THE EVOLUTION OF MORAL PROGRESS

A BIOCULTURAL THEORY

ALLEN BUCHANAN
AND
RUSSELL POWELL

OXFORD
UNIVERSITY PRESS

or in the design of institutions. We cheerfully acknowledge that evolved psychological capacities, interacting with particular social and institutional environments, can pose serious obstacles to using our rationality in ways that result in more inclusive moralities. Indeed, the next chapter offers a model that explains in detail why environments that mirror conditions of the EEA—such as those characterized by great physical insecurity, high parasite threat, severe intergroup competition for resources, and a lack of institutions for peaceful, mutually beneficial cooperation—will tend to be very unfriendly to the development of inclusivist morality.

Evolutionary explanations of morality can thus help to explain why inclusivist attitudes both were a long time coming and remain imperfectly realized today. At the same time, however, this chapter has offered compelling reasons, both theoretical and empirical, to believe that human morality is only weakly constrained by human evolutionary history, leaving the potential for substantial moral progress open. Our point is not that human beings have slipped the "leash" of evolution but rather that the leash is far longer than evoconservatives and even many evolutionary psychologists have acknowledged—and no one is in a position at present to know just how elastic it will turn out to be.

CHAPTER 6

Toward a Naturalistic Theory of Inclusivist Moral Progress

Chapters 4 and 5 argued that evolved human nature is not as formidable an obstacle to moral progress as evoconservatives have thought. Yet evoconservatives do paint a picture of human morality that challenges traditional liberal accounts of moral progress. In particular, they suggest that moral progress in the form of inclusivist morality faces formidable psychological and cultural hurdles, rooted in our evolved nature. This chapter outlines an evolutionary developmental model of inclusivist moral progress that calls into question the seemingly uncontroversial but ultimately misleading assertion that inclusivist morality goes against the human evolutionary psychological grain *tout court* or, as is also sometimes said, that we are "hard-wired" for exclusivist, tribalistic morality. On the account of moral psychological development that we advance, *evolved human nature is both an obstacle to moral progress and an enabler of it, depending upon the environment and the degree to which it resembles certain conditions that were prevalent in the environment of evolutionary adaptation (EEA).*

If our model withstands scrutiny, it will also enable us more confidently to reject another evoconservative/evoliberal claim upon which we cast doubt in Chapter 4: the assertion that although humans are capable of some degree of moral inclusion,

they are now bumping up against the limits of this capacity. To begin to determine whether the limits of inclusivity have been reached, or what their outer bounds might be, it is first necessary to do what the standard evolutionary explanation does not do: provide an account of how inclusivist morality could have developed from, and notwithstanding the constraints of, exclusivist morality. It bears noting that even the most illuminating scientific explanations of human thought and behavior rarely if ever identify a full set of sufficient conditions for some large-scale sociocultural outcome. Accordingly, our aim here is simply to identify important necessary conditions for inclusivist moral progress.

After outlining our evolutionary model of moral psychological development, we will go on to advance three further hypotheses (H1–H3) that draw on and extend this model:

> (H1) Inclusivist morality is a luxury good in the sense that it is only likely to be widespread and stable in highly favorable conditions—namely, those in which the harsh environmental conditions of the EEA have been overcome.
>
> (H2) Inclusivist gains can be eroded if these harsh conditions reappear or if significant numbers of people come to *believe* that they exist.
>
> (H3) A combination of normal cognitive biases and defective social-epistemic practices can cause people wrongly to believe that such harsh conditions exist, especially if there are individuals in positions of power and prestige who have an interest in spreading this false belief.

An Evolutionary Developmental Model of Moral Inclusivity

Our evolutionary explanation of how human beings whose moral capacities were shaped in the EEA could come to have inclusivist moral responses begins, oddly enough, with a much simpler model system: the water flea. Some species of water flea

develop elaborate armor, including defensive spines and helmets, but only if they detect the chemical signatures of predators in the water in which they develop. The development of armor, however, comes at a cost, including reduced locomotion and added energy requirements. As a result, water fleas have evolved a mechanism of conditional expression that enables them to develop armor only when they find themselves confronted with the high probability of a predator-rich environment. Such traits are known in evolutionary biology as "adaptively plastic" traits. The benefit of adaptive plasticity is that it enables a lineage to achieve a better adaptive match across more environments than would be possible if it produced a single phenotype in all environments. Adaptively plastic traits allow organisms conditionally to express alternative character states, depending on which state is most appropriate for the environment at hand. The organism accomplishes this by detecting environmental cues during its development that indicate which character state is ecologically appropriate and then triggering the development of that state.

Our central hypothesis is that exclusivist morality is like flea armor—the result of an adaptively plastic "toggle" that is keyed in to cues of out-group threat that are detected in the environment in which individuals and cultures develop and evolve together. More precisely, exclusivist moral response is a conditionally expressed trait that develops only when cues that were in the past reliably correlated with out-group predation, exploitation, competition for resources, and disease transmission are detected. In the animal world, the adaptively plastic detection of a predation threat can involve not only the detection of pertinent chemical cues, as with the water flea, but also more cognitively sophisticated inspection of predatory types, motivations, and behaviors.

Because humans are linguistic and robustly cultural creatures, the detection of out-group threat can also involve the social transmission of beliefs about out-groups. This can take the form

of explicit and implicit beliefs that individuals come to hold as a result of a combination of personal experience and cultural inculcation through "testimony" broadly understood. The cultural transmission of beliefs about out-groups has the advantage of avoiding the risks of trial-and-error learning but also the disadvantage of increasing the chance of faulty detections. We will return to this important feature of our evolutionary model of moral development in the next chapter, where we emphasize that culture enables the boundaries between groups to be radically redrawn so that, for example, groups *within* societies can become subject to exclusivist moral responses.

