Toward a Naturalistic Theory of Moral Progress*

Allen Buchanan and Russell Powell

Early liberal theories about the feasibility of moral progress were premised on empirically ungrounded assumptions about human psychology and society. In this article, we develop a richer naturalistic account of the conditions under which one important form of moral progress–the emergence of more "inclusive" moralities– is likely to arise and be sustained. Drawing upon work in evolutionary psychology and social moral epistemology, we argue that "exclusivist" morality is the result of an adaptively plastic response that is sensitive to cues of out-group threat that are detected during development. We conclude with a blueprint for reinforcing and extending inclusivist progress.

I. MORAL PROGRESS AND EVOLUTION

A. The Idea of Moral Progress

One of the distinctive features of mainstream liberal thought from the Enlightenment through the nineteenth century was that it secularized the idea of moral progress.[1] A chief problem with early secular conceptions of moral progress, however, was that they claimed to be naturalistic but weren't. They articulated a notion of what moral progress consisted in that was independent of religious tenets, and they insisted that moral progress could be achieved solely by human effort and without divine assistance. But they did not fulfill this naturalistic promise. Instead, like their conservative detractors, liberal political theorists tended to rely upon underevidenced assumptions about human psychology and society. Their factual assumptions were based on folk psychology, flawed attempts to develop empirically based psychological theories, a priori speculation,

and reflections on history hampered both by a lack of information and inadequate methodology.[2] Another defect of some but not all secular conceptions of moral progress was that they claimed, without evidence, that moral progress was inevitable, not merely feasible. Given a near total lack of empirical grounding, the stronger claim that progress was inevitable was even shakier than the claim that it was feasible.

At a minimum, a political theory of moral progress ought to be compatible with psychological and social facts about human beings. A more demanding desideratum is that any such theory must provide an account of how the path of moral progress can be traversed. In particular, the theory must support the conclusion that moral progress is more than logically possible, given an accurate view of the relevant facts. It must show that moral progress is in some significant sense feasible, and it should also supply some guidance as to how moral progress can be achieved. It is too demanding to require that a theory of moral progress be capable of showing how some moral ideal can be fully realized. It may be enough if such a theory can show that substantial progress toward a given ideal is feasible in light of what we know about human beings and societies. Unfortunately, liberal philosophical accounts of moral progress have not met even this modest requirement.

The fact that earlier liberal thinkers failed to achieve naturalistic theories of moral progress is not surprising, of course, given how meager genuine scientific knowledge about human psychology and society was in the nineteenth and early twentieth centuries. Our situation today is more hopeful. For the first time, we are beginning to develop genuinely scientific knowledge about human nature, especially through the development of empirical psychological theories that take evolutionary biology seriously. In addition, the social sciences now provide better information about what sorts of social arrangements are feasible and sustainable and about how social norms arise and change. Accordingly, there are now compelling reasons to revisit the question of moral progress in the light of what we are coming to know about ourselves.[3]

2. Theorists of progress in the English and Scottish Enlightenment, including Hartley, Hume, Smith, and Kames, based their views on psychological assumptions (e.g., that "the association of ideas" was a fundamental feature of the human mind), but their psychological views, like their views about society, lacked rigorous empirical support. In addition, their psychological theories were seriously incomplete, because they lacked an understanding of the full range of what are now called "normal cognitive biases" and how these biases interact with culture to construct morally relevant beliefs—which we later show to be of crucial relevance for any empirically grounded theory of moral progress. For a valuable discussion of the psychological assumptions of these theorists and their bearing on the idea of progress, see Spadafora, *The Idea of Progress in Eighteenth-Century Britain*, 138–48, 163–66, 343–46, 151–52.

3. Needless to say, the effort to ground a theory of moral progress in science must not lapse into overconfidence about what knowledge science has achieved or about our ability,

In this article, we develop some of the philosophical and empirical foundations needed for a more robustly naturalistic theory of moral progress. Our modest "naturalistic" goal is to determine whether and how certain types of putative moral progress are possible, and to assess their limits, within the strictures of our best scientific understandings of human moral psychology and culture. We will focus on one dimension of moral progress, namely, the movement toward increasingly "inclusive" moralities, or expansions of the sphere of beings that are regarded as having moral standing or equal basic moral status. We focus on inclusivist moral progress both because it presents as a strong candidate for an important type of moral progress and because it appears to be in tension with key evolutionary understandings of human moral psychology. We attempt to resolve this apparent tension by developing a richer, dynamic account of the interplay between social and biological forces in the evolutionary development of human moralities. In so doing, we show how philosophical reflections on human morality and society can inform evolutionary theorizing about human nature, and vice versa. Further, we argue that contemporary philosophical theories of how moral progress occurs, which lean heavily on the efficacy of moral reasoning, are insufficient (and insufficiently naturalistic) not because moral reasoning is unimportant in driving moral progress, but because such theories fail to identify circumstances in which inclusivist moral progress can occur and be sustained—and, as a result, they fail to rebuff conservative charges that certain types of moral progress are not realistic for beings like us.

In Section I, we explore some complexities of the idea of moral progress and consider the challenges that evolved aspects of human moral psychology pose to the realization and durability of one important type of moral progress, namely, inclusivist moral shifts. We show that although current evolutionary psychological understandings of human morality do not, contrary to the contentions of some authors, support conservative ethical and political conclusions, they 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 evolutionary psychological and cultural hurdles. Nevertheless, we argue that major moral gains have been, and can continue to be, achieved by modifying the multifaceted cultural and institutional resources that shape moral development and define the moral community. However, taking into account evolved psychological obstacles to moral progress is critical if we are to consolidate and expand

---

at any given time, to identify the best science available. Any attempt to work out a naturalized theory of moral progress must take the scientific hypotheses on which it rests as only provisionally plausible and subject to revision.

upon our moral advances. Knowledge of these obstacles is also crucial if we are to avoid reversions to destructive "exclusivist" moralities that result from the dismantling of the moral-conceptual innovations and cultural achievements that underpin inclusivist moral commitments.

