see Sterelny, *Evolved Apprentice*, supra note 3. For the present purposes, it does not matter whether selection for moral traits can be cashed out at the level of individuals in a group-structured population or at the level of cultural groups proper since in either case a selectionist explanation would be vindicated. We will not consider evolutionary explanations of morality at the level of cultural variants themselves (e.g., so-called memetic theories) since the received selectionist explanation conceives of moral traits as parts of the individual or group phenotype, rather than as units of selection in their own right. Quite apart from their widely discussed conceptual and methodological problems, memetic theories have no clear implications for constraints on the space of moral (and hence moral progress) possibility, and thus we will not address them here.

It is important to emphasize a point about the dynamic nature of adaptation that is often lost in discussions of the evolutionary function of morality. Organisms do not simply adapt to pre-existing ecological niches, much as keys are molded to fit locks. Rather, organisms and their selective environments are co-determinative, in the sense that a lineage's adaptive moves shape the very ecological design problems that it needs to solve.[10] For instance, the evolution of altruism generates a selection pressure for cheaters who can effectively parasitize the evolutionary generosity of altruists, which then results in selection pressures for cheating detection, which in turn results in selection for subtle cheaters, and thus selection for the detection of subtle cheating, and so on. The point is that adaptation is a dynamic, open-ended process, so we should not think of morality as a stable evolutionary key to the fixed ecological lock of cooperation. We will return to the dynamic nature of adaptation in Chapter 7, where we explore the ways in which culturally engineered social environments interact with evolved components of moral psychology to drive moral progress and moral regression.

The Darker Side of Morality

Focusing on the prosocial effects of prehistoric morality can obscure its darker side. Ethnographic work, behavioral studies, and mathematical models of cultural evolution indicate that the development of egalitarian and other altruistic moral norms in moderately sized groups of distantly related individuals whose reputations are harder to monitor hinges on institutionalized moralizing punishment[11]; and the evolution of third-party

[10] See Richard Lewontin (1978), "Adaptation," *Scientific American* 239: 156–169.

[11] Sarah Mathew and Robert Boyd (2011), "Punishment Sustains Large-Scale Cooperation in Prestate Warfare," *Proceedings of the National Academy of Sciences USA* 108(28): 11375–11380; Joseph Henrich et al. (2006), "Costly Punishment Across Human Societies," *Science* 312(5781): 1767–1770; Boyd et al., "The Evolution of Altruistic Punishment," supra note 8; Boehm, *Hierarchy in the Forest*, supra note 3.

punishment in large groups of non-kin appears to pose a higher-order altruism problem that only group-level selection can solve since punishing is often costly to the punishers. Group-level selection, in turn, is only sufficiently strong in the context of frequent and frequently lethal intergroup conflict,[12] where losing groups are extinguished and the individuals composing them are killed, dispersed, absorbed by winning groups, or marginalized to resource-poor areas.

Thus, the high frequency of mortal conflict between prehistoric human groups is a central assumption—and empirical conclusion—of the multilevel selection modeling work on the evolution of altruism. The logical structure of this inference runs as follows: moral norms underpinning cooperation are not sufficiently adhered to in the absence of punishment due to the invasion of free-riding strategies; all known human societies have institutions of punishment that enforce moral norms underpinning cooperation; the evolution of punishment requires sufficiently strong group selection; group selection is only sufficiently strong in the context of frequent lethal intergroup conflict; thus, we can conclude that human life in the late Pleistocene involved frequent antagonistic intergroup interactions.

The idea is that groups that contained more altruists and moralizing punishers, and consequently more cooperative social structures, tended to outperform and "replace" groups with less effective moralities in economic and military contests between groups.[13] Economic advantages of moral groups included higher

[12] Bowles, "Did Warfare Among Ancestral Hunter–Gatherers," supra note 9; Boyd et al., "The Evolution of Altruistic Punishment," supra note 8; Robert Boyd and Peter Richerson (2002), "Group Beneficial Norms Can Spread Rapidly in a Structured Population," *Journal of Theoretical Biology* 215: 287–296.

[13] Sober and Wilson, *Unto Others*, supra note 9. Although punishment may not be necessary for group selection to stabilize cooperative behaviors that do not implicate altruism (the stag hunt game may offer such an example), it

foraging yields (increased success in hunting large game), which in turn supported larger group sizes. Moralities also enhanced warfare capabilities since better cooperation means better coordination in military conflicts and larger group sizes confer a significant advantage in raiding, border skirmishes, and full-scale military conflicts, with victorious groups populating the territories and commandeering the resources of vanquished groups. In addition, moral systems provided more effective dispute resolution, helping to make sure that internal conflicts did not cause the group to dissolve or leave it vulnerable to predation by other groups. Notice that the foraging benefit—the ability to cooperate in hunting large dangerous game—and the warfare benefit—the ability to coordinate military actions against other groups—implicate not only overlapping psychological capacities (such as shared intentionality and anti-free-riding and egalitarian sentiments) but also the ability to develop complex technologies, to improve upon them, and to transmit these manufacturing skill sets faithfully down the generations.

As Kim Sterelny persuasively argues, moral norms likely underpinned the institutions responsible for sustaining and transmitting crucial technological crafts, methods of food preparation, and natural history information in hunter–gatherer bands. Such a scenario would have provided fertile conditions for Darwinian selection to occur in the meta-population of culturally and moral psychologically variable hunter–gatherer bands.

This "how possibly" explanation of the evolution of morality is supported by several converging lines of interdisciplinary research. Although none of them is in itself decisive, taken together they make a strong circumstantial case for the key role

is likely that only the targeted severity of punishment can exert an influence on the payoff matrix sufficient to sustain large-scale participation in warfare, norm enforcement, and other forms of cooperation that are group-beneficial but individually costly and hence vulnerable to free-riding. See Peter Richerson and Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (University of Chicago Press, 2005, pp. 220–225).

of intergroup competition in prehistoric human ecology. First, examinations of the ethnographic, archeological, and evolutionary anthropological records attest to the prehistorical ubiquity of intergroup conflict in hunter–gatherer bands.[14] This conclusion is perhaps not all that surprising: the scarcity of resources during the climatic upheavals of the Pleistocene, which would have triggered competitive intergroup interactions (see Chapter 7), combined with the ultra-coordinated hunting capacities and weapons-making industries of humans during that same time, would have been ripe conditions for intergroup conflict. Although the record of intergroup conflict in the very late Pleistocene and early Holocene is well established, there is still no "smoking gun" of warfare that dates back to the time period in which human morality is thought to have emerged. There are, for example, no cave paintings from the late Pleistocene depicting warfare among bands (though depictions of human forms are in themselves extremely rare and typically schematic), nor are there any fossilized hominids from this period with spear points embedded in their remains (though the human fossil record of this period remains spotty).