As with flea armor, the development of exclusivist moral tendencies has costs. In particular, out-group aggression, antipathy, and distrust—features strongly associated with exclusivist morality—reduce the chances of mutually beneficial interactions with neighboring groups, such as trade, mate exchange, and alliances, and increase the chances of dangerous, belligerent, mutually destructive interactions with foreigners. Because of this evolutionary trade-off, exclusivist tendencies will, according to the adaptive plasticity hypothesis, be tempered in environments in which out-group threats are not detected during development or in which they are counterbalanced by opportunities for cooperation with out-groups. This is not to say that moral developmental environments in which out-group threats are diminished are automatically conducive to deep forms of moral inclusion. To the contrary, there is every reason to think that attitudes toward out-groups would, prehistorically and historically, have been governed by strategic self-interest, rather than genuinely subject-centered considerations. Additionally, in-group favoritism appears to be evolutionarily primitive and hence less culturally and situationally variable than out-group antagonism; and in-group biases (in terms of empathy, trust, cooperative tendencies, etc.) result in very significant forms of discrimination against out-group members even where they do not

translate into active out-group hostility or derogation.[1] Thus, even if moral developmental environments are conducive to prosocial interactions between groups, this does not mean that these interactions will necessarily be guided by robustly inclusivist moral commitments of the sort that characterize recent expansions of the moral community. Moreover, as social psychologist Marilynn Brewer notes, the very fact of in-group/out-group social differentiation creates fertile grounds for intergroup antagonism and conflict: there is a fine line between the absence of trust and active distrust or between a lack of cooperation and active competition.[2] The point, however, is that adaptive moral plasticity makes positive intergroup relations possible, not that it makes them inevitable.

For such an adaptively plastic moral psychological mechanism to have evolved, there must have been reliable periodic selection pressures generated by both exclusivist-friendly and inclusivist-friendly ecological regimes. This picture is supported by research on Pleistocene technology which suggests that long-distance trade, and thus rudimentary markets, predated even the existence of language in hominids. Rigid groupishness or extreme preferences for kin would have made it difficult or impossible to participate in these bartering systems and therefore to reap the fitness-enhancing benefits they conferred. In addition to the trade of material culture, out-marrying and military alliances, which are fairly commonplace activities in hunter–gatherers, require cooperative relationships with out-groups that would have been difficult or impossible to achieve if exclusivist morality were rigidly "hard-wired." In an environment in which opportunities for cooperation with out-groups arose with some regularity,

[1] Gordon Allport, *The Nature of Prejudice* (Cambridge University Press, 1954); 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.

[2] Brewer, "The Psychology of Prejudice," supra note 1, p. 435.

human groups that could appropriately "toggle" (within and across generations) between exclusivist and inclusivist responses based on environmental cues would have had a fitness advantage over groups that were capable of only exclusivist responses.

In using the limited analogy of a toggle, we do not mean to suggest that moral exclusivity and inclusivity are discrete character states that can be switched on and off. To the contrary, they are clearly continuous rather than binary features. That is to say, moralities can be more or less inclusive, and they may be inclusive in some dimensions while being exclusive in others, as a result of the complex interaction of biological, psychological, and cultural forces shaping moral development and evolution. Still, one can speak meaningfully of more and less inclusive moralities and of a relatively exclusive morality coming to be more inclusive and vice versa. Given the spectrum of inclusivity/exclusivity, one might prefer to think of the adaptively plastic moral psychological mechanism we are contemplating here as less like water flea armor and more like adaptively plastic plant growth that allows plants to adjust to angles of sunlight. That is, one may prefer to think of moral exclusivity as less like a toggle and more like a dial that can be adjusted to fit local circumstance. In either case, the thrust of the theory and its ethical implications remain the same.

In fact, our rejection of the evoconservative's pessimistic conclusions about the possibilities for inclusivist moral progress depends only on the thesis that humans possess a flexible capacity for moral response, one that allows for inclusivist responses under certain conditions; it does not depend on the thesis that this capacity is itself an *adaptation*. In other words, even if we are wrong in surmising that the EEA included inclusivist-friendly conditions that were sufficiently pervasive and persistent to create stable selection pressures for inclusivist response, that is consistent with the capacity for inclusiveness being compatible with our evolved nature.

Suppose that very early in the EEA, through the mechanism of genetic mutation, some human beings acquired a

nondiscriminating capacity for what we would now call "prosocial" behavior, a disposition to cooperate with any human being they encountered. This cooperative response might simply be a propensity to reciprocate, to engage in what game theorists call a tit-for-tat strategy, or it might be a propensity for psychological altruism, a disposition to aid others even where there is no prospect of reciprocation. If either sort of cooperative disposition arose in an environment in which human beings existed only in very small groups—more specifically, in families of only one or two generations—then there would have been no reproductive penalty for such a "promiscuous" capacity for cooperation, so long as individuals did not encounter human beings from other groups with which their own group would have to compete for vital resources. In such an environment, any human being one encountered would be highly likely to be kin, and thus an undifferentiating cooperative or altruistic response would work as an effective kin selectionist heuristic. Under these conditions, a "promiscuous" disposition toward cooperative behavior would not reduce an individual's fitness or the fitness of a small group of which she was a member. If early human beings were relatively solitary, existing only in very small kin groups, then so long as they continued in that condition, a "promiscuous" disposition to cooperate would have conferred fitness advantages, even if it would have been disastrous under different conditions in which groups were larger or encounters with other groups were common.

If, however, this situation changed—if groups increased in size and hence intragroup genetic relatedness became more attenuated or if contacts between groups increased—then a promiscuous cooperative or altruistic response would become fitness-reducing, and we would (*ceteris paribus*) expect there to be selection for the emergence of a less "promiscuous," that is, more discriminating disposition to exhibit altruism or cooperation only toward members of one's own group and to adopt (at best) a cautious reciprocation orientation toward members of other groups. As

contact between competing groups increased, groups that developed cultures that sharply distinguished between "us" and "them" would have thrived, and those that were unable to restrict the originally promiscuous response would have been driven to extinction. An implication of this scenario is that the parochiality of human altruism is the result of cultural factors that rein in or demarcate the boundaries of an originally promiscuous inclusivist response, rather than stemming from a conserved "innate" biological disposition toward groupishness that interacts with local cultural systems to generate a more or less discriminating moral response. If this view is correct, then it gives further strong reasons to reject the notion that humans are "hard-wired" for moral exclusivity.