In Section II, we take steps toward developing a naturalistic account of inclusivist moral progress by mapping the complex interaction between evolved components of moral psychology and the social bases of belief, in the development of human moralities. Drawing upon work in evolutionary moral psychology, the psychology of normal cognitive biases, and social moral epistemology, we contend that exclusivist morality may be understood as the result of an adaptively plastic moral response that is sensitive to out-group threat conditions, and that out-group threat conditions are inferred from the presence of certain environmental cues that are detected during human development, such as signs of infectious disease, scarcity of resources, and (especially) enculturated beliefs about out-groups. Unpacking this complex norm of reaction will be crucial to understanding the evolved constraints on moral progress and how they have been, and may be further, overcome.

B. *Inclusivist Morality as a Type of Moral Progress*

A general theory of moral progress could take a more or a less ambitious form. The more ambitious form would be to ground an account of which sorts of changes are morally progressive in a normative ethical theory that is compatible with a defensible metaethics.[4] The less ambitious form would be to identify developments that are widely considered, on reflection, to be clear instances of moral progress and then to explore, on the basis of the best available knowledge about human nature and society, whether or under what conditions they are likely to occur, to be expanded upon, and be sustainable. In what follows we take the more modest path: we set aside metaethical challenges to the notion of moral progress, we make no attempt to ground the claim that certain moralities are in fact better than others, and we do not defend any particular account of what it is for one morality to be better than another. Instead, we assume that the emergence of certain types of moral inclusivity are significant instances of moral progress and then use these as test cases for exploring the feasibility of a naturalized account of moral progress.

There are many types of putative moral progress, ranging from better compliance with valid moral norms to improvements in moral concepts (including understandings of the virtues), moral motivations, moral reasoning, and even conceptions of morality itself. We develop a

---

4. For a sophisticated attempt to meet this more ambitious goal for a theory of moral progress, see Peter Railton, "Moral Realism," *Philosophical Review* 95 (1986): 163–207.

general theory of moral progress elsewhere.[5] Here, we will focus on one plausible type of moral progress: namely, what Peter Singer, borrowing from William Lecky, calls the "expanding circle" of moral concern,[6] or what we have elsewhere referred to as the emergence of "inclusivist moralities."[7] These are moralities that extend moral standing to all human beings and even to some nonhuman animals regardless of their group membership or strategic capacities (i.e., their ability to contribute to or disrupt cooperation). Moral progress in the form of increasingly inclusive moralities consists in expansions of the moral community—the class of beings who are recognized as having moral standing or equal basic moral status—beyond tribal boundaries and the mutually self-serving cooperative relationships between groups. Inclusivist moralities also reject restrictions on membership in the moral community based on gender, race, ethnicity, or species.

Singer can be read as holding that moral progress consists in such expansions of the moral circle.[8] However, this equation is mistaken, for several reasons. First, increasing moral inclusiveness is often but not always an indicator of moral progress. In certain circumstances, moral progress can plausibly take the form of exclusion, or contractions of the moral circle. This is true, for example, in relation to the moral reclassification of objects or entities that have no morally protectable interests, such as sacred artifacts, nonsentient organisms, or abiotic features of the environment like rivers or mountains—at least when according them moral status imposes unacceptable costs on morally protectable beings.[9] Furthermore, greater inclusiveness is not always good. Increases in the strength of inclusivist moral commitments could under some circumstances dilute commitments to fellow group members to the point that

5. For an extended discussion and analysis of types of moral progress, see Allen Buchanan and Russell Powell, "What Is Moral Progress?" (unpublished manuscript, Duke University).

6. Peter Singer, *Expanding the Moral Circle: Ethics, Evolution, and Moral Progress* (Princeton, NJ: Princeton University Press, 2011); William Edward Harpole Lecky, *History of European Morals from Augustus to Charlemagne*, vol. 1, 3rd ed. (New York: Appleton, 1921), http://oll .libertyfund.org/titles/1839 (Online Library of Liberty).

7. Allen Buchanan and Russell Powell, "The Limits of Evolutionary Explanations of Morality and Their Implications for Moral Progress," *Ethics* 126 (2015): 37–67.

8. A more charitable interpretation is that Singer remains agnostic as to whether there are other forms of moral progress. At any rate, he focuses only on "expanding the circle" or what we call inclusivist morality and does not discuss other forms of moral progress.

9. The moral risk of faulty exclusions, which result in "truncated" moralities, may be greater than the moral risk of faulty inclusions, which result in "promiscuous" moralities— since false negatives in relation to moral standing (treating individuals as if they do not have moral standing when in fact they do) will often be more harmful than false positives (treating organisms or entities as if they have moral standing when in fact they do not). Our point here, however, is that both inclusions and exclusions can amount to moral progress or moral regression, depending on the circumstances.

the latter commitments were unacceptably weak from a normative point of view. Indeed, the contemporary debate in political philosophy between cosmopolitans and liberal nationalists is a dispute not about whether all people are of equal moral worth, but about what is proper inclusiveness and what is not. In what follows we focus on examples of inclusiveness that are morally uncontroversial within a broadly liberal perspective and which will be regarded as progressive by cosmopolitans and liberal nationalists alike. We will use "inclusivist morality" only to signal moral attitudes and behaviors that extend moral regard or equal moral worth beyond the narrowest confines of the group, without prejudicing the question that divides cosmopolitans and liberal nationalists.

Another reason that Singer is mistaken if he holds that moral progress simply is the development of increasingly inclusivist moralities is that, as we noted above, there are several other types of changes in human morality—quite apart from expansions and contractions of the moral circle—that constitute prima facie cases of moral progress. Expansions of the moral circle may involve improved moral concepts (such as moral standing), improved moral reasoning (such as the logical extension of valid moral norms to cover individuals who had been arbitrarily excluded from their application), and improved compliance with valid moral norms (such as behavior that is in accordance with the equal basic moral worth of persons). However, there are many types of moral progress that do not involve expansions of the moral circle.