Further, the inferences we are entitled to make about prehistoric human societies and behaviors from observations of modern hunter–gatherer bands—even ones that are organizationally similar to those that existed during the upper Paleolithic—are somewhat limited since living hunter–gatherer lifeways are not necessarily reliable traces of the prehistoric human past. Evidence that chimpanzees (*Pan troglodytes*), the phylogenetically closest living taxon to *Homo sapiens*, regularly engage in violent intergroup conflicts could suggest that the tendency toward

[14] Steven Pinker, *The Better Angels of Our Nature* (Viking, 2011); Bowles, "Did Warfare Among Ancestral Hunter–Gatherers," supra note 9; Boehm, *Hierarchy in the Forest*, supra note 3; Richard Wrangham and Dale Peterson, *Demonic Males: Apes and the Origins of Human Violence* (Houghton Mifflin, 1996); Lawrence Keeley, *War Before Civilization* (Oxford University Press, 1996).

intergroup antagonism was transmitted to humans and chimps from a common ancestor or alternatively that it was arrived at in parallel in *Homo* and *Pan* through adaptation to similar ecological regimes (but see the discussion of bonobos in the next chapter).[15] However, it is unclear whether the chimp "power imbalance" model of intergroup aggression, whether it is grounded in homology or parallelism, can usefully be applied to Pleistocene humans who had different modes of subsistence and weapons that could kill at a distance.[16]

Sterelny suggests that frequent intergroup conflict is unlikely to occur in persistence predators, such as *Homo*.[17] A wide range of evidence indicates that humans are specifically adapted for persistence hunting: pursuing faster prey for extended periods through endurance running and tracking, until the prey becomes exhausted and can be speared at close range.[18] Persistence hunting was probably the dominant mode of big game hunting for humans until the very late invention of projectiles (such as the bow and arrow) and the domestication of horses and dogs. However, other persistence hunters, such as wolves and spotted hyenas, also engage in violent and risky intergroup conflict, so antagonistic behaviors are not inconsistent with this specialized mode of predation. Population genetics also offers somewhat equivocal answers to the question of prehistoric conflict: inferences about human population sizes from comparative genomic data indicate a population holding steady throughout the late Pleistocene—data that are consistent with high levels of intergroup competition[19]

[15] R. W. Wrangham and L. Glowacki (2012), "Intergroup Aggression in Chimpanzees and War in Nomadic Hunter–Gatherers: Evaluating the Chimpanzee Model. *Human Nature* 23: 5–29.

[16] Raymond Kelly (2005), "The Evolution of Lethal Intergroup Violence," *Proceedings of the National Academy of Sciences USA* 102(43): 15295–15298.

[17] Sterelny, *Evolved Apprentice*, supra note 3.

[18] D. M. Bramble, and D. E. Lieberman (2004), "Endurance Running and the Evolution of *Homo*," *Nature* 432: 345–352.

[19] See Samuel Bowles and Herbert Gintis, *A Cooperative Species. Human Reciprocity and Its Evolution* (Princeton University Press, 2011).

but that, as Sterelny points out, can also be explained by extrinsic environmental variables suppressing human population growth.

In his seminal book *War Before Civilization*, Lawrence Keeley's survey of pre-state warfare exploded the politically charged myth of peaceful human prehistory—what he describes as the "thrall of nostalgic delusion" that fueled degeneration theories of civilization that were popular in the academy (see Introduction) and which urged a return to the less hierarchical and allegedly peaceful ways of our hunter–gatherer past.[20] Upon re-examining the archeological and ethnographic records, Raymond Kelly argues that although Keeley is right that homicide and violence were rife in "unsegmented" pre-state societies during the late Pleistocene, intergroup conflicts were of a limited nature during this time. Warfare properly conceived, Kelly maintains, did not emerge until the agricultural revolution, which allowed for the emergence of complex segmented (roughly, differentiated and hierarchically structured) societies equipped with group identities.[21] In a paradoxical twist, and contra received social scientific and behavioral ecological wisdom, Kelly argues that it was not resource scarcity but rather *economic bounty* wrought by the agricultural revolution that created conditions ripe for warfare—since it was only under conditions of surplus "that a society can afford to have enemies for neighbours." Before that time, Kelly contends, warlike groups would have been selected against since warfare was not selectively advantageous. Kelly's theory hinges on the assumption that spears and other close-range weaponry nullified the power imbalances that drove intergroup conflicts in chimpanzees, effectively making warfare too risky in moderately sized human groups; it also assumes that the benefits of prosocial interactions would have almost always outweighed the fitness benefits of antagonistic relations between power-imbalanced groups—both questionable assumptions.

[20] Keeley, *War Before Civilization*, supra note 14.

[21] Kelly, "The Evolution of Lethal Intergroup Violence," supra note 16.

In short, existing ethnographic, archeological, and evolutionary anthropological data are somewhat equivocal on the issue of Pleistocene warfare; and researchers remain divided on the extent to which antagonistic intergroup interactions shaped the ecology of late Pleistocene humans.[22] The question is not so much whether human social evolution was shaped (at some point) by intergroup violence but, rather, how far back in human prehistory intergroup conflict extends and how central a role it played in the evolution of morality.

There is another line of evidence, however, that in our view indicates the centrality of human intergroup conflict in the upper Paleolithic: the impressions or traces of prehistoric ecological regimes left on modern human psychology. There is evidence that core elements of human moral psychology were forged in conflict between moderate to large ethnolinguistic groups.[23] "Parochial altruism," which consists in the combination of in-group favoritism/empathy and out-group antagonism/antipathy, is among the most cross-culturally robust features of human moral psychology and a direct prediction of group selectionist accounts of morality.[24] Ethnocentric bias—a cluster of perceptual, affective, and behavioral biases that favor in-groups over out-groups—emerges rapidly in very young children beginning

[22] M. Lahr et al. (2016), "Inter-Group Violence Among Early Holocene Hunter–Gatherers of West Turkana, Kenya," *Nature* 529(7586): 394–398.

