THE EVOLUTION
OF MORAL PROGRESS

A BIOCULTURAL THEORY

ALLEN BUCHANAN
AND
RUSSELL POWELL

OXFORD
UNIVERSITY PRESS

CHAPTER 7

Naturalizing Moral Regression

A Biocultural Account

The previous chapter sketched an evolutionary model of exclusivist moral psychological development and showed that 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 environment of evolutionary adaptation (EEA) have been overcome. This chapter advances two further hypotheses that draw upon and extend this model: first, that inclusivist gains can be eroded if these harsh conditions reappear or if significant numbers of people come to *believe* that they exist and, second, that 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. Our aim will be not to make a bulletproof case for each of these theses but simply to show that they are important elements of any naturalistic theory of inclusivist moral progress.

Regression and the Demagogic Manipulation of Belief

A theory of moral progress ought to explain not only how progress occurs but also how regression can come about. The key to our naturalistic account of regression is the prediction that inclusivist gains will tend to be eroded if EEA-like conditions return or if enough people come to believe such conditions exist. This hypothesis gains plausibility from the same evidence that supports the "luxury good hypothesis" discussed in the previous chapter—but it is also supported by the fact that, as we have suggested, exclusivist moral responses that were selected for in the EEA can be triggered by people's *perceptions* of their predicament. For the exclusivist moral response to be activated, such perceptions need not be veridical—that is, it is not necessary that competition among groups actually be unavoidably severe or that allowing foreigners into one's society will actually result in deadly epidemics, threaten the stability of existing norms, or undermine cooperation in some other way; all that is necessary is that people come to *believe* this is so.

Importantly, the same resources for cultural innovations that made inclusivist morality possible can also be used to dismantle it. This is precisely what occurs when certain people (such as extremist political elites) have a dominant interest in provoking exclusivist moral responses in others and have the social power and psychological savvy to act effectively on this interest. Those who mobilize exclusivist moral responses can succeed in either of two ways. The first is by directly creating an environment that is, objectively speaking, friendly to exclusivist morality and unfriendly to inclusiveness. This occurs when such individuals provoke highly destructive intergroup conflicts that destroy institutional infrastructures for peaceful interaction and public health or create conditions of severe scarcity and ruthless competition for resources. Alternatively, governments or political leaders can create an environment that is *subjectively* unfriendly to inclusivist morality by persuading enough people that they are living in an environment that mimics the harsh characteristics of the EAA, even when in reality it does not.

One salient tactic common to those who manipulate belief to encourage exclusivist attitudes is to blame social problems on

some external group characterized as a dangerous "other." To take another historical example: many Southerners who resisted school integration and other civil rights gains in the 1950s and 1960s claimed that it was "outside agitators" who were causing these changes, implying that people within southern society, including African Americans, who were satisfied with the status quo were the victims of an aggressive invasion of foreign ideas. The ubiquity and power of this style of discourse make perfectly good sense on our naturalized theory of moral regression, given its emphasis on the potency of representing "the other" as not only alien but also dangerous—if not biologically or physically, then socially. Similarly, prominent figures on the political right blame the decline of American manufacturing on "unfair" trade practices by other countries, with no mention of the role of automation in reducing the number of manufacturing jobs. As we noted in the Introduction, a significant type of moral progress is the recognition that some misfortunes are not the result of the actions of malicious "others" but are due instead to impersonal forces. The tendency to blame all problems affecting Americans on foreigners is a clear and potentially destructive case of regression.

If the manipulators of exclusivist tendencies can succeed in making enough people believe that out-groups pose serious threats, this will not only strengthen in-group ties; it will also elicit out-group antagonism, which in turn can cause people to act in ways that induce reciprocal fear in out-groups—and what began as a misperception of intergroup threat will rapidly become reality. In other words, an initial misperception that another group is hostile can prompt hostile behavior toward that group, which in turn will lead that group to respond in kind, resulting in a spiral of epistemic reinforcement. As political scientist Robert Jervis has shown, even if the initial response prompted by a misperception that the other group is hostile is purely defensive, it may be misinterpreted as aggressive— a dynamic we have seen time and again in, for example, cold

war brinkmanship.[1] In a similar vein, social ostracism causes members of oppressed groups to judge their oppressors as less than human, as well as to infer that their oppressors view *them* as less than human—resulting, again, in the mutual reinforcement of subjective out-group threat.[2] Likewise, the ghettoization of oppressed groups into substandard living conditions serves to "confirm" morally relevant beliefs about out-groups, such as the notion that they are breeding grounds for crime or disease, which in turn are used to justify their social exclusion.