For example, one putative type of moral progress is "proper demoralization," which occurs when behavior that has wrongly been regarded as immoral comes to be seen as inherently morally neutral (examples include, e.g., premarital sex, masturbation, interracial marriage, homosexuality, profit seeking, and lending money at interest). Conversely, "proper moralization" occurs when some types of acts, such as torture and other forms of physical cruelty, are no longer viewed as generally permissible forms of punishment or coercion—or when behaviors once regarded as morally neutral, such as sexual harassment in the workplace, come to be regarded as morally impermissible. These types of moral progress do not implicate expansions of the moral circle. Neither do improvements in how moral virtues are understood, such as when a conception of honor that focuses almost exclusively on taking violent action against supposed insults gives way to one that stresses integrity and honesty and a reluctance to resort to violence.[10] Likewise, there are many important moral concepts apart from our notions of moral standing and moral statuses, and improvements in these concepts are also putative examples of moral progress. One such example relates

10. Kwame Anthony Appiah, *The Honor Code: How Moral Revolutions Occur* (New York: Norton, 2010).

to improvements in our conception of the domain of justice, which has expanded to include the amelioration of social-structural inequalities out of the recognition that institutions are human creations subject (under certain conditions) to modification. Such an expanded conception of justice has the implication that some forms of harm and inequality that were formerly thought to be natural and inevitable—and hence not subject to the constraints of social justice—are in fact within human control and thus potentially subject to moral evaluation.

While there are many types of moral progress that do not consist in expansions of the moral circle, our primary focus in this article is on the inclusivist dimension of moral progress—in part because we agree with Singer that it is a crucial and relatively uncontroversial type of moral progress, and in part because of the apparent tension between moral inclusivity and prevailing evolutionary theories of morality. A central goal of our naturalistic approach is to reconcile normative theorizing about moral progress with our best scientific understandings of human moral psychology and culture.[11]

We will argue that the best available knowledge from evolutionary psychology and the social sciences suggests that although the emergence and preservation of certain forms of inclusiveness face serious impediments, there is no good reason to believe conservative claims that the shift toward greater inclusiveness has reached its limit or is unsustainable. To the contrary, we show that a proper understanding of the psychological and social impediments to inclusivist morality provides resources that we can, and should, draw upon in order to consolidate important moral gains and to achieve further moral progress. Although our discussion will be centered around moral progress in the form of increasing inclusivity, much of what we argue will be applicable to other forms of moral progress as well.

C. Evolved Psychology as a Limitation on Inclusivist Morality

Attempts to draw connections between contemporary evolutionary theories of morality and the possibility of inclusivist moral progress begin with the standard evolutionary psychological assertion that the main contours of human moral capacities emerged through a process of natural selection on hunter-gatherer groups in the Pleistocene—in the so-called environment of evolutionary adaptation (EEA). On the prevailing evolutionary account, morality was selected for promoting productive cooperation within the group, which increased success in competition for resources between groups.[12] The crucial claim, which leads some thinkers to

11. See, e.g., Railton, "Moral Realism."
12. This occurred via a combination of kin, reciprocity, and/or group-level selection mechanisms. See, e.g., Peter Richerson and Robert Boyd, *Not in Our Genes* (Chicago: University of Chicago Press, 2005).

draw a pessimistic inference about the possibility of inclusivist moral progress, is that selection pressures in the EEA favored exclusivist moralities. These are moralities that feature robust moral commitments among group members but either deny moral standing to outsiders altogether, relegate out-group members to a substantially inferior status, or assign moral standing to outsiders contingent on strategic (self-serving) considerations.

The basic idea, therefore, is that human groups in the EEA evolved under ecological conditions that strongly 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,[13] with levels of productivity sufficiently low that sharing resources with out-groups entails dangerously high costs.

(2) The absence of institutions 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.[14]

(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,[15] 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.[16]

Such ecological conditions would have favored moralities underwritten by truncated forms of moral emotions. In particular, sympathy would have been restricted to in-group members, resulting in severe limitations on the capacity for altruistic behavior beyond the confines of the group.[17] Both theory and data suggest that parochialism was a precondition for the evolvability of human altruism, and there is prelimi-

13. Richard Wrangham and Dale Peterson, *Demonic Males: Apes and the Origins of Human Violence* (New York: Houghton Mifflin, 1996).

14. Chris Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press, 2001).

15. Corey Fincher and Randy Thornhill, "Parasite-Stress Promotes In-Group Assortative Sociality: The Cases of Strong Family Ties and Heightened Religiosity," *Behavioral and Brain Sciences* 35 (2012): 61–79; C. D. Navarrete and D. M. T. Fessler, "Disease Avoidance and Ethnocentrism," *Evolution and Human Behavior* 27 (2006): 270–82.

16. Richerson and Boyd, *Not in Our Genes*.

17. Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon, 2012).

nary evidence which suggests that altruism and parochialism evolved in tandem and are mediated by common proximate causes.[18]

What do evolutionary psychological accounts of morality tell us about the nature and plausibility of inclusivist moral progress? Thinkers from a variety of disciplines have taken these evolutionary accounts to support rather bleak prognoses regarding the plausibility of inclusivist moral progress.[19] These "evoconservatives," as we have labeled them, infer from evolutionary explanations of morality that inclusivist moralities are not psychologically feasible for human beings.[20] Some have gone so far as to argue that morally progressive institutional reforms, such as an international order reflecting cosmopolitan principles, are unrealistically utopian.[21] Moreover, liberal proponents of moral progress have done little to block this inference. Some moral philosophers, whom we have elsewhere called "evoliberals," have tacitly affirmed the evoconservative view in arguing that biomedical interventions that enhance human moral capacities are likely to be crucial for major moral progress due to evolved constraints on human moral nature.[22] Such authors have assumed that our evolved psychology is so morally feeble, and our capacity for other-regard is so limited, that the radical step of altering its biological basis may be required.[23]

In a previous article in this journal, we showed that moral progress in the form of inclusivist morality actually exists and plays an important role in human ecology (i.e., is not merely aspirational), even if there is a sense in which expansions of the moral circle go against the evolutionary grain.[24] Second, we showed that a naturalistic account of inclusivist moral progress cannot be derived in any straightforward way from the received selectionist explanation. This is because standard evolutionary

18. For example, studies show that oxytocin and other modulators of empathy enhance intragroup altruism while accentuating exclusivist moral response in competitive intergroup environments. See Herbert Gintis and Samuel Bowles, *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton, NJ: Princeton University Press, 2011).