[23] Melissa McDonald, Carlos Navarrete, and Mark Van Vugt (2012), "Evolution and the Psychology of Intergroup Conflict: The Male Warrior Hypothesis," *Philosophical Transactions of the Royal Society B* 367: 670–679; Mathew and Boyd, "Punishment Sustains Large-Scale Cooperation," supra note 11; Jung-Kyoo Choi and Samuel Bowles (2007), "The Coevolution of Parochial Altruism and War," *Science* 318: 636–640; Helen Bernhard, Urs Fischbacher, and Ernst Fehr (2006), "Parochial Altruism in Humans," *Nature* 442: 912–915; Richerson and Boyd, *Not by Genes Alone*, supra note 13.

[24] Choi and Bowles, "Coevolution of Parochial Altruism and War," supra note 23.

at around three years of age[25]; the expression of ethnocentric bias is cognitively automatic and does not require reward and punishment or explicit acculturation[26]; and as will be discussed in more detail in Chapter 7, there is also evidence that humans have innate tendencies to "essentialize" human groups and to automatically assign moral significance to group membership—which in turn serves to modulate empathy and altruism and thus interactions with other groups. The extent to which in-group and out-group biases are aspects of a single adaptive psychological system, or rather distinct traits that can be "toggled" independently in development and evolution, remains unclear.[27]

Although the "innateness" or "instinctual" nature of in-group/out-group bias has not been established beyond a reasonable doubt (for example, through a "poverty of the stimulus"–style argument), its pan-cultural nature, its rapid acquisition in ontogeny, its intertwining with empathy and altruism, and its consistency with predictions of evolutionary biological theory are at least strongly suggestive that it is genetically prespecified to some degree. This "innateness" conclusion is consistent with observations that ethnocentric bias is robustly scaffolded by culture, that it is shaped by moral norms, and that it is overrideable by executive function or acculturation. Even if in-group/out-group moral psychology is an adaptation to intergroup conflict, as seems highly plausible, this does not definitively pinpoint the relevant time frame of adaptation. It is possible that ethnocentric biases evolved through gene–culture co-evolution in the small

[25] Frances E. Aboud (2003), "The Formation of In-Group Favoritism and Out-Group Prejudice in Young Children: Are They Distinct Attitudes?" *Developmental Psychology* 39(1): 48–60.

[26] Jay J. Van Bavel, Dominic J. Packer, and William A. Cunningham (2008), "The Neural Substrates of In-Group Bias: A Functional Magnetic Resonance Imaging Investigation," *Psychological Science* 19(11): 1131–1139.

[27] Marilynn Brewer (1999), "The Psychology of Prejudice: Ingroup Love or Outgroup Hate?" *Journal of Social Issues* 55(3): 429–444; M. Hewstone, M. Rubin, and H. Willis (2002), "Intergroup Bias," *Annual Review of Psychology* 53:575–604.

evolutionary window that comprises postagriculture human existence—which, if true, would be consistent with Kelly's hypothesis regarding the post-Neolithic origins of war. Though this notion might run contrary to the "gradualism" presupposed by traditional evolutionary biological theory, there is increasing evidence that significant human genetic evolution has indeed occurred on this surprisingly short timescale. Nevertheless, given the pan-cultural distribution, reliable psychological development, and complex proximate neural mechanisms implicated in in-group/out-group biases, it seems more likely that ethnolinguistic bias arose much earlier in the human lineage, and hence that these adaptive psychological configurations contain information about—or traces of—human social ecology as it was in the deep past.

Sterelny is skeptical of Kelly's warless Pleistocene world, but he nonetheless argues, contra Bowles, Gintis, Boyd, Richerson, and their multilevel selection theorist collaborators, that "cooperation and altruism are the fuel of war, but not warfare's child."[28] Yet even if Sterelny is right that basic cooperative capacities predate intense intergroup conflict, perhaps originating as early as *erectus*-grade *Homo* (as inferred from levels of cooperative hunting in *Homo erectus*[29]), and even if, notwithstanding the modeling work alluded to above, punishment can evolve absent a group selection context, this is consistent with morality being co-opted and honed in co-evolution with intergroup conflict during the late Pleistocene. The apex predatory skills for hunting dangerous megafauna are readily transferrable to hunting dangerous weapons-wielding hominins. In short, lethal intergroup conflict may have arisen either subsequent to or directly in connection with the emergence of ultra-cooperation in humans; either way, human morality was selectively shaped—and, if group

[28] Sterelny, *Evolved Apprentice*, supra note 3, p. 190.

[29] Manuel Domínguez-Rodrigo (2002), "Hunting and Scavenging by Early Humans: The State of the Debate," *Journal of World Prehistory* 16(1): 1–54.

selectionist theorists are right, originally forged—in the crucible of intergroup conflict.

A striking feature of the received selectionist explanation, therefore, is that it implies morality is essentially an intragroup affair. The same ecological conditions and selection pressures that made moral traits adaptive would have imposed a fitness cost on extending "evolutionarily excessive" moral consideration to outgroup members. Just as free-riding on in-group members will tend to undermine group performance in a competitive intergroup arena, so too will excessive moral consideration toward members of the out-group. The selectively optimal combination appears to have been reasonably expansive moral consideration toward members of one's in-group (with a caveat for women and children, which we will return to later) and highly strategic—including predatory, antagonistic, and apathetic—behavior toward strangers, who were often distrusted, dehumanized, and delegitimized.[30]

In their groundbreaking theoretical defense of group selection, Elliott Sober and David Wilson take note of this implication for human moral psychology:

> It should be obvious . . . that multilevel selection theory does not lead to the fulfillment of a romantic vision of universal niceness.

[30] McDonald, Navarrete, and Van Vugt, "The Male Warrior Hypothesis," supra note 23; Carlos Navarrete and Daniel Fessler (2006), "Disease Avoidance and Ethnocentrism," *Evolution and Human Behavior* 27: 270–282; Pinker, *The Better Angels of Our Nature*, supra note 14. The claim is not that intraspecific aggression is always adaptive, as the costs of aggression will often outweigh its benefits, nor that cooperation between groups was never fitness-enhancing. Under certain conditions, intergroup hostility can lead to lost opportunity costs, such as the benefits of material trade and mate exchange that would have flowed from non-antagonistic interactions. Nevertheless, patterns of intergroup homicide in pre-state humans, as well as in common chimpanzees, indicate that intergroup predation often reaps evolutionary rewards; and this would have been particularly true for weapons-wielding hominins with the cognitive prowess to make case-by-case risk–benefit calculations.