Evolved Cognitive Biases and Perceptions of Out-group Threat

Recent work in the psychology of normal cognitive biases and errors helps flesh out the idea that misperception can trigger responses that were adaptive in the EEA even when, objectively speaking, the conditions of the EEA no longer obtain. Lawrence Hirschfeld provides impressive empirical work to support the hypothesis that cognitively normal human children exhibit, at a very early age, what might be called an essentializing "natural kinds" ontology with respect to human groups.[3] In simplest terms, children tend to sort the human beings they encounter or hear about into groups and assume that all members of a given group share a hidden essence that determines, in rather rigid fashion, how all members of the group behave. Hirschfeld's point is not that

[1] Robert Jervis, *Perception and Misperception in International Politics* (Princeton University Press, 1976, pp. 58–62).
[2] See Brock Bastian and Nick Haslam (2010), "Excluded from Humanity: The Dehumanizing Effects of Social Ostracism," *Journal of Experimental Social Psychology* 46: 107–113.
[3] Lawrence Hirschfeld, *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds* (Bradford, 1998); S. A. Gelman (2009), "Learning from Others: Children's Construction of Concepts," *Annual Review of Psychology* 60: 115–140; Ilan Dar-Nimrod and Steven J. Heine (2011), "Genetic Essentialism: On the Deceptive Determinism of DNA," *Psychological Bulletin* 137(5): 800–818.

children are born racist but that they do have a psychological disposition to essentialize human groups, which can, given the right environment, provide a template for the development of racist attitudes and behaviors.

Also drawing on a considerable empirical literature, Sarah-Jane Leslie examines a normal cognitive error that may feature in a proximate explanation of how the psychological disposition Hirschfeld documents can result in exclusivist moral responses.[4] She notes that what she calls "generic overgeneralization" occurs when one sees—or believes—that some member of another group has exhibited dangerous or violently aggressive behavior, and as a result one comes to believe that all members of that group will behave in the same way. Hirschfeld's analysis makes this apparent case of hyper-inductions more explicable: if all members of the group share a common deterministic essence, then an observation that one member of the group behaves in a certain way provides a basis for concluding that they all do.

This tendency to essentialize human groups is reinforced by the intergroup asymmetry observed in the so-called fundamental attribution error: people tend to attribute positive in-group behaviors to internal character dispositions and negative in-group behaviors to situational factors, whereas they make the reverse set of attributions in relation to out-group members.[5] Indeed, what is disturbing about generic overgeneralization is that it apparently only applies in connection with highly negative behavior.[6] If a member of another group exhibits commendable behavior, people do not tend to attribute that behavior to all other members of the group.

Evolutionary risk management theory can go some way toward explaining this asymmetry of attribution.[7] Recall that the adaptive plasticity account holds that moral development is shaped by the detection of out-group threat. The detection of out-group threat, in turn, involves probabilistic "judgments" under conditions of uncertainty. In these circumstances, evolutionary theory predicts that certain cognitive biases will evolve as a result of an adaptive error rate asymmetry between false positives and false negatives.

In the EEA, when it came to judgments about whether a stranger was dangerous, the risk attaching to a false negative was much greater than the risk of a false positive. That is to say, a false judgment that a stranger was innocuous could be lethal—and thus would have entailed far greater risks than a false judgment that a stranger was dangerous—which would merely have resulted in lost opportunities from forgoing prosocial interactions with out-group members. Given the paucity of social practices or institutions for mutually beneficial interactions with strangers and given high levels of biological and social parasite threat, a false judgment that a stranger was innocuous could be disastrous to the in-group—and thus would have entailed far greater risks than a false judgment that a stranger was dangerous or not to be trusted. In such an environment, erring on the side of false positives would be adaptive, and hence there would be selection for generic overgeneralization in relation to negatively valenced out-group traits.

One might think that the proclivity to essentialize human groups is simply a byproduct or evolutionary fallout of the adaptive tendency to essentialize the biological world in general.[8] The

[4] Sarah-Jane Leslie (2017), "The Original Sin of Cognition: Fear, Prejudice and Generalization," *Journal of Philosophy* 114(8): 393–421.

[5] Miles Hewstone (1990), "The 'Ultimate Attribution Error'? A Review of the Literature on Intergroup Causal Attribution," *European Journal of Social Psychology* 20(4): 311–335.

