

THE EVOLUTION OF MORAL PROGRESS

A BIOCULTURAL THEORY

ALLEN BUCHANAN
AND
RUSSELL POWELL

OXFORD
UNIVERSITY PRESS

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

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

CHAPTER 8

De-Moralization and the Evolution of Invalid Moral Norms

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

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

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

oversimplified and that the same evolutionary resources that lead to the entrenchment of invalid moral norms can be marshaled to break free of their hold. Here, as in preceding chapters, the aim is to offer not a purely evolutionary theory of moral progress but a *biocultural* one. Evolutionary theory will play two roles in the analysis to follow: first, we will show that certain conservative admonitions against de-moralization rest on a faulty understanding of cultural evolution; second, we will show how sound evolutionary thinking can illuminate the emergence and persistence of invalid (and in some cases highly destructive) moral norms and thereby provide guidance for the kinds of cultural innovations that can help us escape their thrall.

Invalid Moral Norms as Constraints on Liberty

Although morality necessarily involves constraints on liberty, people can mistakenly believe that morality constrains them when it does not. Liberal thinkers of the Enlightenment understood that these “surplus moral constraints” (as we will call them), in the form of invalid moral norms, can impose significant limitations on freedom and flourishing and ought to be overcome.² Surplus moral constraints have both internal and external aspects. Internally, they amount to limitations on an individual’s liberty imposed by conscience. Internal constraints of conscience may be accompanied by external constraints, including not only various sanctions (including punishment) for violating the moral norms in question but also subtler but nonetheless powerful forms of

² The fact that abandoning a moral norm would increase liberty does not, of course, show that this change constitutes moral progress. Abandoning *valid* moral norms might increase liberty but would not be progressive. The topic of this chapter is proper de-moralization—abandonment of invalid moral norms. So far as invalid moral norms constrain liberty, they do so without justification, and removing these constraints counts as moral progress, other things being equal, for two reasons: first, because it is a case of remedying a defective understanding about what morality requires and, second, because (at least from a liberal standpoint) unjustifiable constraints on liberty are to be avoided.

social pressure. Contemporary liberal thinkers have tended to focus chiefly on external constraints and in particular on curtailments of liberty wrought by the coercive power of the state.³ Yet internal constraints of conscience may persist and continue to restrict freedom, even when external sanctions and social pressures have abated. Internalized improper moralization can thus significantly limit an individual’s options for acting and in ways that entail great material and psychological costs, even when external sanctions have been removed.

Emancipation from surplus moral constraints is an important type of moral progress, at least for any conception of moral progress that values liberty. Because surplus moral constraints are unnecessary limitations on liberty, escaping their thrall is a form of emancipation. To the extent that surplus moral norms significantly interfere with liberty, welfare, or other important moral values or duties (such as those associated with justice or beneficence), the de-moralization of those norms, all else being equal, constitutes moral progress.

Instances of proper de-moralization abound and play a prominent role in the catalog of morally progressive developments. Profit-seeking, lending money at interest, premarital sex, homosexual behavior, interracial marriage, masturbation, refusal to die “for king and country,” and virtually all instances of resistance to government authority were once widely thought to be immoral but are no longer so regarded by many people. We will take it for granted, because we are assuming a broadly liberal moral perspective, that these are all cases of *proper* de-moralization—that, at least from a secular liberal point of view, beliefs that these behaviors are morally wrong *per se* or that they warrant

³ Focusing only on external constraints not only obscures the fact that invalid moral norms, if internalized, can unnecessarily limit liberty; it also abets a failure to see that false factual beliefs can limit liberty and at great cost. Allen Buchanan, “Prisoners of Misbelief: The Epistemic Conditions of Freedom,” in David Schmidtz and Carmen Pavel (eds.), *Oxford Handbook of Freedom* (Oxford University Press, forthcoming, pp. 508–524).

institutionalized punishment were unjustified and that coming to realize the falsity of these beliefs is an instance of moral progress. Later, we will suggest that one of the major points of contention between liberals and conservatives is a disagreement about how reliably one can determine when a given instance of de-moralization is a case of *proper* de-moralization. If one is to develop an account of the relationship between liberty and de-moralization, one must be able to determine when internalized moral norms and external sanctions for violating those norms are instances of surplus constraint and when they are not.