> Conflict and competition are not eliminated but merely elevated in the biological hierarchy, where the problem of social dilemmas appears all over again at an even grander (and potentially more destructive) scale.[31]

If the prevailing group selectionist theory is right, then morality not only emerged and co-evolved in a Darwinian crucible of intergroup conflict but it also made large-scale human conflict possible by amplifying internal cooperation and by carving up the moral community and the scope of altruistic norms along in-group/out-group boundaries.

Still, it is important not to overstate the degree of conflict and the lack of cooperation between human groups in the environment of evolutionary adaptation. There is evidence of a significant degree of trading, exogamy, military alliances, and other forms of cultural exchange among even geographically distant cultural groups. Some local early human evolutionary environments may have been, for a number of reasons, more amenable to peaceful relations among groups than others; and evidence suggests that human moral psychology exhibited sufficient flexibility to allow them to take advantage of these conditions, perhaps in the form of minimally inclusivist moralities. This is our first serious indication that the evoconservative view that human moral nature is "hard-wired" for tribalistic morality is simplistic. Further, the hypothesis that we are hard-wired for exclusivist morality is in tension with the psychological findings noted above, namely that in-group favoritism does not automatically result in uniform out-group aggression and antagonism and that there is greater cultural variation in degrees of out-group antagonism than there is in degrees of in-group favoritism. We will return to the adaptive flexibility of prehistoric human morality in greater detail in Chapter 7, where we advance an alternative evolutionary model

[31] Sober and Wilson, *Unto Others*, supra note 9, p. 174.

of the development of human moral psychology. If the theory we advance is right, then it is misleading to say that human beings are "hard-wired" for exclusivity; it is more accurate to say that humans have an adaptively plastic capacity to develop either exclusivist moralities or inclusivist moralities, depending upon certain crucial features of the environment in which moralities develop and evolve.

Accordingly, we can restrict our delineation of the explanandum to the psychological and social mechanisms that dispose human beings to demarcate the moral community in particular ways—and this more fine-grained delineation allows for a meaningful adaptationist analysis of the trait. Nevertheless, to say that the above adaptationist account is the received *selectionist* explanation of morality is not to say that it is the received *explanation*. Some prominent moral psychologists and philosophers of science argue that allegiance to specifically moral norms is an evolutionary byproduct of adaptive tendencies toward norm compliance in general[32] or that certain moral norms are byproducts of moral emotions and nonmoral capacities.[33] To further complicate matters, when some theorists maintain that morality did not evolve, what they mean is that it did not evolve through *gene-based* selection (including, perhaps, gene–culture co-evolution), although they are open to the possibility that specific moralities could have been *culturally* selected for.

There is also the vexed conceptual problem of how to delineate properly moral norms from those typically thought of as social conventions. "Do not rape" appears to be a

[32] See, e.g., Edouard Machery and Ron Mallon, "The Evolution of Morality," in J. M. Doris (ed.), *The Moral Psychology Handbook* (Oxford University Press, 2010, pp. 3–46); Chandra Sripada and Stephen Stich, "A Framework for the Psychology of Norms," in P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Culture and Cognition* (Oxford University Press, 2006).

[33] Ibid.

qualitatively different sort of norm than "Use the small fork for salads," even if they both have normative force and provide reasons for acting or refraining from acting in particular ways. Theorists have proposed a number of ways in which the moral–conventional distinction might be drawn, including (1) the content of norms (e.g., moral norms are harm-based, whereas conventional norms do not implicate the interests of others), (2) the affective reactions produced by norm violations (e.g., guilt, anger, indignation, and perhaps disgust result from the violation of moral norms, whereas conventional norm violations provoke weaker or no emotional responses),[34] and (3) the subjective justification of norms (e.g., in the minds of moral agents, conventional norms are grounded in social practice, whereas moral norms are grounded in considerations that are authority/practice-independent).[35] In addition, studies of normal[36] and abnormal[37] moral psychological development have been interpreted as providing evidence that humans have specialized, innate moral faculties that are distinct from their generic normative capacities.

Many theorists remain skeptical, however, that the moral–conventional distinction can be sustained, given that pancultural studies have shown that conventional norm violations can also provoke powerful emotional and institutional responses and that harm-based moral judgments are sometimes and in some cultures viewed as authority-dependent.[38] It does

[34] Sean Nichols (2002), "Norms with Feeling: Towards a Psychological Account of Moral Judgment." *Cognition* 84: 221–236.

[35] Nicolas Southwood (2011), "The Moral/Conventional Distinction," *Mind* 120: 761–802.

[36] E. Turiel, *The Development of Social Knowledge: Morality and Convention* (Cambridge University Press, 1983).

[37] R. Blaire (1995), "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition* 57: 1–29.

[38] D. Kelly, S. Stich, K. J. Haley, S. J. Eng, and D. M. T. Fessler (2007), "Harm, Affect, and the Moral/Conventional Distinction," *Mind & Language* 22: 117–131; see also Machery and Mallon, "The Evolution of Morality," supra note 32.

not matter for present purposes whether there is a proper subset of norms that are distinctively moral; indeed, our working definition of morality, like that of most cultural evolutionary theorists, does not rely on the moral–conventional distinction. The key claim here is that the human capacity for norm acquisition and implementation (including motivations for adherence and enforcement)[39] is likely adaptive and was selected for its ability to coordinate action and support cooperation within groups, with specific norms culturally selected for these effects. If particular moral judgments stably and substantially contributed to cooperation (e.g., judgments with regard to in-group harm or free-riding), then we might expect gene–culture co-evolution to select for genetic factors that make the expression of those judgments more likely. Thus, even if the moral–conventional distinction is not vindicated, evaluative judgments often thought to be distinctively moral may play an especially important and culturally ubiquitous role in mitigating selfish tendencies, resolving potentially destabilizing intragroup conflicts, and motivating punishment.