[6] Leslie, "The Original Sin of Cognition," supra note 4.

[7] M. G. Haselton and D. Nettle (2006), "The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases," *Personality and Social Psychology* 10: 47–66; M. Haselton, D. Nettle, and P. Andrews, "The Evolution of Cognitive Bias," in D. M. Buss (ed.), *The Handbook of Evolutionary Psychology* (Wiley and Buss, 2005, pp. 724–746).

[8] See F. J. Gil-White (2001), "Are Ethnic Groups Biological 'Species' to the Human Brain?" *Current Anthropology* 42: 515–554.

tendency to essentialize the biological world appears to be innate and is most likely an adaptation.[9] Very young children, across very disparate cultures, come to form essentializing beliefs about species of animals, for example—that is, they attribute the properties of animals of a single species or type to a common, immutable essence. This tendency may be fallacious from a scientific perspective, but it serves as a reasonably good evolutionary heuristic, given that conspecifics will tend to behave in characteristic, predictable ways. So perhaps ethnic "essentializing" implicates the same cognitive faculties that identify and characterize biological species. If the tendency to essentialize living things were the whole story, however, then people would essentialize positive traits of out-groups as much as they do negative traits. Yet as we have seen, there is a fundamental asymmetry in this regard. Thinking of out-groups as natural kinds and attributing their negative (but not positive) behavior to internal, immutable character dispositions possessed by every member of the group looks very much like an evolutionary biological heuristic for managing out-group threat. Like conspecifics, co-ethnics share many properties that are not evident from superficial inspection—and given the asymmetric cost between false negatives and false positives, generic overgeneralization may not only allow for the successful prediction of individual behavior but also help avoid the risks that attend intergroup interactions.

The Social (Mis)Construction of Belief

Once properly fleshed out, an adaptive plasticity account of exclusivist morality tells us *part* of what we need to know about how to increase the probability that moral progress will persist and grow. Much of the remaining part of what we need to know is supplied by integrating our knowledge of evolved cognitive

[9] See G. A. Gelman, *The Essential Child: Origins of Essentialism in Everyday Thought* (Oxford University Press, 2003).

biases, discussed above, with social moral epistemology. By "social epistemology" we mean the critical evaluation of alternative social practices and institutions with regard to their efficacy and efficiency in promoting true or justified beliefs.[10] Social *moral* epistemology focuses on the social promulgation of beliefs that tend to be crucial for moral judgment, moral reasoning, and the moral emotions.[11] How do social moral-epistemic practices interact with cognitive biases and other evolved features of human moral psychology to impede or facilitate the development of inclusivist morality?

Inclusivist shifts, we have suggested, are the result of cultural innovations that can flourish and be sustained only under a narrow range of moral developmental environments, making inclusivist morality a luxury good. We have further suggested that the moral bridges that these cultural innovations provide can be dismantled using the same materials that were used to construct them: human psychology and culture. It will prove valuable to home in now and elaborate on how some individuals can use these resources to manipulate the beliefs of others in such a way as to trigger exclusivist moral responses, thereby reversing the gains that constitute an important form of moral progress.

The vast literature on genocides and ethnic cleansings, as well as that on eugenic forced sterilizations, demonstrates that those who mobilize others to commit violations of basic human rights on a massive scale often rely on a technique that involves "dehumanization of the other."[12] Dehumanization is one type of

[10] Social epistemology focuses on the social norms and processes by which some individuals come to be regarded as experts in various domains of knowledge, on how individuals come to seek expertise and to identify experts, and, more generally, how beliefs are socially promulgated.

[11] Allen Buchanan (2002), "Social Moral Epistemology," *Social Philosophy & Policy* 19(2): 126–152.

[12] N. Haslam, "Dehumanization: An Integrative Review," *Personality and Social Psychology Review* 3: 252–264; and Jonathan Glover, *Humanity: A Moral History of the Twentieth Century* (Yale University Press, 2001).

so-called delegitimizing belief,[13] a class of extremely negatively valenced attributions to out-groups that serve to exclude them from the moral community.[14] Dehumanization involves classifying out-groups as subhuman, either by identifying them as non-human animals with lesser or no appreciable moral status (such as vermin or insects) or by identifying them as negatively valenced supernatural entities (such as evil demons).