Paradigmatic cases of de-moralization present as clear cases of moral progress in the liberal tradition. For example, if one is homosexual but has internalized a norm that brands all homosexual behavior as sinful and morally abhorrent and if one adheres to this norm, then one may experience great psychological suffering due to the self-inflicted frustration of one's most basic needs. These needs include not just sexual satisfaction but also the intimacy and deep attachment of partnership—needs that can be frustrated even after homosexual behavior has been decriminalized and other formal external sanctions have been removed; and if the internalized norms of conscience have sufficient psychological inertia, these needs may continue to be denied even in the absence of any external constraints at all. Alternatively, if a person violates the moral prohibition of homosexuality and engages in homosexual acts in order to meet these basic human needs, then he may experience haunting shame and guilt.

Similarly, if one believes that any perceived insult to one's honor requires violent retaliation, one may put oneself at lethal risk by initiating a duel or feel compelled to engage in otherwise violent behavior that runs contrary to one's basic values, thereby risking self-alienation, unnecessary trauma, and guilt. Likewise, if people in a society refrain from profit-seeking or from lending money at interest on the grounds that these vital economic behaviors are immoral, the result may be the perpetuation of a state of economic underdevelopment, with disastrous consequences for

human welfare, liberty, and justice. Some have argued that the persistence of moral norms against profit-based lending partly explains the economic underdevelopment of the Middle East.⁴

One final example: in a society in which it is widely believed that it is wrong for women to engage in independent economic activity outside the home, women will not only be barred from important paths to flourishing but also remain so dependent upon men and have so little influence on the political process that there may be little prospect of eliminating the grosser abuses of patriarchal society, including domestic abuse and honor killings. In short, the costs of surplus moral constraint, and accordingly the benefits of proper de-moralization, can be extraordinarily high. That is why proper de-moralization is an important form of moral progress.

Why Do Invalid Moral Norms Evolve and Persist?

At this point, a puzzle looms: if some supposed moral constraints are so costly to obey because they are clearly irrational, destructive, or bigoted, then why did they come about in the first place and why do they persist? Consider, for example, biblical prohibitions on planting more than one kind of crop in a field or wearing garments with more than one kind of fiber, on women trimming the edges of a man's beard, on simmering a young goat in its mother's milk, or, in some cultures, a norm against eating fish that results in avoidable malnourishment or one that requires men to gorge on protein-rich foods while depriving women of the same nutrients.

Such norms seem to be irrational limitations on liberty at best and destructive of human welfare at worst. How did these norms come to be institutionalized and internalized by large numbers of people and to persist despite their costs and apparent lack

⁴ Timur Kuran, *The Long Divergence: How Islamic Law Held Back the Middle East* (Princeton University Press, 2012).

of sound grounding in morality or prudence? Having a theory of the origination and persistence of improper moralizations may help us develop effective strategies for overcoming them. Evolutionary explanations of morality—especially cultural evolutionary explanations—may provide some of the resources necessary for constructing such a theory. In what follows, we will consider several evolutionary explanations of how invalid moral norms arise, proliferate, and persist.

Surplus Moral Norms as the Result of Adaptation

Any population exhibiting variation and heredity is an evolving system; and if the trait variations it exhibits are causally connected to differential survival and reproduction, then the population can not only evolve (that is, change in its distribution of heritable traits over time) but also produce adaptations.⁵ Culture is an evolving system that has been shown to produce adaptations. Indeed, on dominant accounts of the evolution of cultural capacities in the genus *Homo*, culture was *designed* by natural selection to serve as a parallel channel of non-genetic inheritance for the accumulation of cultural adaptations.⁶ It thus makes sense to query whether any, and if so what proportion, of such apparently invalid moral norms are adaptations. It also makes sense to ask, accordingly, whether the mere fact that a moral norm is an adaptation has any epistemic bearing on the question of its validity.

As we saw in earlier chapters, the prevailing evolutionary selectionist explanation holds that in the environment of evolutionary adaptation (EEA), there were strong selective pressures for the coevolution of moral psychology and moral culture that supported “thick” moral relations among in-group members but

xenophobia, distrust, and reciprocity-contingent attitudes toward out-group members. This “thin” moral consideration extended to out-group members culminated in a refusal to accord them equal moral status and, in extreme cases, the denial of any moral standing whatsoever.