Chapter 8 will revisit the diverse origins, functions, and effects of social norms. Our focus until then will be on the evolution of normativity in its especially weighty forms. Our aim is to evaluate neither the standard selectionist account nor its detractor theories. Instead, we will argue that even if a selectionist explanation of certain aspects of morality could be given along the lines sketched above, *whether it is grounded in cultural group selection or reciprocity or some combination of the two*, this would still leave much of contemporary morality beyond the scope of evolutionary explanation altogether. This, in turn, will show that morality is not constrained by evolution to the degree that evoconservatives and others might suppose.

[39] Sripada and Stich, "A Framework for the Psychology of Norms," supra note 32.

The Evoconservative Logic

It is important to recognize that many proponents of the standard selectionist explanation of morality do not subscribe to a conservative brand of politics, nor have they suggested that the evolutionary explanations they give, if vindicated, would have any conservative moral or political implications. Philip Kitcher, for instance, maintains that although morality has the evolutionary function of solving cooperation failures within groups, its emergence prompted an ongoing ethical discussion, which due to our deliberative faculties can go in any number of directions, including inclusivist ones.[40] Likewise, one upshot of Boyd and Richerson–type models of cultural evolution is that punishment can theoretically stabilize any norm, including more inclusive ones, regardless of whether it is group-beneficial.[41]

Nevertheless, authors from a variety of disciplines have inferred from the received selectionist explanation of morality that the content of human morality is seriously constrained—particularly in relation to the scope of other-regard. These evoconservatives contend that the ecological challenges our distant ancestors faced generated selection pressures for evaluative tendencies that limited effective moral commitments to members of one's own kin, group, tribe, or nation—and that these putative facts about human evolutionary history significantly constrain the shape of plausible moralities and the scope of other-regarding concern. This, in turn, is thought to suggest that cosmopolitan and other inclusivist moral principles are not appropriate or realistic for beings like us.

Stephen Asma, for instance, stresses the moral importance of tribal biases, arguing that moral emotions "cannot stretch indefinitely to cover the massive domain of strangers and nonhuman

[40] Personal communication; see also Chapter 1.

[41] Robert Boyd and Peter Richerson (1992), "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups," *Ethology and Sociobiology* 13: 171–195.

animals," given that our other-regarding dispositions were limited by evolutionary design to our "affective communities" of kith and kin.[42] U.S. appellate judge and legal theorist Richard Posner, in debates with moral philosopher and animal welfare proponent Peter Singer, defends species-based moral discriminations by appealing to similar evolutionary considerations.[43] International law theorists Jack Goldsmith and Eric Posner contend that it is a mistake to try to create an international legal order grounded in cosmopolitan moral principles because "we should not expect individual altruism to extend to people who are physically and culturally more distant"—and they argue that such biopsychological plausibility constraints on the moral obligations of individuals apply with equal force to institutions.[44] Francis Fukuyama, a prominent conservative bioethicist and political theorist, holds that political orders and social norms must be grounded in a substantive conception of human nature that pays heed to our evolved biases toward kin and in-group, as well as to the evolutionarily evidenced limitations of our capacity to sympathize with all human beings.[45] Leading psychologist Jonathan Haidt, who has stressed the moral psychological significance of in-group loyalty, expresses a related view:

> It would be nice to believe that we humans were designed to love everyone unconditionally. Nice, but rather unlikely from an evolutionary perspective. Parochial love—love within groups—amplified

[42] Stephen Asma, "The Myth of Universal Love," *New York Times*, January 5, 2013; Stephen Asma, *Against Fairness* (University of Chicago Press, 2012, pp. 45–46).

[43] Richard Posner and Peter Singer, "Animal Rights: A Debate," *Slate*, June 2001.

[44] Jack Goldsmith and Eric Posner, *The Limits of International Law* (Oxford University Press, 2005, p. 212).

[45] Francis Fukuyama, *Our Posthuman Future* (Farrar, Straus and Giroux, 2002, pp. 127–128).

by similarity, a sense of shared fate, and the suppression of free riders, may be the most we can accomplish.[46]

Whether evolutionary limits on love significantly constrain morality depends, of course, on the extent to which behaving morally toward others requires love. Although Haidt does not directly address this question, his statement occurs within the context of reflections on what we can expect by way of moral behavior, so it seems fair to interpret him as suggesting that the character of our evolved morality does not bode well for the possibility of inclusivist morality.

Larry Arnhart, a proponent of the "Darwinian right," goes further in arguing that not only does an evolutionary perspective on human nature bolster conservative views vis-à-vis the limitations of human altruism but "we can judge political regimes as better or worse depending on how well they satisfy the evolved desires of human nature."[47] Thus, evoconservatives believe that there are significant evolved psychological constraints on the shape of human morality, that these constraints are essentially fixed, and that they result in a scope of other-regard that is effectively restricted to in-groups.

The chief "improvement" of evoconservatism over traditional conservative philosophies is that it appeals to contemporary evolutionary psychology to ground its empirical claims about the moral limitations of human nature. Evoconservatives hold that the content of morality—in particular, the scope of moral duties and the class of beings who are recognized as having moral standing—is severely constrained due to evolutionary history. This in turn limits the set of social practices and institutions that are feasible. Highly inclusivist social arrangements, such as an international order exemplifying cosmopolitan principles of justice, would then be unattainable or at least unsustainable. The

[46] Haidt, *Righteous Mind*, supra note 3, p. 245.
[47] Larry Arnhart, *Darwinian Conservatism* (Imprint Academic, 2005, p. 84).

evoconservative lesson, then, is that attempts at moral reform that pay inadequate heed to evolved constraints on human other-regard not only are ultimately futile but also proceed at great peril since they are likely to destroy the value of existing moral practices and the institutions grounded in them.

There is a much weaker evoconservative claim that might be distinguished here. This weaker view holds that selectionist explanations of morality imply limited sympathy or feelings of positive regard for distant strangers but that this psychological claim in itself has no conservative political implications. That is to say, it acknowledges that humans may develop effective institutions and cultural practices that allow them to *treat* distant strangers as being worthy of moral consideration, even equal consideration, even if they are incapable of "loving" them (to use Haidt's words) or their compassion is attenuated under certain conditions, such as mass-scale humanitarian tragedies.[48] In other words, social practices and institutions may produce inclusivist morality, or a broadened range of what Sober and Wilson have called "behavioral altruism," without unlimited compassion or love. Some of the writers discussed above (including Haidt) are unclear as to whether they are only making the psychological claim or also making the mistake of assuming that if the psychological claim is true, then conservative moral or political conclusions follow. Some, including Posner and Goldsmith, clearly make the mistaken inference from the former to the latter. As we will see, to do so is to fail to appreciate how cultural developments, in particular institutions, can expand our capacities for behavioral altruism and shift human moral psychology in inclusivist directions.