19. Stephen Asma, *Against Fairness* (Chicago: University of Chicago Press, 2012), chap. 1, 45–46; Haidt, *The Righteous Mind*; Jack L. Goldsmith and Eric A. Posner, *The Limits of International Law* (Oxford: Oxford University Press, 2006); Larry Arnhart, *Darwinian Conservatism* (Charlottesville, VA: Imprint Academic, 2005); Francis Fukuyama, *Our Posthuman Future* (New York: Farrar, Straus & Giroux, 2002).

20. Buchanan and Powell, "The Limits of Evolutionary Explanations of Morality."

21. Goldsmith and Posner, *Limits of International Law*, 212.

22. For a discussion and critique of this view, see Russell Powell and Allen Buchanan, "The Evolution of Moral Enhancement," in *Human Enhancement: The Moral Challenge*, ed. Julian Savulescu, Steve Clarke, and Alberto Giubilini (Oxford: Oxford University Press, forthcoming).

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

24. Buchanan and Powell, "The Limits of Evolutionary Explanations of Morality."

accounts of morality are incapable of explaining these inclusivist trends either as adaptations, or as meaningful by-products of adaptations, whether these are taken to have emerged under EEA conditions or in response to more recent environments.

In what follows, we sketch 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. If this model withstands scrutiny, it will also enable us to reject another evoconservative/evoliberal claim: namely, 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.

D. Contemporary Accounts of Moral Progress and Their Limits

Before sketching such an account, it is helpful to see the explanatory gaps in existing philosophical theories of inclusivist moral progress. Thus, we will first show why recent theories of moral progress, though they provide part of the story, prove inadequate to the task of explaining inclusivist moral progress, and then we will show why these explanatory deficits translate into normative deficits.

Singer attributes the move toward greater inclusion to the human capacity to reason to moral truths even when these are not conducive to evolutionary fitness.[25] We agree with Singer that the capacity for reasoning will be an important part of any adequate explanation of inclusivist moral progress. More specifically, we think that such an explanation will assign a key role to what we call "open-ended normativity": the capacity to make explicit the norms one has hitherto been following and subject them to rational criticism and revision. One way this happens is when critical reflection leads to the recognition that existing

25. Katarzyna de Lazari-Radek and Peter Singer, "The Objectivity of Ethics and the Unity of Practical Reason," *Ethics* 123 (2012): 9–31; Singer, *Expanding the Moral Circle*, chap. 5. Enlightenment thinkers (including the French Encyclopedists) and nineteenth-century liberals who advanced doctrines of progress also exhibited a rather naive faith in the efficacy of reason, without fully appreciating the need to consider the particular social and other environmental factors that modulate its exercise. John Stuart Mill, for example, appears to have put too much faith in the efficacy of freedom of expression under conditions of widespread literacy, not sufficiently appreciating the ways in which cultural forces (e.g., media) and normal cognitive biases can interact to produce and disseminate false beliefs.

norms are being applied inconsistently or are arbitrarily restricted in their scope, which in turn provides reasons to revise them.[26] However, simply saying (as Singer does) that inclusivist moral commitments are a product of the human capacity for reasoning is not a sufficient basis on which to naturalize moral progress, for three reasons.

First, the capacity for reasoning is much older than the emergence of inclusivist morality, so reasoning alone is not sufficient for the emergence of inclusivist morality. Again, this is not to deny that reasoning is a crucial component of inclusivist trends—it is, rather, to say that the operation of reasoning of the right sort on a sufficiently large scale in the moral realm has only occurred under some conditions, and Singer's account is incomplete because it fails to consider or spell out these conditions. Second, at present the penetrance of inclusivist morality is quite uneven, with different human beings and different human cultures exhibiting inclusivist commitments to a greater or lesser degree; and yet there is no reason to believe that these interpersonal and intercultural differences in penetrance are the result of differences in the general capacity for reasoning. Third, the exercise of human reason can sometimes contribute to expansions of the moral circle, but in other cases it plays a significant role in contracting the circle in ways that lead to moral regression. This occurs, for example, when people judge that some human beings do not count morally on the basis of false premises about natural differences between groups of humans (such as blacks and whites or men and women), or due to mistaken ideas about which characteristics qualify one for equal moral status or for moral standing more generally. It also occurs when reasoning is used in an ad hoc or confabulatory way to justify preexisting moral judgments about out-groups that are motivated by negative affects like disgust, fear, or distrust. Thus, while Singer is right to say that reasoning plays an important role in the development of more inclusive moral commitments, he lacks an account of the developmental conditions under which such reasoning occurs and is likely to become sufficiently pervasive to result in large-scale moral progress notwithstanding resistance from evolved exclusivist tendencies. He also lacks an account of why the capacity for reasoning has often failed to be exercised in such a way as to achieve greater inclusiveness.

Just as reasoning functions differently in different social environments, so too does the capacity for self-scrutiny that open-ended normativity requires. In an insightful analysis, Michele Moody-Adams argues that a precondition for moral progress is that individuals be in a position

---

26. For a path-breaking analysis of how individuals come to realize that they hold inconsistent views about the morality of particular behaviors, and how they come to resolve these inconsistencies in morally progressive ways, see Richmond Campbell and Victor Kumar, "Moral Reasoning on the Ground," *Ethics* 122 (2012): 273–312.

to scrutinize their own values and the social practices that implement them: "One of the most important tasks of constructive moral inquiry," she argues, "is to . . . break down the common human resistance to self-scrutiny."[27] However, she does not specify the conditions under which this breakdown in the resistance to self-scrutiny is likely to occur, and thus her explanation of moral progress, like Singer's appeal to reason, is critically incomplete.