The first and most obvious implication of this evolutionary account of the origins and persistence of invalid moral norms is that individual and cultural susceptibilities to bigoted, xenophobic, and unjustified discriminatory norms may have straightforwardly adaptive roots. The reason for this is simple: invalid norms are often adaptive, and valid norms are often maladaptive; that is to say, we have good reason to believe that the validity of moral norms is not determined by, and often fails to track, their biocultural fitness. As we saw in Chapter 5, the inclusivist anomaly is an evolutionary anomaly precisely in virtue of its failure to track biocultural fitness. The “adaptive plasticity” model of moral psychological development proposed in Chapter 6 and elaborated on in Chapter 7 further fleshes out this Darwinian explanation of invalid moral norms, particularly in relation to exclusivity. According to that model, exclusivist psychological tendencies and cultural norms arise in response to cues of out-group threat permeating the environment in which moralities develop and evolve. This adaptively plastic system, we argued, interacts with normal cognitive biases (such as group essentialism, generic overgeneralization, epistemically flawed cognitive dissonance resolution, etc.), as well as the cultural construction of morally relevant beliefs about out-groups, to produce and sustain environments that are conducive to invalid exclusivist moralities. Although some EEA environments would have allowed for the development and evolution of more inclusive moralities, the model indicates that many arbitrarily discriminatory moral systems will arise as biocultural moral adaptations, or else as evolutionary “misfires” of these adaptive faculties because the out-group threat cues to which they are responding are not veridical. At the same time, the model suggests that environmental

⁵ For a classic statement of the necessary conditions for adaptation, see Richard Lewontin (1978), “Adaptation,” *Scientific American* 239(3): 157–169.

⁶ See Peter Richerson and Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (University of Chicago Press, 2005).

alterations that reduce cues of out-group threat make it possible to break free of invalid moral norms, or at least exclusivist ones.

Not all adaptive moral norms are straightforwardly functional, however. Some moral norms that appear irrational or arbitrary are in fact mechanisms for delineating group membership, coordinating group action, signaling cooperative intent, and/or maintaining group cohesion. This is, for example, how hunter-gatherer religious rituals are now widely understood.⁷ Norms requiring distinctive attire, body modification, hair growth, or ritual participation draw a boundary around the group in such a way as to reduce the risk of exposure not only to out-group biological parasites but also to “alien” ideas, norms, and behaviors that could destabilize cooperation within the group (see Chapter 6). The biblical requirement of male circumcision and the prohibition on simmering a young goat in its mother’s milk, mentioned earlier, may be instances of this phenomenon: circumcision is a costly device to signal a distinctive group identity, and the prohibition on simmering a goat in its mother’s milk banned participation in a ritual of the competing Canaanite religion. Various cultural dietary restrictions (such as prohibitions on pork or beef consumption) may also serve to demarcate group boundaries and serve as similar costly signals of cooperative intent, though they do so by co-opting disgust mechanisms that readily react to animal products that are prone to microbial contamination.⁸

Other seemingly irrational moral norms have an even less obvious effect on the fitness of cultural groups. Because cooperation, at least on a fairly large and complex scale, requires coordination through the following of norms and because internalization of norms improves compliance and reduces the costs of achieving

⁷ Russell Powell and Steven Clarke (2012), “Religion as an Evolutionary Byproduct: A Critique of the Standard Model,” *British Journal for the Philosophy of Science* 63(3): 457–486.

⁸ Daniel Fessler and Carlos Navarrette (2003), “Meat Is Good to Taboo: Dietary Proscriptions as a Product of the Interaction of Psychological Mechanisms and Social Processes,” *Journal of Cognition and Culture* 3(1): 1–40.

it, it is important for individuals to develop the disposition to follow moral norms automatically, as it were. Given that this is so, the internalization of some norms may be functional only insofar as they contribute to the habit of obedience to authority, a habit that has significant fitness payoffs in the case of other, directly functional norms. Here an analogy may be helpful. In basic military training, considerable time is devoted to what American soldiers used to call “monkey drills”—learning to execute rather complex movements on the parade ground that are of no use in combat. The standard explanation for why military authorities devote so much time and energy to such apparently functionless behavior is that it helps form the habit of immediate, unreflective obedience to orders.