The strong evolutionary constraints view has much more radical implications than those who endorse it acknowledge. If human morality is explainable according to the selectionist logic

[48] See D. Västfjäll, P. Slovic, M. Mayorga, and E. Peters (2014), "Compassion Fade: Affect and Charity Are Greatest for a Single Child in Need," *PLOS One* 9(6): 100115.

that evoconservatives endorse, then it is an understatement to say that inclusivist morality is a nonstarter. It implies that the scope of moral consideration *tout court* is very limited, not just the scope of *equal* basic moral consideration. In other words, it implies that is it implausible not only to expect people to regard all human beings as worthy of *equal* basic moral consideration but also to expect people to regard many human beings as worthy of *any moral consideration at all.*

As the quotes above indicate, there is an unsatisfying vagueness in the evoconservative stance. In fact, at least four evoconservative claims can be distinguished.

1. Any "morality" that is inclusive is practically ineffective and merely aspirational (because human moral emotions, such as sympathy or love, are "hard-wired" by evolution to be quite limited in their scope).
2. Inclusivist elements of morality, to the extent that they exist, are not durable (because the strong exclusivist, that is, intragroup, nature of human moral responses will inevitably undermine inclusivist developments).
3. The limits of inclusivist morality have already been reached or soon will be (because we are already at or near the end of the "evolutionary leash" on human culture).
4. Any effort to realize inclusivist ideals or norms will encounter serious resistance from the exclusivist tendencies that were selected for in the remote human past (even if the durability of such norms could be secured in theory).

We are sympathetic to the fourth evoconservative claim, albeit with certain important qualifications that we will elaborate in Chapter 7, where an alternative evolutionary model of moral psychological development is outlined. But we hasten to add that the fourth claim has no concrete practical implications for any particular inclusivist proposal for institutional reform or instance of individual moral development. It is one thing to say that those who wish to expand the moral circle should recognize that what they

propose may go against the evolutionary psychological grain; it is quite another to say that any particular move in the direction of greater inclusion is doomed to failure. If evoconservatism were restricted to claim 4, it would not be a very interesting view. It would be about as helpful as the warning to "proceed with caution" in developing new technologies: a trite admonition to be mindful of risk that supplies no specific guidance as to when risk is unacceptable or how to determine when risk is justified or whether it might be mitigated.

Consequently, we will focus on claims (1), (2), and (3). It is these assertions that make the evoconservative view interesting and which, if true, make it a serious threat to the project of developing a theory of moral progress that gives a prominent place to increases in inclusiveness. In the next chapter we provide a systematic critique of all three evoconservative claims, thereby clearing the way for a cogent naturalistic theory of moral progress along the dimension of inclusiveness.

Before doing so, however, it is important to bring to the fore certain common misconceptions about adaptation that might load the dice in favor of the strong evolutionary constraints view. As we noted earlier, modern evolutionary science rejects the Aristotelian notion that species' natures, if these exist in any meaningful sense, are fixed essences.[49] However, the concept of human nature operational in evolutionary moral psychology gets its explanatory purchase only insofar as it takes human nature to consist of a cluster of moral psychological traits that are highly "developmentally canalized," that is, robust across diverse cultural contexts. Just how developmentally robust putative

[49] For philosophical analyses of the scientific utility of the concept of human nature, compare David Hull (1986), "On Human Nature," *Proceedings of the Philosophy of Science Association* 2: 3–13, and Tim Lewens, "Human Nature: The Very Idea," *Philosophy & Technology* 25(4): 459–474, with Edouard Machery (2008), "A Plea for Human Nature," *Philosophical Psychology* 21: 321–329, and Grant Ramsey (2012), "Human Nature in a Post-Essentialist World," *Philosophy of Science* 80(5): 983–993.

elements of human moral nature are remains hotly contested. Nevertheless, it is important to recognize that even if one can establish that certain features of morality are adaptations, this does not in itself tell us how malleable these features are. Some plants, for example, have an adaptive propensity to grow tall in a crowded forest but wide in an open field. Likewise, the fact that some trait is an adaptation—the result of selection acting on heritable variation—does not imply that the trait is "innate" since cultural evolution can produce adaptations that are transmitted through mechanisms of social learning. By the same token, the fact that certain features of moral psychology are innate (that is, genetically prespecified) does not imply that they are unalterable or even difficult to modify through enculturation. The amenability of a given trait to environmental alteration is a separate contingent question from whether the trait is an adaptation, whether it is innate, and whether it has a genetic or cultural basis (or both).

The Received Evolutionary Account Supports a "Strategic" Conception of Morality

The foregoing discussion shows that the received selectionist explanation does not, in itself, make any explicit claims or license any strong inferences about constraints on the shape of human morality. However, it could be read to suggest, in line with the evoconservative inference, that the only sort of morality that humans are capable of engaging in, in any sustained and robust way, is what we referred to in Part I as *morality as cooperative group reciprocity*—or the strategic conception of morality.[50] Recall that according to morality as cooperative group reciprocity theories, moral standing is something that members of a cooperative group confer on one another—and only on one another. Individuals excluded from this reciprocal

[50] See David Gauthier, *Morals by Agreement* (Clarendon Press, 1989).

arrangement have no moral standing at all, and hence there are no moral duties constraining how out-group members should be treated. Moral standing is conferred only on individuals who can either disrupt or contribute to cooperation—that is, on the basis of "strategic capacities" relative to a cooperative scheme (though this need not involve explicit strategic calculations in every case). The strategic conception has a radical implication: it denies moral standing to individuals of other groups, and to individuals *within* the group, if they lack the ability to harm or benefit the group, as is the case with severely disabled individuals; and it may relegate individuals with limited strategic capacities to lower moral statuses.

The strategic conception of morality neatly accords with, and is arguably central to, evolutionary theories of morality. It is not surprising, therefore, that evolutionary theorists have explicitly linked the selectionist account to a strategic, prudence-based theory of morality, such as that of David Gauthier.[51] Prominent evolutionary theorists have argued that to understand morality, one must view societies as populations of individuals seeking their own self-interest.[52] Even evolutionary theorists who acknowledge that contemporary human morality is not confined solely to morality as cooperative group reciprocity still view reciprocity relations and social coordination functions as dominating contemporary moral behavior.[53] If they are right, then the possibilities for moral progress in the form of inclusivity are severely limited because robustly inclusive moralities do not make moral standing or status depend upon cooperative group membership or the capacity to reciprocate.