Richmond Campbell and Victor Kumar advance a much more detailed and empirically grounded account of one kind of reasoning that can result in moral progress: the identification and resolution of inconsistent moral responses, which they argue occurs through the interaction of intuitive and deliberative systems that guide moral judgment.[28] But, like Singer and Moody-Adams, they do not explain the conditions under which this is likely to occur. Nor do they provide an account of why people, and even entire cultures, can fail to identify moral inconsistencies or to work through them rationally, instead employing various cognitive-affective strategies to minimize the dissonance that arises from exposure to logically contradictory information, without revising their preexisting beliefs. As an instance of the latter, all-too-common phenomenon, consider the following case. In the racist culture of the American South, it was commonly thought that blacks were of inferior intelligence. When confronted with a black person who clearly showed high intelligence, there was a ready strategy for resolving the contradiction—not by discarding or reducing confidence in the generalization about the intelligence of blacks, but by explaining away the apparently disconfirming case: it was said that the black person in question "must have some White blood."[29]

Peter Railton develops a very different account of moral progress that aligns reasonably well with our notion of inclusivist moral progress, save for one crucial respect. Railton holds that moral progress occurs when the interests of certain group members in a society are disregarded, resulting in social frustration and alienation, which in turn leads to disruptions in the social order that prompt changes in moral norms and behavior that accommodate those previously discounted interests, and thereby restore social stability. On Railton's functionalist account, this

27. Michele M. Moody-Adams, "The Idea of Moral Progress," *Metaphilosophy* 30 (1999): 168–85, 175.

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

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

process of disruption, moral change, and equilibrium can lead to the emergence of progressive cosmopolitan norms that govern relations within and between peoples. A crucial component of Railton's theory of moral progress, which causes it to diverge sharply from ours, is that it requires that the individuals whose interests are disregarded are capable of some significant "degree of mobilization," with egalitarian changes in morality serving to restore social equilibrium.[30] This functionalist explanation of inclusivist shifts is inadequate because it overlooks the fact that expansions of the moral circle are often driven by a regard for individuals who have no strategic capacities that rise to a level that would prompt moral change. This is true, for example, in the case of moral protections for nonhuman animals and for vulnerable groups such as children and people with severe disabilities. It is also true for the proliferation of subject-centered human rights culture,[31] as well as for concerns about the global poor—individuals who have little capacity to disrupt the cooperation between wealthy nations that leaves them severely disadvantaged. Railton's theory of how expansions of the moral circle occur cannot account for these important cases of inclusivist moral progress.

A naturalistic account of how inclusive moral commitments emerge despite the evolved parochiality of human moral emotions, judgments, and norms cannot appeal solely to capacities for reason or self-scrutiny or to strategic self-serving relations between groups. It must identify the conditions under which reasoning capacities are capable of fostering inclusivist commitments, even in the absence of strategic motivations.[32] This, in turn, requires understanding how the capacity for cultural innovations can create developmental environments in which moral reasoning and self-scrutiny can flourish—and how this cultural scaffolding can be dismantled in ways that lead to moral regression.

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

30. Railton, "Moral Realism," 194–95.
31. See Allen Buchanan, "Justice as Reciprocity versus Subject-Centered Justice," *Philosophy and Public Affairs* 19 (1990): 227–52.
32. One might assert that inclusivist morality is a dispositional trait that humans have long (or always) possessed, but that the conditions necessary for its expression only manifested quite recently in human history. Even if one were content to describe such highly flexible, nonselected behaviors as conditionally expressed traits (which we are not), our point is that any naturalized account of inclusivist moral progress would need to identify the difference-making conditions for their expression.

derstandings of moral psychology and culture, bringing a diverse body of interdisciplinary research under a single, unified causal-explanatory framework.

II. NATURALIZING MORAL PROGRESS

In the remainder of this article, we lay the groundwork for a fuller naturalistic theory of inclusivist moral progress. We first sketch an evolutionary developmental account of exclusivist morality, and then advance three further hypotheses:

> (H1) Inclusivist morality is a luxury good in the sense that it is only likely to be widespread 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.

Our aim will not be to make a decisive case for each of these three hypotheses but simply to show that they offer a plausible starting point for a naturalistic theory of inclusivist moral progress.

Although any plausible evolutionary explanation of the emergence of morality does not tell us that inclusivist gains are merely aspirational, or that they must decay, or that they have already reached their limit, when properly fleshed out it does tell us part of what we need to know about how to increase the probability that moral progress will persist and grow. The remaining part of what we need to know, we will contend, is supplied by the integration of evolutionary theory with social moral epistemology and the psychology of normal cognitive biases. 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.[33] Social moral epistemology focuses on the social promulgation of beliefs that tend to be crucial for moral judgment, moral

33. 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.

reasoning, and the moral emotions.[34] Below we will examine how social epistemic practices regarding morally crucial beliefs interact with cognitive biases and other evolved features of human moral psychology to impede or facilitate the development of inclusivist morality.

A. The Evolutionary Development of Exclusivist Morality

If we assume that the standard evolutionary account of the emergence of morality is more or less correct as far as it goes, then we may conclude that it is unrealistic to expect inclusivist morality to be widespread and sustainable in circumstances that mimic the harsh conditions of the EEA. There are both ontogenetic and evolutionary reasons, respectively, for this conclusion—that is, reasons that advert to the way in which exclusivist morality develops in individuals (ontogeny), and reasons that advert to the way in which exclusivist moralities are selected for and inclusivist moralities eliminated over longer timescales (evolution).

Understanding the evolutionary development of exclusivist morality requires understanding the role of group identity in normal human moral psychology. It is well established that in-group/out-group biases are among the most cross-culturally robust of human psychological traits, and these biases can easily be manipulated in laboratory and field study investigations.[35] At the same time, there is strong evidence that the development of exclusivist moral tendencies—or what evolutionary psychologists refer to as "in-group assortative sociality," which is associated with ethnocentric, xenophobic, authoritarian, and conservative psychological orientations—is sensitive to environmental cues that are indicative of out-group threat. Such threats may relate to the transmission of infectious disease,[36] competition over scarce resources, external physical dangers,[37] or beliefs and practices that are dissonant with in-group values and thus imperil group cohesion.[38] Cues that are associated with out-group threat

34. Allen Buchanan, "Social Moral Epistemology," *Social Philosophy and Policy* 19 (2002): 126–52.

35. Donald E. Brown, *Human Universals* (New York: McGraw-Hill, 1991). It is well known that the mere fact of group membership, even when the groupings are temporary and essentially meaningless, can generate these psychological dynamics. Henri Tajfel and J. C. Turner, "The Social Identity Theory of Intergroup Behavior," in *Psychology of Intergroup Relations*, ed. Stephen Worchel and William G. Austin (Chicago: Nelson-Hall, 1986), 7–24.