The first step in the delegitimization process, however, is to convince people that some people are the "other"—members of a distinct group that is significantly different from one's own, and different in ways that warrant hostile actions toward its members. Even if, as Hirschfeld and others have shown, there is an innate disposition to sort people into groups, how the sorting plays out depends on how children and adults are acculturated. For example, in Nazi Germany, children were taught to identify Jews by the shape of their supposedly distinctive noses. Once a group is identified—or rather constructed—the next step is to create the perception that they are less than human, or more like beasts, with respect to, for example, their reduced reasoning capacities, their tolerance of pain, their lack of uniquely human moral emotions, their tendency to transmit infectious disease, and so on. If the out-group is thought to lack traits like rationality, this precludes entertaining the possibility that intergroup conflicts could be resolved through reason-based negotiations.

For example, Nazi propaganda, in political speeches, textbooks, and cartoons, portrayed Jews as a deadly bacillus infecting society and as plague-carrying rats. Similarly, propaganda that fueled the Rwandan genocide referred to Tutsis as cockroaches. From the standpoint of manipulating beliefs in order to trigger

---

[13] D. Bar-Tal, *Shared Beliefs in a Society: Social Psychological Analysis* (Sage, 2000, pp. 121–122).

[14] See also S. Opotow (1990), "Moral Exclusion and Injustice: An Introduction," *Journal of Social Issues* 46: 1–20.

exclusivist moral responses, these dehumanizing metaphors kill two birds with one stone: they activate the parasite threat response that triggers disgust, fear, and other negatively valenced emotions that modulate out-group antipathy, while at the same time removing the impediment to harsh treatment of the other that the recognition of the other's humanity erects. Exclusion from the moral community results in what Albert Bandura has called "moral disengagement,"[15] which allows individuals to treat out-group members in ways that are inconsistent with their humanity and which would otherwise trigger moral inhibitions. A similar moral disengagement function can be attributed to sanitized euphemisms, which are often coupled with parasite stress triggers—such as referring to mass murder as ethnic or political "cleansings" or "purges."

Ironically, the rhetoric of dehumanization is a back-handed tribute to a fundamental gain in inclusiveness: if most people did not regard other human beings, as such, as deserving of basic moral consideration, it would not be necessary to instill the belief that some people are subhuman in order to mobilize violence toward them. The use of dehumanization and contamination metaphors to foster intergroup hatred or to justify aggression toward out-groups, therefore, is an excellent example of a technique that causes people to regress toward the exclusivist moral responses that were more uniformly typical of human beings before the synergism of improved environmental conditions, open-ended normativity, and cultural innovation did their progressive work.

Mobilizers of ethnic and racial hatred exhibit an impressive working knowledge of both normal cognitive biases and social epistemology. They use existing social-epistemic institutions such as the media and government information agencies, as well as norms of epistemic deference to medical personnel, scientists, teachers, and in some cases clerics, in order to exploit

---

[15] Albert Bandura (2002), "Selective Moral Disengagement in the Exercise of Moral Agency," *Journal of Moral Education* 31: 101–119, p. 109.

normal cognitive biases (such as generic overgeneralization and responses to parasite threat) in order to activate exclusivist moral responses that dismantle culturally constructed bridges to inclusion.

Another historical example will reinforce this conclusion. In the Third Reich public schoolteachers were issued a teachers' manual in which they were instructed to teach children not only facts but also values. They were told to instill in their pupils the Golden Rule—an impressive thought experiment, which, if properly applied, can reduce the risk of exclusivist moral responses by encouraging one to put oneself in the other's place. However, this instruction came with an important proviso: that it was to be made clear to students that the Golden Rule only applies to racial comrades.[16] The teachers were also instructed to help students learn to distinguish racial comrades from inferior types and to understand just how dangerous and subhuman Jews in particular are. Here we have an example of a deliberate educational effort to disable a cultural innovation that fosters inclusiveness, in this case the Golden Rule. This effort proceeds, moreover, by exploiting the psychological dispositions that Hirschfeld, Leslie, Haslam, Bandura, and others identify, as well as the social-epistemic resources of the society in which it occurs—in particular, the patterns of deference to supposed experts, such as schoolteachers, who have an especially formative influence on the child's moral education.