It is notoriously difficult to make reliable inferences about natural history—in this case, about the social ecological conditions of early humans—particularly when we have no direct evidence to consult. There remains some question as to whether the social world of early humans was more like that of the relatively solitary orangutans or rather more like the highly social and group-structured societies of chimps and bonobos. How might such an inference be made? Biologists often rely on a model lineage to infer the presence or absence of some trait in a target lineage.[3] One natural thing to do would be to compare *Homo sapiens* to its closest living "sister taxa"—namely, the lineage that includes chimpanzees and bonobos—and then infer that any trait shared between humans and chimpanzees/bonobos was present in and transmitted continuously from their most recent common ancestor. Such phylogenetic inferences are based on the principle that a trait present in two closely related existing species (e.g., extant chimpanzees and humans) can be inferred to have existed in and been faithfully transmitted from their most recent common ancestor. This hypothesis postulates fewer causes, or

[3] Arnon Levy and Adrian Currie (2014), "Model Organisms Are Not (Theoretical) Models," *British Journal for Philosophy of Science* 66(2): 327–348.

character state changes, than the alternative hypothesis, namely that the shared traits arose independently in extant groups from a common ancestor that did not possess the trait. This basic phylogenetic analysis licenses the defeasible inference that the last common ancestor of humans and chimpanzees was highly gregarious and highly groupish in its social ecological orientation, quite unlike the solitary and more evolutionarily distant orangutan. This phylogenetic inference is defeasible because it only holds barring compelling evidence to the contrary—a key proviso that we will return to shortly.

However, there is an important wrinkle in the phylogenetic inference when it comes to imputing social ecological properties beyond the bare fact of gregariousness to the last common ancestor of humans and chimps. Recall that the proposed model contends that exclusivist moral psychology evolved by natural selection under conditions of out-group threat, which arose in part from competition over scarce, scattered resources that tended to spark intergroup conflicts. Similar evolutionary–ecological explanations have been given for the stark differences we see between chimpanzees and bonobos (pygmy chimps) in their tendencies toward intergroup aggression, or what might loosely be referred to as "proto-exclusivist" moral psychology. For chimpanzees, resources are few and far between, and as a result violent intergroup conflict is common, with one group often raiding and brutally killing members of competing groups (patterns mirrored, to some extent, in human hunter–gatherer groups, as discussed below). Bonobos, in contrast, have abundant, static resources and, as a result, enjoy relatively peaceful intergroup relations (although they retain the presumably ancestral trait of gregariousness).

Given this divergence between chimps and bonobos in proto-exclusivist morality, we are faced with what, at first blush, appears to be a phylogenetic inference impasse: since humans are equal in evolutionary distance to chimps and bonobos, phylogenetic data do not speak in favor of or against imputing proto-exclusivist ecological conditions or capacities to early humans. We

are left instead to infer the social circumstances of early humans by drawing upon other sources of evidence, such as paleoclimatic, archeological, and cross psychological data (against the theoretical backdrop of evolutionary theory). But we need not give up on the phylogenetic inference so quickly, for there is compelling evidence that the bonobo social condition—marked by reduced levels of intergroup aggression, increased tolerance, and passive coping strategies during competitive feeding interactions (in contrast to the dominance orientation of chimpanzee and human males)—is in fact the "derived" (rather than ancestral) condition. If that is right, then the ancestral social state of the human–chimp/bonobo common ancestor more likely resembled the chimp condition than the bonobo condition and thus can be imputed, on the parsimonious assumption of faithful common ancestry, to the target lineage (namely, early humans).

Why think that the peaceful bonobo condition is derived? The reason is that evolutionary anthropologists have identified the anatomical signature of selection for reduced aggression—a syndrome of phenotypic traits that includes reductions in cranial capacity and tooth size, shortening of the face, floppy ears, and depigmentation of body parts—juvenile-like traits that are regularly observed in domesticated species and that appear to be a byproduct of selection on regulatory genes or physiological systems that produce reduced aggression.[4] It turns out that bonobos exhibit these features of evolutionary "self-domestication," suggesting that the peaceful nature of bonobo society is a derived condition that evolved in response to a highly localized ecology (namely stationary, bountiful resources), while the proto-exclusivist chimp condition is ancestral and adapted to a broader range of ecological circumstances. In essence, self-domestication syndrome is a "trace" of selection processes that

[4] Brian Hare, Victoria Wobber, and Richard Wrangham (2012), "The Self-Domestication Hypothesis: Evolution of Bonobo Psychology Is Due to Selection Against Aggression," *Animal Behavior* 83: 573–585.

took place in the prehistoric past.[5] The self-domestication hypothesis strongly suggests that the disparities we observe between chimp and bonobo aggression are the result of adaptive evolution in bonobos that occurred after their geographic isolation and phylogenetic split from chimps. A phylogenetic analysis thus supports the (defeasible) parsimonious inference that early human conditions were more chimp/bonobo-like than orangutan-like and more chimp-like than bonobo-like.

One problem with phylogeny-based inferences, however, is that there are innumerable ways that the traits of target and model lineages can be parsed, and thus more distant lineages may in some cases be more appropriate models for a given trait of the target. Thus, baboons, gorillas, chimpanzees, and even gibbons have been used to model the evolution of particular traits thought to exist in early humans. Although some have argued that relatively solitary apes, such as orangutans, are a better model of early human societies,[6] this ignores not only the phylogenetic data discussed above but also a large body of evidence from evolutionary anthropology and archeology establishing the scattered, variable nature of Pleistocene resources[7] and the ubiquity of organized warfare in pre-state societies.[8] It also overlooks

[5] For a discussion of the epistemic role of traces in the historical sciences, see Adrian Currie, *Rock, Bone and Ruin: An Optimist's Guide to the Historical Sciences* (MIT Press, 2018).

[6] See Alexandrea Maryanski and Jonathan Turner, *The Social Cage: Human Nature and The Evolution of Society* (Stanford University Press, 1993).

[7] See Chris Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard University Press, 2001); Kim Sterelny, *The Evolved Apprentice: How Evolution Made Humans Unique* (MIT Press, 2012); Peter Richerson and Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (University of Chicago Press, 2005).