36. See Fincher and Thornhill. "Parasite-Stress."

37. See Blake M. Riek, Eric W. Mania, and Samuel L. Gaertner, "Intergroup Threat and Outgroup Attitudes: A Meta-Analytic Review," *Personality and Social Psychology Review* 10 (2006): 336–53; Muzafer Sherif and Carolyn Wood Sherif, "Ingroup and Intergroup Relations: Experimental Analysis," in *Social Psychology*, ed. Muzafer Sherif and Carolyn Wood Sherif (New York: Harper & Row, 1969), 221–66.

38. Miles Hewstone, Mark Rubin, and Hazel Willis, "Intergroup Bias," *Annual Review of Psychology* 53 (2002): 575–604.

have been shown to trigger negatively valenced moral emotions, such as fear, anger, and disgust, which in turn lead to aversive intergroup attitudes and behaviors. In contrast, exclusivist moral tendencies are attenuated in populations inhabiting environments in which cues of out-group threat are absent.

An evolutionary explanation for these patterns is that exclusivist moral psychology is what is known in biology as an "adaptively plastic" trait. Selective environments can change rapidly over time and space: they can vary from generation to generation, and over geographical ranges in spatially distributed populations. The benefit of adaptive plasticity is that it enables a lineage to achieve an adaptive match across more environments than would be possible if it produced a single phenotype in all environments. Adaptively plastic traits allow organisms to conditionally 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. For instance, some water fleas develop armor (defensive spines and helmets) only if they detect the chemical correlates of a predator in the water in which they develop.

According to an account of exclusivist morality as a conditionally expressed (adaptively plastic) trait, the suite of attitudes and behaviors associated with exclusivist tendencies develop only when cues that were in the past highly correlated with out-group threat are detected. But like the development of elaborate flea armor, which is physiologically expensive and can impede important functions, the development of exclusivist moral tendencies has costs. These include the opportunity costs that flow from forgoing the benefits of material trade, mate exchange, and alliances with neighboring groups, as well as the risk of dangerous, belligerent interactions with foreigners. As a result of this evolutionary trade-off, the adaptive plasticity hypothesis predicts (and there is evidence to support the claim) that exclusivist tendencies will be tempered in environments in which out-group threats are not detected.

"Detection" in adaptively plastic systems is essentially an evolutionary epistemic heuristic that picks up on cues that were in the past statistically associated with the presence of a particular ecological regime, such as a predator-rich environment. This can involve sampling of the environment for chemical cues of (e.g.) predators, as in the case of the water flea, but it can also involve the inspection of predatory types, motivations, and behavior.[39] Importantly, in the case of beings with

---

39. Thomas J. DeWitt, Andrew Sih, and David Sloan Wilson, "Costs and Limits of Phenotypic Plasticity," *Trends in Ecology and Evolution* 13 (1998): 77–81.

language, detection 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 (more on this in Sec. II. *C*).

Fincher and Thornhill offer an adaptive plasticity hypothesis to explain the robust, cross-cultural link between in-group assortative sociality and parasite stress.[40] We suggest expanding this account to incorporate other cues that are indicative of out-group threat, such as competition for scarce resources and, especially, socially constructed beliefs about out-groups. We argue below that social epistemic practices and 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. The precise developmental pathway through which detections (whether veridical or not) lead to the relevant plastic moral response is unknown; all that matters for purposes of this article, however, is that such pathways exist and are activated with some reliability.

Although adaptive plasticity is common in nature, it is not the standard way through which selection produces an adaptive match between organism and environment. In most cases, the adaptive match is achieved not by detecting cues in the environment during development, but via a process of natural selection that results in a trait that is adaptively canalized (i.e., does not have alternative character states that depend on specific environmental cues for their development). If environments hostile to the development of inclusive moralities are stable for a sufficiently long time span (as they likely were for much of human history), then (ceteris paribus) exclusivist moral tendencies and their

40. Fincher and Thornhill, "Parasite-Stress." The fact that a disposition is exhibited by a wide range of human cultural populations does not prove that there has been selection for that response profile. Nevertheless, support for an adaptive explanation of in-group assortative sociality comes from fields as wide-ranging as paleontology, archeology, ethnography, evolutionary anthropology, and social and experimental psychology, which demonstrate both the trait's ubiquity and its ecological fit. Fincher and Thornhill's work is the most rigorous to date to test a specifically "plastic" account of adaptive in-group assortative sociality. However, to make a definitive case for adaptation that goes beyond an empirically constrained "how possibly" explanation, we would need direct observations of fitness differences among competing groups that vary in their tendencies toward group-ishness, and such observations are epistemically inaccessible for practical and ethical reasons. Modeling to determine whether observed traits match theoretically optimal values provides a possible, albeit still inconclusive, avenue for testing adaptive hypotheses. The best we are entitled to, therefore, is a highly plausible, empirically supported inference of adaptation.

associated cultures will be stabilized, and inclusivist tendencies and cultural innovations driven to extinction, by selection. We might call this the "adaptive canalization" account of the origins and maintenance of exclusivist morality.[41]

For example, in environments in which there is severe competition for resources, in which the threat of foreign biological and/or social parasites is high, in which there is no institutional infrastructure for peaceful, mutually beneficial cooperation among the groups, and in which low productivity makes sharing resources with out-group members dangerously costly, there will be selection for moral tendencies and norms that help to ensure fairness and altruism toward members of one's own group but that permit or even require harsh treatment of members of other groups. In contrast, if all of these conditions are alleviated, better treatment of out-group members becomes evolutionarily "affordable," and possibly even fitness enhancing, given the benefits of prosocial intergroup relations in such favorable environments. A moral framework that allowed one to think of out-group members as objects of respect would facilitate cooperative interactions (such as opportunities for trade and alliances) that would not be available to those with highly exclusivist moral tendencies. Such reciprocal attributions of moral standing may have been the evolutionarily plausible beginnings of more inclusivist moralities.

Yet as we remarked earlier, we do not believe that a biological or cultural selection process has driven the trend in inclusivist moral progress. We maintain, rather, that the development of inclusivist moralities has become possible because selection pressures against inclusivist morality have been relaxed, and because conditions that trigger the adaptively plastic exclusivist moral response have been substantially ameliorated.