Perhaps the clearest example of how the perception of out-group threat can dismantle culturally constructed inclusivity is the nationalist version of social Darwinism that appears to have played a significant role among the causes of the Second World War. According to this ideology, nations are locked in an inevitable struggle of unlimited violence in which the only alternatives are domination or subjugation and ultimately extinction. This

[16] Claudia Koontz, *The Nazi Conscience* (Harvard University Press, 2003, p. 119).

view gained popularity in the countries that came under the sway of fascism (Italy and Germany) and militarism (Japan) and, perhaps in response to the spread of the Great Depression through global trade and financial networks, was combined with a belief in economic autarchy. This is the view that a country must control within its own borders all the natural resources required for its economy to function or to function well enough for it to succeed in the Darwinian struggle against other nations.

There is an impressive social science literature that builds a strong case for the conclusion that leaders who accepted the nationalist social Darwinist claims about international relations were biased toward "preventive" aggression and that they typically attempted to justify striking first on grounds of perceived "necessity."[17] The necessity here is rational, though only conditionally so: given the requisite premises about the inevitability of violent conflict among nations and assumptions about the existential risk that attaches to losing, it is rational for each nation to attempt to strike first before its potential opponent becomes powerful enough to dominate. And given the economic autarchy view, one must engage in wars of aggression to command more and more resources, given the premise that if one does not do so, other nations will use them against one.

The hyper-realist picture of international relations painted by nationalist social Darwinism has been thoroughly exploded in the international relations literature for several decades now, and the doctrine of autarchy has disappeared from respectable economic discourse. What matters, however, is not whether these views are true but whether they are *believed* to be true. To believe them is, in effect, to believe that we are living in the harsh environment characteristic of the EEA, with this modification: the relevant groups are not small batches of hunter–gatherers or

[17] Ian Kershaw, *Fateful Choices: Ten Decisions That Changed the World, 1940–41* (Penguin, 2013, pp. 274, 277).

hunter–pastoralist tribes but nations.[18] Acting on their nationalist social Darwinist and economic autarchy views, the leaders of fascist Italy and Germany, along with their ally militarist Japan, disabled the existing institutional infrastructure for peaceful cooperation in international relations (including the League of Nations and the Hague Conventions) and thereby created an environment that more closely approximated their own distorted vision. Given the environmental sensitivity of human morality, it is hardly surprising that once the aggressors succeeded in creating a harsher, more dangerous international environment, it became difficult if not impossible for their opponents to cleave to their own inclusivist moral commitments. Indeed, war propaganda in the democracies often indulged in the same dehumanization techniques their enemies used, in part to rationalize barbaric actions against civilian enemy populations, as in the case of Allied terror bombing of German and Japanese cities.

Case Study: Eugenics

Reflection on the eugenics movements of the late nineteenth and early to mid-twentieth centuries further bolsters the conclusion that our naturalized theory provides valuable resources for understanding regression and for appreciating certain recurring commonalities among otherwise quite different cases of regression. The eugenics movement was in fact highly heterogeneous—there were positive and negative, radical and reformist, liberal and conservative eugenicists. Yet there were five widely held if not

---

[18] Nations are already examples of inclusiveness: they are "imagined communities" that manifest strong ties among veritable strangers. The destructiveness of nationalism when combined with social Darwinism illustrates an important point: developments that in themselves might be viewed as instances of progress, such as the transcendence of cramped "tribal" identities in favor of larger communal identities like nationality, need not be progressive, all things considered, depending upon what other moral developments have occurred or failed to occur.

universally endorsed themes: (1) that the most serious social ills, from poverty and crime to drunkenness, "promiscuity," and child neglect, are the deterministic result of a cluster of traits found in some "genetically inferior" individual human beings; (2) that these traits are hereditary and are inherited in a straightforward fashion; (3) that those human beings who have these clusters of hereditary traits are, as a result of some of the traits themselves, reproducing at a much higher rate than are people with "good" genes (or germplasm, to use the earlier term); (4) that private philanthropy and the welfare state are fostering the reproduction of individuals with these deleterious packages of traits by buffering them against evolutionary selection pressures that would otherwise have eliminated them from the gene pool; and (5) that if there is not a radical change in human reproductive patterns rather soon—that is, unless the higher reproductive rate of the people with deleterious genes is not stemmed—major social ills will worsen to the point where civilization itself is imperiled.[19]

Eugenic discourse fits the template we have delineated: a certain group (in this case, those with supposedly defective germplasm) is characterized as "other," as dangerous, and as the bearers of diseases (eugenicists talked of the vertical transmission of disease, from generation to generation, and described those with defective genes as agents of infection). Reflection on eugenic discourse also shows how exclusion can be, as it were, internalized: for eugenicists, the dangerous "others" are not foreigners, members of another society; they are among us and constitute a growing proportion of the members of our society. In a subsequent work we intend to explore in detail this phenomenon of the internalization of exclusion, that is, the ways in which discourse, individual