[8] See Lawrence Keeley, *War Before Civilization* (Oxford University Press, 1996); Samuel Bowles (2009), "Did Warfare Among Ancestral Hunter–Gatherers Affect the Evolution of Human Social Behaviors?" *Science* 324(5932): 1293–1298; 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.

extensive research in social psychology which strongly suggests that humans appear to have innate dispositions to form groups, to essentialize them, and to make moral discriminations that closely track group membership (see further discussion in the next section). In sum, phylogenetic analyses, data about the chaotic nature of Pleistocene climate and ecology, archeological and ethnographic records of coordinated warfare in modern and prehistoric hunter–gatherer groups, and the existence of apparently "innate" group identity-formation capacities and cross-cultural in-group/out-group psychological dynamics license a reasonably strong inference: early humans lived in intensely social, highly cooperative, and rigidly group-structured environments.

Nevertheless, for the purposes of our argument against the evoconservatives, it is not necessary to take a position in the dispute as to whether human societies in the earliest stages of the EEA were smaller and more solitary or consisted of larger, competing groups. If the former scenario (call it the "solitary origins" account) turns out to be correct, then we can explain how, as encounters among groups began to occur and groups became larger, there would have been selection among genetic dispositions for a more parochial altruistic response, as well as among cultural groups for cultural innovations that reined in the originally promiscuous altruistic or cooperative response so that it extended only to members of one's own group. If the latter scenario (call it the "social origins" account) turns out to be correct, then we can explain that cooperative responses, right from the start, must have been discriminating rather than promiscuous but also point out that groups that developed the capacity to override the propensity to react with hostility toward strangers would have gained a fitness advantage under circumstances in which there were increasing opportunities for mutual benefit through exogamy, military alliances, and long-distance trade. Either scenario can explain why human beings have the capacity for inclusivist moral responses (under certain

conditions) and hence why it is misleading to say that human beings are "hard-wired" for exclusivist morality. And if it is true that humans are not hard-wired for exclusivist morality, then the prospects for further moral progress in the dimension of inclusivity look brighter than evoconservatives are willing to admit.

One main point of the preceding line of argument bears emphasis. Regardless of whether human beings originally had a nondiscriminating or a discriminating cooperative response, it is easy to understand how current human beings can have a capacity to respond in either an inclusive or an exclusive fashion, depending upon the circumstances—and it is also not hard to see why cultural innovations can either enhance inclusion or produce regressions toward exclusivity. If humans originally had a nondiscriminating disposition to cooperate (with any human beings), then there would have been selection for cultural innovations that served to restrict that disposition, in other words, selection for the capacity for exclusivist responses. If humans originally had a discriminating cooperative disposition, then, as environmental conditions changed to create more opportunities for mutually beneficial cooperation among groups, there would have been selection for cultural innovations that moderated or even overrode the disposition for exclusivity. And if environmental conditions changed again to evoke EEA-like threat cues (either through an objective deterioration of the conditions for mutually beneficial cooperation among groups or through the widespread misperception that this has occurred), then the disposition for exclusivist responses would become ascendant. To summarize: regardless of which account of the origins of human altruistic response that one accepts, there is good reason to believe that humans, rather than being hard-wired for exclusivity, have a momentous moral plasticity—a capacity for both exclusivist and inclusivist response—that is shaped by biology, culture, or (most likely) both.

Evidential Support for the Adaptive Plasticity Model

For the reasons discussed above, the adaptive plasticity model of moral psychological development is not beyond dispute. The solitary origins account, which postulates the initial evolution of an unconditional altruistic response, is not theoretically implausible. But the theoretical plausibility of a hypothesis does not give us sufficient reason to believe it is true. Unlike the solitary origins account, the thesis that human beings are pre-culturally disposed toward groupish moralities and that the development of this disposition is triggered by specific cues of out-group threat is supported by a wide range of evidence from history, psychology, biology, anthropology, and cognitive science.

For instance, evolutionary psychologists Corey Fincher and Randy Thornhill propose an adaptive plasticity hypothesis to explain the strong cross-cultural correlation between what they call "in-group assortative sociality," which is associated with ethnocentric, xenophobic, authoritarian, and conservative psychological orientations, and parasite stress.[9] They marshal a formidable amount of evidence in support of the claim that in-group bias tends to develop when signs of infectious disease are detected during human moral development, whereas less xenophobic attitudes and behaviors (or, on our terminology, more inclusivist ones) tend to emerge when cues of infectious disease are absent. We expand this account to include other signs of out-group threat broadly construed, such as competition for scarce resources and, especially, socially constructed beliefs about out-groups. In so doing, we carve out a more fundamental role for culture in our adaptive plasticity account of human moral psychology. In particular, we argue below that social-epistemic practices and

[9] Corey Fincher and Randy Thornhill (2012), "Parasite-Stress Promotes In-Group Assortative Sociality: The Cases of Strong Family Ties and Heightened Religiosity," *Behavioral and Brain Sciences* 35:61–79; C. D. Navarrete and D. M. T. Fessler (2006), "Disease Avoidance and Ethnocentrism," *Evolution and Human Behavior* 27: 270–282.

evolved cognitive biases can interact to result in faulty detections of out-group threat, resulting in the development of exclusivist tendencies in circumstances that are otherwise conducive to inclusivist morality.

It is well established that in-group/out-group biases are among the most cross-culturally robust of human psychological traits—biases that can easily be manipulated in laboratory and field study investigations.[10] The mere fact of group membership, even when the groupings are temporary and essentially meaningless, has been shown to generate these moral psychological dynamics.[11] Cues that are associated with out-group threat—including the transmission of infectious disease,[12] competition over scarce resources, external physical dangers,[13] and beliefs and practices that are dissonant with in-group values and thus imperil group cohesion[14]—have all been shown to trigger negatively valenced moral emotions, such as fear, anger, and disgust, which in turn lead to increasingly aversive intergroup attitudes and behaviors. In contrast, the adaptive plasticity hypothesis predicts that exclusivist moral tendencies tend to be attenuated in populations inhabiting environments in which cues of out-group threat are absent, and the evidence supports this prediction, too. The precise developmental pathway through which detections (whether veridical or not) lead to the relevant plastic moral response is unknown; all that matters for the purposes of our theory, however,

[10] Donald Brown, *Human Universals* (McGraw Hill, 1991).