B. Inclusivist Morality as a Luxury Good (Hypothesis 1)

Although 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) robust health-care infrastructures, reductions in crime, rule of law, property rights, literacy,

41. As we shall see, the fact that human moral tendencies can change so rapidly and predictably, and exhibit so much variation even within a single geopolitical range, speaks strongly in favor of the "adaptive plasticity" account (and against the "adaptive canalization" theory) of exclusivist morality.

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 regarding exclusivist morality is right, then moral progress and the above sociopolitical developments do not merely have a common cause: rather, 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.[42] Individuals exhibiting these psychological orientations are less able or willing to acknowledge that they hold logically contradictory beliefs and less motivated to iron out logical contradictions within their belief system. This, in turn, makes it difficult or impossible to subject one's values and cultural practices to critical scrutiny (Sec. I.*D*), thus impeding inclusivist moral development and perhaps moral progress more generally. Although we have focused on moral inclusivity, insofar as other forms of moral progress (such as improvements in our understandings of virtues, moral concepts, and morality itself) plausibly rely on effective moral reasoning, they too may be luxury goods.

By the same token, cultural innovations that alleviate conditions that trigger exclusivist responses act to break this 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 is true, for example, of the division of labor and improvements in agricultural technologies that greatly increased the social surplus, the development of institutional infrastructures for peaceful, mutually beneficial cooperation among groups, and improvements in medicine and public health that dramatically reduced parasite stress. Second, cultural innovations in the form of new moral norms and more sophisticated moral reasoning can reshape moral responses, but only if economic conditions are sufficiently favorable and there is a reasonable degree of physical security. It is beyond the scope of this article 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.

Consider, first, a familiar phenomenon in contemporary American and European culture. Many people in these regions have overcome or

42. John T. Jost, Jack Glaser, Arie W. Kruglanski, and Frank J. Sulloway, "Political Conservatism as Motivated Social Cognition," *Psychological Bulletin* 129 (2003): 339–75.

at least significantly moderated their disgust reactions toward homosexuals, and this attitude grows more prevalent with each succeeding generation. Once homosexual behavior was no longer criminalized, it became more feasible for gays and lesbians to be open about their sexual preferences, and people began to realize that there were significant numbers of nonheterosexuals in their society. It became easier to see gays and lesbians as people who were "just like us" in virtually all other dimensions of personality and social interaction. Many people eventually came to view their disgust reactions as irrational and—perhaps more importantly—as inflicting undeserved stigmatization on gays and lesbians and contributing to wrongful discrimination against them. In this instance, certain cultural innovations—namely, changes in the law, fostered by judicial review of legislation in the light of an increasingly expansive legal understanding of unconstitutional discrimination (with gender and eventually sexual orientation coming to be considered constitutionally protected classes)—helped create conditions under which people could experience gays and lesbians in a new light and come to see that penalizing and stigmatizing them for their sexual preferences was inconsistent with their avowed values.[43] Similar accounts could be given for the rights of women to vote and to compete on merit for coveted professional and social leadership positions. These examples suggest that reasoning can alter moral judgments when it operates in certain inclusivist-friendly institutional environments and is guided by moral norms that are internalized via early enculturation.

A second example, or rather a set of examples, of cultural innovations that contributed to increased inclusivity, is documented by historians of the 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 a pervasive racist ideology. To accomplish this they employed a number of techniques that evidenced a remarkable grasp of both human psychology and what we now call 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 demonstrated their rationality. They also developed sophisticated techniques for evoking sympathy for the suffering of slaves. For example, antislavery 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 addi-

---

43. This discussion of changing responses to homosexuals is an elaboration of one provided by Campbell and Kumar, "Moral Reasoning on the Ground," 287.

tion, taking advantage of existing norms of epistemic (and moral) deference to clergymen, they worked to win over parish priests and even provided them with "canned" antislavery sermons.[44] Perhaps the most important cultural innovation that contributed to the success of abolitionism was the printing press, along with the great increase in literacy in the decades prior to the founding of the movement—which dramatically amplified the effects of the aforementioned cultural innovations.[45]

Many abolitionists apparently were motivated in part by the belief that slavery was incompatible with Christianity, but their strategies for mobilizing antislavery 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.

For a third example, consider the techniques employed by advocates for the better treatment of animals. These include the distribution of films 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 and at the same time reveal inconsistencies in our moral responses and behavior.

---

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

45. 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 previous 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* (Cambridge, MA: Harvard University Press, 1997).

The key point is that these inclusivity-advancing cultural innovations are only likely to arise 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 nonhuman 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.[46] Second, periods of severe economic downturn correlate with increases in xenophobic and racist behavior, particularly when out-groups (including minorities within larger populations) are salient.[47]

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.[48] 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.[49]

C. Naturalizing Moral Regression: Cognitive Biases and Perceptions of Out-Group Threat (Hypothesis 2)

Let us now turn to our prediction that inclusivist gains will tend to be eroded if EEA-like conditions return or if individuals come to believe such conditions exist. This hypothesis gains plausibility from the same evidence that supports hypothesis 1, but also from the fact that, as we have suggested, exclusivist moral responses that were selected for in the EEA

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

47. Lincoln Quillian, "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-immigrant and Racial Prejudice in Europe," *American Sociological Review* 60 (1998): 586–611. For reviews of empirical connections between perceptions of out-group threat and negative intergroup attitudes, see Riek, Mania, and Gaertner, "Intergroup Threat and Outgroup Attitudes"; and Jost et al., "Political Conservatism as Motivated Social Cognition."

48. Ervin Staub, "Genocide, Mass Killing and Violent Conflict: Prevention and Reconciliation," in *The Oxford Handbook of Intergroup Conflict*, ed. Linda Tropp (Oxford: Oxford University Press, 2012), 274–90.

49. Howard Markel and Alexandra Minna Stern, "The Foreignness of Germs: The Persistent Association of Immigrants and Disease in American Society," *Milbank Quarterly* 80 (2002): 757–88.

can be triggered by people's perceptions of their predicament. For the exclusivist moral response to be activated, it is not necessary that such perceptions are 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, they 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 EEA.