---

[19] For discussions, see Russell Powell (2015), "In Genes We Trust: Genetic Engineering, Eugenics and the Future of the Human Genome," *Journal of Medicine and Philosophy* 40(6): 669–695; A. Buchanan, D. Brock, N. Daniels, and D. Wikler, *From Chance to Choice: Genetics and Justice* (Cambridge University Press, 2001, chapter 2).

behavior, and social policy can cast certain groups within society as dangerous "others" or as having less than full basic moral status. We intend to build on the work in this volume to construct a naturalistic theory of ideology.

Historians of eugenics have emphasized that a "public health" model was central to eugenic thinking. The idea was that the vertical transmission of disease could not be stemmed by individual healthcare but required large-scale social policy changes, either to encourage the "fitter" types to reproduce more or to encourage or force the "unfit" to reproduce less (or preferably not at all) or both. Because it was thought that the disproportionate proliferation of deleterious genes would result in the destruction of civilization, eugenic thinking appealed to what Michael Walzer in a quite different context calls the idea of a "supreme emergency." In a lethal plague in which the transmission of disease is horizontal (from person to person existing at the same time), extraordinary measures, including policies that are coercive, may be necessary. To halt the spread of infection, individuals may have to be quarantined, travel prohibited, mandatory vaccination programs initiated, and so on. In brief, such a state of emergency may license infringements of individuals' rights that would be impermissible under ordinary conditions. Similarly, the eugenicists argued, the ordinary moral rules, including those implicating individual rights, are abrogable when the vertical transmission of disease threatens catastrophe for all of humankind.[20]

Eugenic discourse also manifests dehumanization techniques that play on the disgust response: those with deleterious genes are likened not only to plague-carrying vermin but also to sewage polluting the public water supply.[21] Just as important, eugenic rhetoric also appealed to another threat cue from the EEA: the danger of social parasites, free-riders, or "useless eaters" who will

soak up resources without contributing. Eugenic literature often featured elaborate calculations of how much social wealth would be absorbed by some number of generations of "defectives" and even went so far as to predict that unless such parasitism was stopped society would be reduced to a condition of primitive scarcity.[22]

Each of these instances of moral regression could be elaborated in greater detail, and no doubt our desire to present them concisely slips over certain nuances and complications. Nevertheless, our characterizations are sufficiently accurate to illustrate the power of our naturalistic theory to help illuminate at least some important forms of moral regression. Understanding the ways in which the EEA shaped human moral capacities helps explain both why regressive phenomena as different from one another as those we have described all appeal to certain ideas and metaphors—such as disease, scarcity, free-riding, and degeneration—and why such appeals are so motivationally potent.

What's New?

At this point, one might object that we have merely stated what everyone knew already—for example, that eugenicists, Nazis, and more recent genocidaires dehumanized their victims and that unscrupulous politicians foment conflict by playing on fears of "the other." It is important to emphasize, therefore, that the naturalized account of moral progress and moral regression that we have developed here is not an attempt to reinvent the wheel. It is true that some of the processes of social change we have focused on have already been characterized by psychologists, historians, and sociologists, for example, in the Holocaust studies literature. Our contribution is to provide an explanatory framework that *unifies*

---

[20] Allen Buchanan (2007), "Ethics, Beliefs, and Institutions: Eugenics as a Case Study," *Journal of Political Philosophy* 15(1): 22–45.
[21] Ibid.

[22] See, for example, *Eugenics Catechism* (American Eugenic Society, 1926), which calculates the social costs of the continued reproduction of the fictional Jukes family.

and deepens this diverse body of interdisciplinary work, relating it to evolutionary understandings of human nature and linking it to philosophical discussions of moral progress. In particular, we have shown how normal cognitive biases, existing patterns of epistemic deference, and evolved mechanisms of conditional moral expression can work together to produce forms of moral behavior that philosophers and other thinkers have characterized as regressive.