[51] For example, Kim Sterelny and Ben Fraser (2017). "Evolution and Moral Realism," *The British Journal for the Philosophy of Science* 68(4): 981–1006.

[52] Alexander, *Biology of Moral Systems*, supra note 8, p. 3.

[53] See, e.g., Joyce, *The Evolution of Morality*, supra note 1, chapter 4; Gibbard, *Wise Choices, Apt Feelings*, supra note 3.

Why Accounts of Moral Progress That Simply Appeal to Reasoning Are Inadequate

As we have seen, evoconservatives infer from prevailing evolutionary accounts of morality that inclusivist moralities are not psychologically feasible for beings like us—and that morally progressive institutional reforms, such as an international order reflecting cosmopolitan principles, are unrealistically utopian.[54] Importantly, liberal proponents of moral progress have done little to block this inference. In fact, some moral philosophers, whom we dub "evoliberals," have tacitly affirmed and reinforced the evoconservative view. Evoliberals argue that if there is to be significant progress in dealing with serious problems now facing humanity, it will be necessary to undertake biomedical interventions that enhance human moral capacities in order to remove or at least relax evolved constraints on human moral nature.[55] The assumption here is that our evolved psychology is so morally feeble, and in particular that the human capacity for other-regard is so limited, that the radical step of altering its biological basis may be morally required. We will critique the evoliberal view in the final chapter of this volume. Our point here is simply that some liberal thinkers who discuss moral progress have accepted key evoconservative assumptions.

Other recent discussions of moral progress, which tend to rely heavily on the efficacy of moral reasoning, have done little to deflect the evoconservative and evoliberal challenges to cultural moral reform. Though improvements in moral reasoning are part of the story (see Part III), focusing solely on reason ultimately proves inadequate to the task of explaining how inclusivist moral progress is possible, given the kinds of evolved beings that we

[54] Goldsmith and Posner, *Limits of International Law*, supra note 44.

[55] See Ingmar Persson and Julian Savulescu (2012), "Moral Enhancement, Freedom and the God Machine," *Monist* 95(3): 399–421; Ingmar Persson and Julian Savulescu, *Unfit for the Future: The Need for Moral Enhancement* (Oxford University Press, 2011).

are. As we will show later, these *explanatory* deficits translate into *normative* deficits: an inadequate understanding of how moral progress comes about supplies incomplete guidance for how to sustain and achieve more of it.

Eminent evolutionists of the nineteenth and twentieth centuries, such as Darwin, Huxley, and Simpson, believed that the capacity for reason enabled human beings to escape their base biological natures. This view is echoed, albeit in more sophisticated forms, by some contemporary moral and political philosophers. For example, Peter Singer, whose view was discussed in detail in Chapter 1, attributes the move toward greater inclusion to the human capacity to reason to moral truths even when doing so is not conducive to evolutionary fitness.[56] We agree with Singer and his collaborators that the capacity for reasoning will be an important part of any adequate explanation of inclusivist moral progress. More specifically, such an explanation will assign a key role to what we earlier called "open-ended normativity": the capacity to make explicit the norms one has hitherto been following and subject them to rational criticism and revision. One way this happens is that critical reflection leads to the recognition that existing norms are being applied inconsistently or are arbitrarily restricted in their scope, which in turn provides reasons to revise them.[57] However, simply saying (as Singer does) that inclusivist

[56] K. D. Lazari-Radek and Peter Singer (2012), "The Objectivity of Ethics and the Unity of Practical Reason," *Ethics* 123: 9–31; Peter Singer, *The Expanding Circle: Ethics, Evolution, and Moral Progress* (Princeton University Press, 2011). Enlightenment thinkers (including the French encyclopedists) and nineteenth-century liberals who advanced doctrines of progress also exhibited a rather naive faith in the efficacy of reason, without fully appreciating the need to consider the particular social and other environmental factors that modulate its exercise. John Stuart Mill, for example, appears to have put too much faith in the efficacy of freedom of expression under conditions of widespread literacy, not sufficiently appreciating the ways in which cultural forces (e.g., media) and normal cognitive biases can interact to produce and disseminate false beliefs.

[57] For a pathbreaking analysis of how individuals come to realize that they hold inconsistent views about the morality of particular behaviors and how they come to resolve these inconsistencies in morally progressive ways, see

moral commitments are a product of reasoning is not a sufficient basis on which to naturalize moral progress, for three reasons.

First, reasoning is much older than the emergence of inclusivist morality, so reasoning alone is not sufficient for the emergence of inclusivist morality. Again, this is not to deny that reasoning is a crucial component of inclusivist trends—it is, rather, to say that the operation of reasoning *of the right sort* and *on a sufficiently large scale* in the moral realm has only occurred under some conditions, and Singer's account is incomplete because it fails to consider or spell out these conditions. To that extent, his account has limited value both for explaining how moral progress has occurred and for understanding how to sustain it.

Second, at present the penetrance of inclusivist morality is quite uneven, with different human beings and different human cultures exhibiting inclusivist commitments to a greater or lesser degree; and yet there is no reason to believe that these interpersonal and intercultural differences in penetrance are the result of populational differences in the general capacity for reasoning.

Third, the exercise of human reason can sometimes contribute to expansions of the moral circle, but in other cases it plays a significant role in contracting the circle in ways that lead to moral regression. This occurs, for example, when people judge that some human beings do not count morally on the basis of false premises about natural differences between groups of humans (such as blacks and whites, men and women, heterosexuals and homosexuals) or due to mistaken ideas about which characteristics qualify one for equal moral status or for moral standing more generally. It also occurs when reasoning is used in an ad hoc or confabulatory way to justify pre-existing moral judgments that are motivated by negative affects like disgust, fear, or distrust.[58]

Richmond Campbell and Victor Kumar (2012), "Moral Reasoning on the Ground," *Ethics* 122(2): 273–312.

[58] Jonathan Haidt (2001), "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108: 814–834.