If the manipulators of moral exclusivist tendencies can succeed in making enough people believe that out-groups pose a lethal threat, those people will tend to act in ways that induce reciprocal fear in out-groups—and what began as a misperception will rapidly become reality. 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 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 (e.g.) cold war brinkmanship.[50] 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.[51] 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

50. Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976), 58–62.

51. See Brock Bastian and Nick Haslam, "Excluded from Humanity: The Dehumanizing Effects of Social Ostracism," *Journal of Experimental Social Psychology* 46 (2010): 107–13.

grounds for crime or disease, which in turn are used to justify their social exclusion.

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.[52] In simplest terms, children tend to sort the human beings they encounter or hear about into groups and to 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 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.[53] 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-induction 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.[54] Indeed, what is disturbing about generic overgeneralization is that it apparently only applies in connection with highly negative behavior.[55] If a member of another group exhibits commendable

52. Lawrence A. Hirschfeld, *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds* (Cambridge, MA: Bradford, 1998).

53. Sarah-Jane Leslie, "The Original Sin of Cognition: Fear, Prejudice and Generalization," *Journal of Philosophy* (forthcoming).

54. Miles Hewstone, "The 'Ultimate Attribution Error'? A Review of the Literature on Intergroup Causal Attribution," *European Journal of Social Psychology* 20 (1990): 311–35.

55. Leslie, "The Original Sin of Cognition."

behavior, we 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.[56] 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. Given the paucity of social practices or institutions for mutually beneficial interactions with strangers, and high levels of biological and social parasite threat, a false judgment that a stranger was innocuous could be lethal to the in-group—and thus would have entailed far greater risks than a false judgment that a stranger was dangerous, which would merely have entailed lost opportunities from forgoing prosocial interactions. 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.

Thinking of out-groups as natural kinds and attributing their negative 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.[57] Because co-ethnics share many properties that are not evident from superficial inspection, generic overgeneralization may support adaptive inductions without the risks that attend intergroup interactions.

D. Manipulating the Social Construction of Belief (Hypothesis 3)

We have suggested that inclusivist shifts are the result of normative 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 these moral bridges can be dismantled using the same materials that were used to construct them: human psychology and culture. We now want to focus on how some

---

56. Martie G. Haselton and Daniel Nettle, "The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases," *Personality and Social Psychology* 10 (2006): 47–66; Martie G. Haselton, Daniel Nettle, and Paul W. Andrews, "The Evolution of Cognitive Bias," in *The Handbook of Evolutionary Psychology*, ed. David M. Buss (Hoboken, NJ: Wiley & Sons, 2005), 724–46.

57. Ethnic "essentializing" may implicate the same cognitive faculties that identify and characterize biological species. See Francisco J. Gil-White, "Are Ethnic Groups Biological 'Species' to the Human Brain?" *Current Anthropology* 42 (2002): 515–54.

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 one 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."[58] Dehumanization is one type of "delegitimizing belief,"[59] a class of extremely negatively valenced attributions to out-groups that serve to exclude them from the moral community.[60] Dehumanization involves classifying out-groups as subhuman, either by identifying them as nonhuman 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 them. 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 (e.g.) 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. From the standpoint of manipulating beliefs in order to trigger 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

---

58. Nick Haslam, "Dehumanization: An Integrative Review," *Personality and Social Psychology Review* 3 (2006): 252–64; and Jonathan Glover, *Humanity: A Moral History of the Twentieth Century* (New Haven, CT: Yale University Press, 1999).

59. Daniel Bar-Tal, *Shared Beliefs in a Society: Social Psychological Analysis* (Thousand Oaks, CA: Sage, 2000), 121–22.

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

the other's humanity erects. Exclusion from the moral community results in what Albert Bandura has called "moral disengagement,"[61] 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 also often coupled with parasite stress triggers (e.g., references to ethnic or political "cleansings" or "purges").

In one sense, 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 sanitized 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 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 normal cognitive biases (such as generic overgeneralization and responses to parasite threat), with the aim of activating 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.[62] 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 dis-

---

61. Albert Bandura, "Selective Moral Disengagement in the Exercise of Moral Agency," *Journal of Moral Education* 31 (2002): 101–19, 109.

62. Claudia Koontz, *The Nazi Conscience* (Cambridge MA: Harvard University Press, 2003), 119.

able 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 view gained popularity in the countries that came under the sway of fascism (in Italy and Germany) and militarism (in 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 a social science literature that builds an impressive 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 "necessity."[63] 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 one's potential opponent becomes powerful enough to dominate. And given the economic autarky 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 autarky 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-

---

63. Ian Kershaw, *Fateful Choices: Ten Decisions That Changed the World, 1940–41* (London: Penguin UK, 2013), 274, 277.

gatherers or hunter-pastoralist tribes, but rather nations.[64] 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 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, the democracies, in war propaganda, 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.

It is worth emphasizing 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 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 conditional moral dispositions can work together to produce forms of moral behavior that philosophers and other thinkers have characterized as progressive (and 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 are 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. We do not aim to provide an explanation of moral progress from the bottom up, but rather to integrate evolved psychological mechanisms, cognitive biases, and social moral epistemic practices into a dynamic

---

64. 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 such as nationality, need not be progressive all things considered, depending upon what other moral developments have occurred or failed to occur.

developmental account that does not reduce fundamentally to any one of these phenomena.

We do not purport to offer an account that encompasses every important facet of moral progress, nor do we expect the account we provide to explain every aspect of the instances of moral progress it endeavors to explain. Our aim, rather, is to provide an empirically constrained 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 ethnic conflict, etc.), or simply repackage these observations in biological terms. It brings them under a unified causal-explanatory umbrella with respect to which philosophical and scientific theories of human moral nature play mutually informing roles.

III. FUTURE DIRECTIONS

Evolutionary accounts 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. This account 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 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 out-group 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 our 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, depends 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: it is only activated by certain cues that may or may not be present in the developmental 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.[65]

We have 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, we believe, 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

---

65. 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 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.

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.

We suspect that many contemporary liberal political thinkers have faith in moral progress, even though they rarely speak of it. By taking steps toward a naturalized theory of at least one important dimension of moral progress, we have shown that their faith can be replaced with rational hope.