We have also shown that existing theories of racial and ethnonational behavior are not only consistent with the prevailing evolutionary explanation of the origins of human morality but in fact enriched by it. The naturalistic account of moral progress we have proposed is by no means "reductionistic" or "scientistic." It is no more fundamentally an evolutionary explanation than it is a social scientific, historical, or philosophical one. Its aim is to integrate evolved psychological mechanisms, cognitive biases, and social moral-epistemic practices into a dynamic developmental account of morality that does not reduce fundamentally to any one of these phenomena. Unlike views of human culture that could be seen as "biologically imperialistic," our account takes culture—and cultural innovation—seriously and conceives of human morality as only loosely constrained by its evolved genetic moorings.

We do not purport to offer an account that encompasses every important facet of moral progress or regression, let alone one that provides generalizable sufficient conditions for any instance of it; nor do we expect our account to explain every aspect of the instances of moral progress or regression to which it is applied. Our goal, rather, is to provide an empirically constrained and informed model that ties together a diverse range of observations about human moral thought and behavior by recourse to a few organizing principles and idealized causal mechanisms. This account does not merely restate a list of widely documented dispositions (e.g., intergroup violence is triggered by resource scarcity; individuals with disease and disability have often been excluded

from the moral community; altruism is modulated by group membership; people tend to form racial and ethnic stereotypes; dehumanization of the out-group can facilitate interethnic violence, etc.); nor does it simply repackage these observations in bio-conceptual garb. Rather, it brings these diverse phenomena under a unified causal-explanatory umbrella, with philosophical and scientific theories of human nature playing mutually informing roles.

Theories give data meaning. Observations only count as data in relation to some hypothesis, and what we perceive as data depends heavily on our background theories.[23] As Tooby and DeVore state, "Models (or theories) are organs of perception: they allow new kinds of evidence and new relationships to be perceived."[24] The model we propose not only explains known patterns of data and the links between them but also is likely to reveal entirely new sources of evidence that corroborate—or compel us to elaborate, modify, or abandon—elements of the model.

One might skeptically query whether any single observation could falsify our theory; but falsifiability is no longer treated as the gold standard for theory adjudication. The question, rather, is whether the proposed model adds to our explanatory toolkit. To be deemed inadequate, one would need to show that most of what we want to explain in the domain of moral progress and moral regression is not amenable to the explanatory tools our theory provides. To the contrary, as we have seen, the model offers a range of novel explanatory insights in this domain.

Further, our account takes the interaction between biology and culture seriously: it holds that threat cue detection can be faulty

[23] Niles Eldredge and Stephen Jay Gould, "Punctuated Equilibria: An Alternative to Phyletic Gradualism," in T. J. M. Schopf (ed.), *Models in Paleobiology* (Freeman, Cooper, 1972, pp. 82–115).

[24] J. Tooby and I. DeVore (1987), "The Reconstruction of Hominid Behavioral Evolution Through Strategic Modeling," in Warren G. Kinzey (ed.), *The Evolution of Human Behavior: Primate Models* (SUNY Press, 1987, pp. 183–237, p. 184).

yet still provoke exclusivist responses and that human beings can manipulate one another's beliefs in ways that create false positives for threat detection. At the most general level, these techniques for belief manipulation can be called cultural innovations, but our analysis goes further than that by emphasizing that whether manipulation succeeds will depend upon the social-epistemic environment—whether certain individuals have incentives for manipulating the beliefs of others and whether they can rely on effective communication technologies (from the printing press to Twitter) and exploit established patterns of epistemic deference to do so. The fundamental point is that the same cultural innovations that allow people to expand the moral circle can be used to dismantle the bridges that have with great effort and over a long period of time achieved successive expansions.

Evolutionary explanations of the origins of morality provide us with the beginnings of an account of what must occur if moral progress in the form of greater inclusiveness is to continue and be sustained. Our evolutionary model suggests that those who value this form of moral progress should support efforts to (1) alleviate the harsh conditions characteristic of the EEA wherever they still exist, (2) avoid regression to EEA-like conditions or to perceptions of those conditions where more favorable circumstances now prevail, and (3) prevent those who would mobilize exclusivist moral responses from using social-epistemic resources to dismantle the cultural innovations that have been instrumental in expanding the moral circle.