[11] H. Tajfel and J. C. Turner, "The Social Identity Theory of Intergroup Behavior," in S. Worchel and W. G. Austin (eds.), *Psychology of Intergroup Relations* (Nelson-Hall, 1986, pp. 7–24).

[12] See Fincher and Thornhill, "Parasite-Stress," supra note 9.

[13] See B. M. Riek, E. W. Mania, and S. L. Gaertner (2006), "Intergroup Threat and Outgroup Attitudes: A Meta-Analytic Review," *Personality and Social Psychology Review* 10(4): 336–53; M. Sherif and C. W. Sherif, "Ingroup and Intergroup Relations: Experimental Analysis," in M. Sherif and C. W. Sherif (eds.), *Social Psychology* (Harper & Row, 1969, pp. 221–266).

[14] Richerson and Boyd, *Not by Genes Alone*, supra note 7; Hewstone, Rubin, and Willis, "Intergroup Bias," supra note 1.

is that such pathways exist and act with some reliability. The more out-group threat cues that are present in the developmental environment, the stronger the statistical biasing toward exclusivist moral tendencies will be.

If the capacity for moral responses is an adaptively plastic trait, then the common assertion that exclusivist morality is "hard-wired" in humans and that inclusivity goes against the grain of our evolved moral nature is extremely misleading because it wrongly suggests that exclusivist dispositions are invariant across all environments. Further, the concept of an adaptively plastic trait can accommodate a more nuanced understanding of what the EEA was like. Even though it is true that the EEA lacked developed institutions for mutually beneficial cooperation among groups (including preeminently a developed market economy), there is evidence that limited cooperation among groups sometimes occurred and may have been commonplace in some locales. Intergroup cooperation in the EEA, as noted above, included exogamy (marrying members of other groups), trade (sometimes over long distances), and military alliances. If humans were hard-wired for exclusivist morality—if they uniformly reacted with fear and hostility to strangers and failed to show any consideration for their interests—exogamy, trading, and intergroup alliances could not be sustained to the degree that they were. The key point here is that the EEA was not uniformly and thoroughly inimical to cooperative and even respectful relationships. While it is likely that in most cases the EEA was overall rather unfriendly to inclusivist moral responses, there were clearly local exceptions.

An evolved moral psychology that included an adaptively plastic capacity to respond to strangers would have been more fitness-enhancing than one that was hard-wired for exclusion. To that extent, one would expect that selection pressures in the EEA would have favored the development of a flexible capacity for both inclusivist and exclusivist responses that is responsive to local ecological demand. Groups that developed this adaptive moral plasticity would have had a fitness advantage over those

that did not, so the capacity for conditional moral expression would spread in the human population.

Critics of evolutionary psychology have rightfully cautioned that one cannot reliably infer from a particular adaptive pattern of behavior that there is a specific organ or cognitive mechanism "designed" by natural selection to produce that behavior.[15] However, the "how possibly" explanation that we offer to account for the evolution of moral inclusivity is not a wildly speculative or empirically irrefutable "just so" story, concocted in an ad hoc manner that dismisses non-adaptive explanations out of hand. To the contrary, the model we propose is empirically constrained in that it has the benefit of broad-based evidential support and takes non-adaptive—indeed non-biological—factors seriously.

The basic in-group/out-group dynamics around which our model is built are robustly cross-cultural and develop predictably very early in individual development—which is indicative of some degree of pre-cultural specification. Furthermore, there is a great deal of experimental, sociological, and historical evidence to support the plasticity thesis. Recall that developed market economies make peaceful, mutually beneficial behavior among people from different groups—call them "strangers"—possible and give people incentives to engage in it. Norbert Elias and others have argued that once these opportunities for peaceful, mutually beneficial relations among strangers become pervasive, there is cultural selection for behavior (as well as attitudes and motivations to support the behavior) that signal the willingness to cooperate peacefully and on terms of reciprocity.

Joseph Henrich and collaborators provide experimental evidence to support this hypothesis in experiments involving the ultimatum game.[16] In the ultimatum game, the investigator

[15] Elizabeth Lloyd (1999), "Evolutionary Psychology: The Burdens of Proof," *Biology and Philosophy* 14: 211–233.

[16] Joseph Henrich et al. (2010), "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment," *Science* 327(5972): 1480–1484.

authorizes one subject to choose how much of some significant amount of money to share with another participant (a stranger) in a one-shot (ephemeral) interaction. The second participant stipulates beforehand whether he or she will accept or refuse allocations of particular sums; if the actual offer is rejected, then each player receives zero.

The results are strikingly different, depending on whether the population playing the game includes individuals from developed market economies or individuals from societies in which markets are not developed. People from populations with high levels of market integration are much more likely to exhibit a commitment to treating strangers fairly; in contrast, individuals from communities that lack robust market economies tend to have little compunction against dividing the money in extremely unfair ways. The explanation for these differences is not, of course, that human populations differ in their innate moral psychology—but rather that certain moral norms have proliferated through cultural selection in human populations that possess robust institutional frameworks for cooperation. These experiments support our main thesis that human beings are not hard-wired for exclusivist moral responses but instead possess an adaptively plastic trait: the capacity to modulate their responses depending upon the environmental context in which their moral capacities develop. In this case, the exclusivist moral response depends on whether individuals encounter one another in the context of developed market relations and hence are able to see one another as mutually beneficial cooperators, or rather as strangers who are to be feared at worst and disregarded at best.

Recent work by Victor Kumar lends additional credibility to the adaptive plasticity hypothesis.[17] He argues that the disgust response evolved as a proximate mechanism of exclusion: persons

[17] Victor Kumar (2017), "Foul Behavior," *The Philosophers' Imprint* 17(15) https://quod.lib.umich.edu/p/phimp/3521354.0017.015/—foul-behavior?view=image.

who are objects of disgust tend to be avoided and excluded from valuable social relationships. When one individual reacts with disgust toward another, he or she exhibits the emotion of disgust through characteristic body language, facial expression, and vocal intonations. These all serve as signals to others that the object of disgust is to be avoided and thereby excluded—in effect relegating the object of disgust to something like the status that strangers typically had in the EEA. Suppose that Kumar is right about the social meaning and function of the disgust reaction—suppose that disgust is an "othering" or outcasting mechanism, a response that signals that the object of disgust is to be excluded in some important way. Clearly, whether one human being reacts with disgust to another human being can depend on how the first individual has learned to see the second individual and that, in turn, can be shaped by the first individual's culture. This further supports our view that it is a mistake to see exclusivist responses as hard-wired: instead, they are adaptively plastic, and whether an exclusion response occurs—and toward whom—is subject to modification by culture.