Thus, while Singer is right to say that reasoning plays an important role in the development of more inclusive moral commitments, he lacks an account of the developmental conditions under which the right sort of reasoning occurs and is likely to become sufficiently pervasive to result in large-scale moral progress notwithstanding resistance from evolved exclusivist tendencies. He also lacks an account of why the capacity for reasoning has often failed to be exercised in such a way as to achieve greater inclusiveness.

Just as reasoning functions differently in different social environments, so too does the capacity for self-scrutiny that open-ended normativity requires. Chapter 1 discussed Michele Moody-Adams's argument that a precondition for moral progress is that individuals be in a position to scrutinize their own values and the social practices that implement them: "One of the most important tasks of constructive moral inquiry," she argues, "is to . . . break down the common human resistance to self-scrutiny."[59] However, she does not specify or even indicate the conditions under which this breakdown in the resistance to self-scrutiny is likely to occur or under which self-scrutiny diminishes; and thus, her explanation of moral progress, like Singer's appeal to reason, is incomplete.

Richmond Campbell and Victor Kumar advance a much more detailed, illuminating, and empirically grounded account of one kind of reasoning that can result in moral progress: the identification and resolution of inconsistent moral responses, which they argue occurs through the interaction of intuitive and deliberative systems that guide moral judgment.[60] But like Singer and Moody-Adams, they do not explain the conditions under which this is

[59] Michele Moody-Adams (1999), "The Idea of Moral Progress," *Metaphilosophy* 30(3): 168–185, p. 175.

[60] Campbell and Kumar, "Moral Reasoning on the Ground," supra note 57. They argue that in contrast to deductive reasoning from principles, moral inconsistency reasoning is a dedicated moral system that emanates from distinct cognitive-affective pathways (p. 296).

likely to occur, so their view, too, is insufficiently naturalized. Nor do they provide an account of why people, and even entire cultures, can fail to identify moral inconsistencies or to work through them rationally but instead employ various cognitive-affective strategies to minimize the dissonance that arises from exposure to logically contradictory information without revising their pre-existing beliefs. As an instance of the latter, all too common phenomenon, consider the following case. In the racist culture of the American South, it was commonly thought that blacks were of inferior intelligence. When confronted with a black person who clearly showed high intelligence, there was a ready strategy for resolving the contradiction—not by discarding or reducing confidence in the generalization about the intelligence of blacks but by explaining away the apparently disconfirming case: it was said that the black person in question "must have some White blood."[61] This was a gross failure of moral consistency reasoning among people who possessed normal capacities for reasoning.

Contemporary philosophical theories of how moral progress occurs, which tend to lean heavily on the efficacy of moral reasoning, are inadequate—not because moral reasoning is unimportant in driving moral progress (it is very important) but because such theories fail to identify the circumstances under which moral reasoning is likely to contribute to inclusivist moral progress. More importantly, such theories have not provided an empirically grounded account of the general conditions under which moral progress is likely to occur and to be sustainable. As a result, they fail to rebuff conservative charges that certain types of moral progress are not realistic for beings like us. Moreover,

[61] This example is drawn from the personal experience of one of the authors, but this sort of cognitive dissonance resolution, which accounts for the resilience of false beliefs about out-group individuals, is familiar to those acquainted with the racist culture of the American South as it existed well into the twentieth century.

they fail to take seriously enough the conservative claim that moral reasoning in itself has limited causal efficacy in driving moral progress. We have in mind a more optimistic picture of the role of moral reasoning than conservative thinkers tend to paint. Nevertheless, to focus on moral reasoning while neglecting the biological and social conditions under which moral reasoning can flourish, as liberal theorists have been wont to do, is to gloss over explanatory components that are crucial to any genuinely naturalistic theory of moral progress.

In sum, a naturalistic account of how inclusive moral commitments emerge despite the evolved parochiality of human moral emotions, judgments, and norms cannot simply appeal solely to capacities for reason or self-scrutiny or to strategic self-serving relations between groups. It must identify the conditions under which reasoning capacities are exercised in such a way as to foster inclusivist commitments even in the absence of strategic motivations.[62] This, in turn, requires understanding how the capacity for cultural innovations can create moral developmental environments in which valid moral reasoning and self-scrutiny can flourish—and how this cultural scaffolding can be dismantled in ways that lead to moral regression.

To be fair, current philosophical theories of inclusivist moral progress do not purport to offer complete explanations—and, like them, we make no pretentions of doing so. Our aim, however, is not merely to plug explanatory gaps in existing theories of moral progress by providing supplementary empirical details. Rather, it is to sketch a model of inclusivist moral progress that not only is consistent with but affirmatively draws upon and

[62] One might assert that inclusivist morality is a dispositional trait that humans have long (or always) possessed but that the conditions necessary for its expression only manifested quite recently in human history. Even if one were content to describe such highly flexible, nonselected behaviors as conditionally expressed traits (which we are not), our point is that any naturalized account of inclusivist moral progress would need to identify the difference-making conditions for their expression.

unifies current biological and social scientific understandings of moral psychology and culture.

This chapter has articulated both the evoconservative view and the evolutionary account of the origins of human morality on which it is premised. The next chapter argues that evolutionary explanations of morality are limited in certain crucial respects that make the pessimistic inferences that evoconservatives draw from it invalid. More specifically, it argues that the received evolutionary explanation of morality cannot account for robustly inclusivist features of contemporary human morality and that this "inclusivist anomaly" indicates that the strong evolutionary moral constraints view is mistaken.

CHAPTER 5

The Inclusivist Anomaly and the Limits of Evolutionary Explanation

One major flaw in the evoconservative appeal to evolutionary theory is that contemporary morality, as experienced and exhibited by significant numbers of people and embodied in social practices and institutions, is strikingly more inclusive than one would expect if selectionist explanations were the whole story, or even most of it. In other words, from a selectionist perspective, inclusivity is highly anomalous. This chapter will first highlight four aspects of this inclusivity, drawing upon empirical evidence that strongly suggests that inclusivist morality is not a rare, exceptional, or merely academic phenomenon. It will then show that none of these aspects can be explained by the received selectionist account of the origins of morality or by alternative evolutionary accounts.

The Inclusivist Anomaly

The first feature of contemporary human morality that is anomalous from the standpoint of the received evolutionary account of morality is that significant numbers of people now regard at least some non-human animals as proper subjects of moral consideration; that is, they believe that there are moral constraints on how we are to treat animals, constraints that do not derive