Concretely, the first task requires reducing the incidence of infectious disease, creating conditions of greater physical security in many parts of the world (including in microenvironments in developed nations), fostering economic development to increase social surpluses, and creating institutional structures that link groups in peaceful, mutually beneficial cooperation. The second and third tasks involve not only solidifying objective conditions that are friendly to the development of inclusivist morality but also protecting inclusivist cultural innovations against efforts to

dismantle them by those who create perceived conditions of outgroup threat. The protective effort will need to draw on the same resources that regressive forces utilize: knowledge of evolved human psychology (including normal cognitive biases) and an appreciation for how socially promulgated beliefs can influence our conditional moral responses (social moral epistemology). Any naturalized account of moral progress will therefore need to be informed by evolutionary psychology and social moral epistemology—not only to provide a realistic account of how (and how much) inclusivist moral progress is possible but also to supply practical guidance on how best to achieve and sustain the moral progress of which we are capable.

Evolutionary explanations of morality that stress the predisposition toward exclusivist morality do not show that inclusivist morality is impossible. Nor do they show that inclusivist gains made thus far have reached their limit or are unsustainable. They do indicate, however, that whether the gains made thus far will be sustained and whether further gains can be achieved depend on the environment in which our moral powers develop and operate. A key upshot is not simply that exclusivist morality is a predisposition rather than an inevitability. It is that the exclusivist predisposition is itself *conditional*: this disposition is only activated by certain cues that may or may not be present in the developmental evolutionary environment. In that sense, it is too strong to say that inclusivist morality goes against our evolved grain; instead, it is more accurate to say that *under certain conditions* inclusivist morality goes against our evolved grain. The task that lies before us is to spell out these conditions in greater detail.[25]

---

[25] At the outset of this inquiry, we noted that even if rejecting extreme forms of exclusivist morality is uncontroversially progressive, it should not be assumed that greater inclusivity is, even on a liberal account, always better. Nor should we assume, even if the adaptive plasticity account were right, that human moral capacities could be stretched indefinitely along the dimension of inclusivity without incurring significant moral costs. Therefore, a problem

Our analysis has scrupulously avoided any suggestion that some elite should, naturalistic theories in hand, take it upon themselves to guide humanity toward moral progress. The dangers associated with misuse of the notion of moral progress and with claims of moral expertise are amply illustrated in human history. Nonetheless, we believe that it is a mistake to respond to these dangers by refusing to explore the possibility of a naturalized theory of moral progress. The better course is to develop an account of how some of the most important putative instances of moral progress (and regression) have occurred and then, armed with that explanatory framework, address the question of how abuses of the notion of moral progress can best be avoided. Indeed, many such abuses can be understood (and perhaps ultimately mitigated) by recourse to an explanatory framework like the one we have sketched here. If it turns out that the risk of abuse is intolerably and unavoidably high, then perhaps "moral progress" should remain conspicuously absent in liberal discourse. Absent such a showing, however, we will continue to remain open to the possibility that a theory of moral progress may eventually reclaim its rightful place at the heart of liberal political theory.

---

remains: under what circumstances will human beings be able to determine when greater inclusiveness is progressive and when it is regressive? In particular, a theory of inclusivist moral progress should shed light on the circumstances in which the capacity for open-ended normativity is likely to be exercised in such a way as to give inclusivity its due without giving short shrift to special moral ties. Another important task is to spell out the implications of our thesis for attributions of moral praise and blame. If individuals live in an environment that is hostile toward sustaining inclusivist moral commitments, then their violation of inclusivist moral principles may be less blameworthy. It may still be the case, however, that such individuals have obligations to try to change the environment so that they are able to adopt and honor more inclusivist moral commitments.

CHAPTER 8

De-Moralization and the Evolution of Invalid Moral Norms

Thus far our naturalistic theory of moral progress has focused on moral inclusivity. However, as Part I makes clear, there are many other important types of moral progress—and we believe that human evolutionary history both constrains and enables progress in some of these dimensions, too. The present chapter illustrates this point by examining moral progress in the form of proper de-moralization, which occurs when behavior thought to be morally impermissible rightly comes to be seen as morally neutral or even commendable.

In what follows, we explain why proper de-moralization is a paradigmatic type of moral progress, why improper and even outright destructive moral norms evolve and persist, and how invalid moral norms can be identified and overcome.[1] We will also construct and critique another "evoconservative" challenge to moral reform, in this case one that appeals to cultural evolution in arguing that de-moralization is a risky, hubristic endeavor that is likely to have unintended bad consequences. Once again, we will show that these evoconservative assertions are fatally

---

[1] Arguments in this chapter are drawn from Allen Buchanan and Russell Powell (2017), "De-Moralization as Emancipation: Liberty, Progress, and the Evolution of Invalid Moral Norms," *Philosophy & Social Policy* 34(2): 108–135.