Adaptive Plasticity and the Limits of Cultural Malleability

This does not mean that exclusivist responses are infinitely malleable through cultural influences. If moral developmental environments prominently feature certain threat cues that were pervasive in the EEA, then there may be limits to the formative influence of culture. For example, if members of another group exhibit extreme hostility, are seen to carry deadly diseases, or are imposing values that threaten to severely undermine cooperation in one's own group, then it will be extremely difficult for cultural innovations, including cosmopolitan commitments, to overcome reactions of fear, hostility, and disgust toward them. It will likewise be difficult for cosmopolitan norms to take root and to be sustained if there are widespread *perceptions* of these conditions,

even if these perceptions are not veridical (see Chapter 7 for an extended discussion of this crucial point).

By the same token, one should not underestimate the difficulty of altering entrenched cultural moral systems even in favorable moral developmental environments. Even if biological and social conditions are ripe for the development of a more inclusivist morality, the inertia of cultural evolutionary hangovers can prevent or delay the development of more inclusive moral norms. For example, the significant differences in cultural conceptions of honor between northern populations in the United States and those in the American South have been explained as the result of these regions being settled by peoples with different cultural moral systems adapted to different historical ecologies.[18] In particular, Scotch-Irish livestock herders were the predominant settlers of the South, whereas peasant farmers from Germany, England, and the Netherlands were the chief settlers of the North. Livestock herding is robustly associated with hyper-masculine, honor-based cultures around the world because it typically occurs in rugged, lawless regions of countries where theft and other forms of predation are commonplace—and where violent reactions serve as a necessary deterrent in the absence of an effective police force.[19] Despite being arguably ill-suited for the ecological conditions of twentieth-century America, the honor culture of the American South has been slow to change and southern moralities have struggled to become more inclusive. To the extent that cultural moral demographics of the South have begun to shift in inclusivist directions, the relocation of relatively culturally inclusive Northerners to the South has likely played a significant role. The developmental interconnectedness of certain norms and values in a cultural web can make them difficult to modify, and

[18] R. E. Nisbett and D. Cohen, *Culture of Honor: The Psychology of Violence in the South* (Westview Press, 1996).

[19] Stefan Linquist (2015), "Which Evolutionary Model Best Explains the Culture of Honour?" *Biology and Philosophy* 31: 213.

individuals already primed for exclusivist responses are liable to react in psychologically hostile ways to the introduction of potentially destabilizing moral norms.

The fundamental point is not that inclusivist progress is easy to achieve but rather that, from the standpoint of both theory and experimental evidence, the adaptive plasticity hypothesis fares better than the hard-wired hypothesis. And this matters greatly for the prospects of moral progress, as will soon become clear.

Advantages of a Biocultural Account of Moral Development and Evolution

The central idea of the naturalized theory we have proposed is that whether the toggle (or dial) of the adaptively plastic capacity moves toward exclusion or inclusion depends on whether certain threat cues are salient during moral development. It is vital to emphasize that this is no more an environmental determinist view than it is a genetic determinist view. The claim is that favorable environments—ones in which the harsher conditions prevalent in the EEA are muted—create a space for the development of inclusivist responses but do not ensure it. Whether the potential for inclusivist morality is realized depends, as we shall see, upon a number of factors, including the presence of incentives for developing cooperative relationships with strangers (which markets preeminently provide) and on cultural innovations of various sorts, including communication and transportation technologies that link previously separated groups, techniques for perspective-taking, reductions in parasite threat, and improvements in moral concepts and moral reasoning.

This chapter has articulated the main outlines of a naturalized theory of moral progress that takes the evolutionary history of human moral capacities seriously, while avoiding the error of underestimating the power of culture when it fosters the exercise of the capacity for open-ended normativity. The key to this account is the hypothesis that human beings evolved an adaptively

plastic capacity to develop either exclusivist or inclusivist moral responses and corresponding social practices and institutions, depending upon the environment and whether it mimics—or is thought by its inhabitants to approximate—the harsh conditions of the EEA.

The chief advantages of such a biocultural theory are these. First, unlike the moral hard-wiring story, the adaptive plasticity account is compatible with important facts about morality as it now exists for many human beings and is reflected in significant social practices and institutions. In other words, the various inclusivist phenomena that present as anomalies for the hard-wiring view are perfectly consistent with the adaptive plasticity view. So our theory can explain what the rival theory cannot. Our theory can also explain everything that the rival theory can explain—and it can explain these features better. This is because our theory gives a more informative explanation of why exclusivist moral responses occur when they do, while at the same time explaining why they can give way to more inclusivist responses under certain conditions. Second, by relying on the adaptive plasticity hypothesis, our theory beats the hard-wiring theory at its own game: it tells a more convincing evolutionary story. Given that there were some opportunities in the EEA for intergroup cooperation, selection would be expected to have favored an adaptively plastic capacity over an inflexible or hard-wired capacity that would have resulted in lost opportunities for fitness-enhancing intergroup cooperation.

To summarize the discussion thus far: early human groups evolved under ecological conditions that commonly favored the development of exclusivist morality and severely penalized inclusivist tendencies. Such conditions included:

(1) Severe competition for resources among scattered, weakly genetically related groups, with levels of productivity sufficiently low that sharing resources with out-groups entails dangerously high costs.

(2) The absence of institutions (in particular markets and security arrangements) to facilitate peaceful, mutually beneficial cooperation *among* groups—in contrast to the existence of efficacious institutions *within* hunter–gatherer societies to prevent powerful individuals from monopolizing resources and exploiting vulnerable individuals.
(3) High risk of infection by biological and social parasites: these include pathogens carried by members of foreign groups to which one's own group had little or no immunological resistance and human "social parasites" whose integration into a host group risked undermining social cohesion through free-riding or a lack of familiarity with or commitment to host group norms.