Similarly, some of the seemingly excessive and nonfunctional moral rules found in the Bible or in the taboos of premodern societies recorded by anthropologists may be only indirectly functional: they may serve chiefly or exclusively to cultivate the disposition to follow supposedly authoritative norms, a disposition that can have considerable fitness benefits. In addition, as Norbert Elias has emphasized in his monumental book *The Civilizing Process*, compliance with some apparently nonfunctional norms may promote cooperation and even reduce the incidence of violence if they serve as social signals of self-restraint, or readily observable proxies for “prosocial” dispositions.⁹

Surplus Moral Norms as the Result of Evolutionary Mismatch

Other apparently invalid and indeed outright destructive moral norms can be explained as “evolutionary hangovers”—remnants of moral responses that were perhaps functional in the EEA but

⁹ Norbert Elias, *The Civilizing Process Sociogenetic and Psychogenetic Investigations*, 2nd ed., revised, illustrated (Wiley, 2000); Steven Pinker, *The Better Angels of Our Nature* (Viking, 2011).

are unnecessary in and discordant with the current environment. Modern human ecology is far removed from the hunter-gatherer lifeways that characterized the vast majority of human evolutionary history and in which core components of human moral psychology purportedly evolved (see Chapter 11). Thus, as Stephen Pinker notes, some of “our ordeals come from a mismatch between the sources of our passions in evolutionary history and the goals we set for ourselves today.”¹⁰

Evolutionary moral mismatch can take two forms. The first, which we will refer to as the “Pleistocene hangover,” is a mismatch between the “innate” psychological dispositions that solidified in the EEA and the modern ecological environment that our evolved prehistoric psychologies must navigate. A classic example is the human fondness for sweet foods and aversion to bitter foods. Sweetness generally indicates the high energy density of a food source and is a good proxy for vitamin C content; in addition, primates are able to store fructose as fat, which can then be tapped for crucial calories in times of food shortage. In contrast, bitterness (especially in plants) tends to indicate the presence of natural pesticides and other sources of toxicity and is associated with low-quality foods. High-energy sweet foods, such as fruit, berries, and honey, tend to be rare and comprised only a very small—and highly desired—portion of early hunter-gatherer diets (which consisted mainly of meat and tubers). Humans have thus inherited an evolved penchant for sweet foods and their associated neurochemical pathways of reward and an aversion to bitter foods. In the modern human environment, however, in which there is a superabundance of calories and effectively unlimited access to high-energy foods, the sweet tooth adaptation (and the adaptive aversion to bitter plant foods) may result in obesity, diabetes, and other serious damage to long-term human health.

¹⁰ Steven Pinker, *The Blank Slate* (Penguin Classic 2002, p. 219). For a more recent articulation of the ‘mismatch hypothesis,’ see Ronald Giphart and Mark Van Vugt, *Mismatch: How Our Stone Age Brain Deceives Us Every Day* (Robinson 2018).

Other Pleistocene hangovers involve specifically *moral* psychology. For example, leadership-followership psychology was critical to coordinating activities of prehistoric human groups, such as hunting, foraging, war-making, and resolving internal disputes¹¹—particularly as human societies expanded to sizes that make strictly egalitarian decision-making cumbersome and inefficient in real-time conflicts.¹² However, leadership-followership psychology may pose grave risks in the modern world, where state-level conflicts involve hundreds of millions of people and deploy powerful weaponry that can have irrevocable consequences for generations far into the future. For instance, there is a well-documented human tendency to gravitate toward authoritarian, hawkish, masculine, and charismatic leaders in times of actual or perceived intergroup conflict.¹³ In the EEA, such aggressive posturing and “rally-round-the-flag” proclivities may have been adaptive; but in the modern world of interstate brinkmanship and terrorism with weapons of mass destruction, such “hawkish” virtues may have devastating costs. Moreover, elites may exploit this prehistoric moral psychology by provoking intergroup conflicts or by engendering perceptions of intergroup threat—or of a threat “from within” (see Chapter 11)—in order to consolidate power. Indeed, we seem to be witnessing these demagogic dynamics in contemporary U.S. electoral politics (see Preface and Chapter 10).