Such ecological conditions would have generally favored moralities underwritten by truncated forms of moral emotions. In particular, sympathy would have been circumscribed to in-group members, resulting in severe limitations on the capacity for altruistic behavior beyond the confines of the group.[20] Indeed, theory suggests that parochialism was a precondition for the evolvability of human altruism, and there is evidence that altruism and parochialism are mediated by a common proximate developmental cause. Studies show that oxytocin, a neurotransmitter that modulates empathy, not only enhances intragroup altruism but also accentuates exclusivist moral response, particularly in competitive intergroup environments (see Chapter 11 for a more detailed discussion).[21] Altruism and exclusivism appear, therefore, to be two sides of the same adaptive coin—and this makes good sense if, as the received evolutionarily view suggests, morality was forged in the crucible of intergroup conflict.

[20] Jonathan Haidt, *The Righteous Mind* (Pantheon, 2012).
[21] Samuel Bowles and Herbert Gintis, *A Cooperative Species. Human Reciprocity and Its Evolution* (Princeton University Press, 2011).

Nevertheless, this is far from the whole story. According to the model we propose, in environments in which out-group threat cues are attenuated or absent, adaptively plastic moral response mechanisms permit the evolution of more inclusivist moral responses, which can be shaped by social and in particular institutional conditions. This brings us to the thesis that inclusivist morality is in effect a "luxury good."

Inclusivist Morality as a Luxury Good

Chapters 5 and 6 argued that the view that there are strong evolutionary constraints on morality overlooks the existence of a remarkable human characteristic: the capacity for open-ended normativity—a capacity that is crucial to understanding the development of more inclusive moralities and other types of moral progress as well. While evolutionary developmental environments have favored varying degrees of exclusivity over the course of human history, conditions amenable to the exercise of open-ended normativity and hence to the development of more inclusivist moralities appear to be rare. In particular, they seem to be connected to a range of recent sociopolitical developments that have taken place predominantly in highly resourced populations. Such developments include (inter alia) healthcare and public health infrastructures, reductions in crime, rule of law, property rights, literacy, and the emergence of markets, to name a few. There is an important sense, therefore, in which inclusivist morality is a luxury good.

If the adaptive plasticity hypothesis of exclusivist morality is right, then moral progress and the above sociopolitical developments do not merely have a common cause: they are reciprocal causes of one another. Conditions of infectious disease, physical insecurity, interethnic conflict, and low rates of productivity seed exclusivist moral responses, which in turn feed back into the exacerbation and perpetuation of the conditions that trigger

exclusivist tendencies. Furthermore, there is a link between exclusivist psychological orientations and mental rigidity, closed-mindedness, dogmatism, and fear of uncertainty.[22] Individuals exhibiting these psychological orientations are less able or willing to critically examine assumptions underlying their moral worldview, to perceive the complexities of moral problems, to acknowledge that they hold logically contradictory beliefs, or to be motivated to iron out logical contradictions within their belief system. These traits, in turn, make it difficult or impossible to subject one's values and cultural practices to critical scrutiny, thus impeding inclusivist moral development and perhaps moral progress more generally. In other words, it is likely that the same environmental conditions that impede the development of inclusivist morality inhibit other forms of moral progress as well.

Although the focus of our inquiry is on moral inclusivity, we surmise that other forms of moral progress (such as proper demoralization and improvements in our understandings of virtues, moral concepts, and morality itself) may also be luxury goods. This would be the case if, for example, it turns out that these moral improvements rely upon prior improvements in moral reasoning—and if, as we have suggested, the efficacy of moral reasoning as a significant contributor to moral progress requires favorable conditions.

By the same token, cultural innovations that alleviate conditions that trigger exclusivist responses act to break the vicious spiral, creating an environment in which inclusivist morality can flourish. Cultural innovations can modify evolved moral responses in two ways. First, they can remove or ameliorate the harsh conditions of the EEA. This has been accomplished, for example, by the division of labor and improvements in agricultural technologies that greatly increased the social surplus, thus

[22] John T. Jost et al. (2003), "Political Conservatism as Motivated Social Cognition," *Psychological Bulletin* 129: 339–375.

reducing the cost of sharing resources with strangers; by the development of institutional infrastructures for peaceful, mutually beneficial cooperation among groups; and by improvements in medicine and public health that dramatically reduced parasite stress.

Second, cultural innovations in the form of new moral norms, more sophisticated moral reasoning, and new techniques for perspective-taking can reshape moral responses; but this is likely to occur on a large scale only if economic conditions are sufficiently favorable and there is a reasonable degree of physical security, both actual and perceived. It is beyond the scope of this book to develop a comprehensive account of how cultural innovations can, under favorable conditions, result in moral progress for significant numbers of people and in such a way as to change social practices and institutions. Instead, we will simply offer a few illustrations of how progress in the form of inclusiveness has occurred. All of the illustrative cases fit the luxury goods hypothesis: the morally progressive change occurred on a large scale only recently and under conditions favorable to the development of inclusivist moral responses. That is, in each case progress was achieved on a significant scale in societies that had already attained high levels of physical security and material abundance.

Case Example: Abolition

A remarkable example, or rather set of examples, of cultural innovations that contributed to increased inclusivity is extensively documented by historians of the British abolitionist movement. In order to convince people that slavery was a wrong and such a serious wrong as to require legal prohibition, with all the economic costs this entailed, British abolitionists had to overcome or at least weaken the racist ideology that supported slavery and to overcome the complacency of those who were not deeply racist but simply turned a blind eye to the

evil.[23] To accomplish these goals, abolitionists employed a number of techniques that evidenced a remarkable grasp of both human psychology and what is now called "social epistemology." To counter the belief that Africans were not sufficiently rational to possess natural rights, including the right to liberty, abolitionists arranged extensive speaking tours and funded books and journal articles in which freed African slaves publicly demonstrated their rationality. They also developed sophisticated techniques for evoking sympathy for the suffering of slaves. For example, anti-slavery societies sent artists, under false pretenses, to travel on slave ships and to make detailed drawings of the unspeakable conditions to which slaves were subjected in the Middle Passage, which were later copied and distributed widely. In addition, taking advantage of existing norms of epistemic (and moral) deference to the clergy, they worked to win over the clergy and even provided them with "canned" anti-slavery sermons.[24] Perhaps the most important cultural innovation that contributed to the success of abolitionism was the printing press, along with a great increase in literacy in the decades prior to the founding of the movement—which dramatically amplified the effects of the aforementioned cultural innovations.[25]

[23] In Chapter 4, we note that much of the opposition to emancipation was not explicitly racist. Predictions that freed slaves would engage in violence against their former masters, and forecasts of economic ruin were perhaps as powerful as outright racism, at least in the case of British abolitionism.

[24] For an accessible account of abolitionist techniques that draws upon and synthesizes much primary scholarship, see Adam Hoschild, *Bury the Chains: Prophets and Rebels in the Fight to Free an Empire's Slaves* (Houghton Mifflin Harcourt, 2005).

[25] Some scholars have argued that the development of the novel helped some people to broaden their empathy and extend their sympathy to foreigners, to women, and to members of other social classes. This technique for fostering inclusivity, like the ones previously noted, depended on the great cultural innovation of literacy plus printing. See, for example, Martha Nussbaum, *Cultivating Humanity: A Classical Defense of Reform in Liberal Education* (Harvard University Press, 1997).

Many abolitionists apparently were motivated in part by the belief that slavery was incompatible with Christianity, but their strategies for mobilizing anti-slavery sentiment and political action included techniques that operated independently of explicitly religious appeals. Indeed, there have been countless places and times in which religious beliefs have served to justify and reinforce exclusivist moralities and drive moral regressions, including slavery. The fact that religious beliefs and motivations have resulted in both uncontroversial moral progressions and uncontroversial moral regressions suggests that there are other difference-making factors at play in driving these moral trajectories. A naturalized account of abolitionist successes does not deny the importance of religious belief and motivation but instead explains how a combination of favorable circumstances, evolved psychological responses, the capacity for open-ended normativity, and social-epistemic practices enabled religious activists to bring about one of the greatest instances of moral progress.

Case Example: Animal Welfare

For a second example, think of the techniques employed by advocates for the better treatment of animals. These include the distribution of films and television spots depicting the mistreatment of animals in laboratories, "factory" farms, and meat-processing plants (similar to abolitionist artists depicting the horrors of the Middle Passage), as well as the dissemination of scientific information to show that animals used in experimentation and food production experience pain and fear much as humans do. Through direct appeals to emotions by offering descriptions and images of animal suffering and by changing our beliefs about the capacity of animals for suffering, these techniques extend our sympathy while at the same time revealing the inconsistencies in our moral responses and behavior. The case of progress in the

treatment of non-human animals is especially encouraging because the impetus has come not from the oppressed community itself but from outside. It is a remarkable illustration of the capacity for inclusivist moral thinking and behavior.

Evidence for the "Luxury Good" Hypothesis

The key point is that these inclusivity-advancing cultural innovations are only likely to arise, become pervasive, and take root under highly favorable socioeconomic conditions. Our hypothesis that inclusivist morality is a luxury good fits the historical evidence, in several respects. First, significant penetrance of inclusivist moral commitments in human populations, such as the extension of moral regard to non-human animals and the condemnation of slavery, is a rather recent phenomenon and appears to correlate, roughly, with the remarkable gains in productivity that began in Britain and western Europe in the mid-eighteenth century.[26] While it is true that vegetarianism has been practiced by some members of some Asian religious cultures—Jains, Hindus, and Buddhists—a more general shift in attitudes toward the treatment of animals, translated into widespread legal and institutional reform, is relatively recent and appears to have been initiated mainly in societies of relative abundance and security. Further, it is not clear that the practice of vegetarianism in these Asian religious traditions indicates the recognition that non-human animals have moral status on their own account; instead, eating them is avoided in order to escape the cycle of reincarnation. Second, periods of severe economic downturn correlate with increases in xenophobic and racist behavior, particularly when out-groups

[26] Gregory Clark, *A Farewell to Alms: A Brief Economic History of the World* (Princeton University Press, 2009); and Robert William Fogel, *The Escape from Hunger and Premature Death, 1700–2100* (Cambridge University Press, 2004).

(including minorities within larger populations) are salient.[27] Third, in conditions of great physical insecurity and where the institutional infrastructure for peaceful, mutually beneficial relations among groups has broken down—as in the case of failed states or war zones—group ties strengthen, while hostility toward and distrust of out-groups increase.[28] Fourth, outbreaks of deadly infectious diseases (whether the recent Ebola epidemic or the Mexican typhus outbreak in the early twentieth century) tend to evoke disproportionate fears among significant numbers of people, including those in developed nations far from the site of the outbreak, disposing them to adopt unusually harsh policies toward foreigners and immigrants within their own borders.[29]

In later chapters we will elaborate the luxury good hypothesis. Here we wish only to emphasize that although moral progress in the form of inclusion is only likely to occur and be sustained in environments that do not feature the harsh conditions of the EEA, there are additional necessary conditions for progress. In particular, a complex social-epistemic environment is needed. The case of British abolitionism, to which we will recur later, nicely illustrates this point: British society in the late eighteenth and early nineteenth centuries was not only more materially prosperous and physically secure than ever before; it also featured impressive communication technologies made effective by unprecedented levels of literacy as well as political conditions that included freedom of expression and the responsiveness of government to public opinion.

[27] Lincoln Quillian (1998), "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe," *American Sociological Review* 60(4): 586–611.

[28] Linda Tropp, ed., *The Oxford Handbook of Intergroup Conflict* (Oxford University Press, 2012, p. 116).

[29] H. Markel and A. M. Stern (2002), "The Foreignness of Germs: The Persistent Association of Immigrants and Disease in American Society," *Milbank Quarterly* 80(4): 757–788.

Proceeding on the assumption that a general theory of moral progress should illuminate moral progress in the form of inclusion, this chapter has proposed an alternative evolutionary model of moral psychological development and evolution and has demonstrated how this naturalized theory helps to flesh out more satisfying explanations of a number of historical gains in inclusion. The next chapter elaborates on the biocultural dimensions of the theory and shows that it provides valuable insights into how regression toward moral exclusivity comes about.