A second type of evolutionary moral mismatch is that between evolved cultural moral systems and the modern ecological

¹¹ Mark Van Vugt et al. (2008), “Leadership, Followership, and Evolution: Some Lessons from the Past,” *American Psychologist* 63(3): 182–196.

¹² Although early human societies are generally thought to have been rather egalitarian, subordination to the *temporary* authority of a powerful male (so-called Big Men) in times of armed conflict also seems to have been common. Ibid.

¹³ Mark Van Vugt et al., “Evolution and the Social Psychology of Leadership: The Mismatch Hypothesis,” in C. Hoyt, D. Forsyth, and A. Goethals (eds.), *Social Psychology and Leadership* (Praeger, 2008), pp. 267–282.

environment. Consider, for example, the differences in cultural conceptions of honor between American populations in the North and those in the South, discussed in Chapter 6. Southerners are far more likely to respond to insults or affronts to their honor with violence—a cultural difference that is explained, as we have already seen, by the fact that the primary settlers of the South were livestock herders with a hyper-masculine, honor-based culture. Honor cultures, which are robustly associated with livestock herding throughout the world, appear to be cultural adaptations to rugged, lawless regions of countries where there is little or no institutional recourse to prevent theft and other forms of predation.¹⁴ In modern environments, however, cultures of honor impose significant surplus moral constraints and tend to involve improper moralization as well. Imported into grand conflicts between powerful states and combined with prehistoric leadership-followership psychology discussed above, a culture of honor can cause spiraling, destructive intergroup conflicts; make peaceful resolutions harder to come by; and create conditions in which the critical scrutiny of moral norms—or what we referred to earlier as the capacity for “open-ended normativity” (see Chapter 5)—is unable to gain sufficient purchase.

An example of a costly evolutionary moral mismatch that may implicate both innate dispositions and cultural moral norms concerns the treatment of homosexuality. It has been suggested that prohibitions on homosexual sex and, even more so, on homosexual partnership may have been selected for because of their contribution to higher fertility rates in small, vulnerable societies (such as the biblical Israelites) whose survival depended upon achieving high fertility. In addition, this prohibition may have facilitated an efficient division of labor between men and women (big game hunting/warfare, on the one hand, childcare/local foraging, on the other), and this specialization may have

¹⁴ R. E. Nisbett and D. Cohen, *Culture of Honor: The Psychology of Violence in the South* (Hachette, 1996).

been crucial for survival for much of human evolutionary history. Prohibitions on homosexual sex and the enforcement of stereotypical gender roles are, of course, no longer critical for group survival; to the contrary, they can be highly detrimental to group success. Population increases can be economically disadvantageous, and the ability of women to enter the workforce and to be able to compete for desirable positions has a significant positive impact on economic development. In modern human ecology, therefore, prohibitions on same-sex partnerships and the enforcement of strict gender roles seem to do far more harm than good—and, from the standpoint of morality properly conceived, this harm is unjust because it falls disproportionately on women and sexual preference minorities.

Although norms can sometimes change rapidly, evolutionary investigations of such norm–environment mismatches show that formerly adaptive norms can have substantial inertia, even when societies find themselves in ecological circumstances to which the norms are ill-suited. What explains this cultural staying power? One possibility is that it simply takes time for new cultural variants to emerge and become sufficiently frequent in the population for cultural copying biases to drive them to fixation (more on the dynamics of cultural transmission below). Another possibility is that some norms are “culturally entrenched”—a cultural analog of developmental constraint in biology.¹⁵ A norm is culturally entrenched if it is causally connected to other aspects of a cultural tradition web, such that the norm cannot be altered without the costly alteration of many other aspects of the web, resulting in the norm’s selective preservation. Norms that implicate group identity or moral identity, for example, are likely to

¹⁵ For an extended discussion of different types and causes of generative entrenchment, see W. C. Wimsatt, “Entrenchment and Scaffolding: An Architecture for a Theory of Cultural Change,” in L. Caporael, J. Griesemer, and W. Wimsatt (eds.), *Developing Scaffolding in Evolution, Cognition, and Culture* (MIT Press, 2013), pp. 77–105.

be preserved at great cost because they occupy a central, highly connected position in the cultural web. A key problem for cultural evolutionary research is to identify what the philosopher of science William Wimsatt calls “escape mechanisms” that allow for deep modifications of entrenched cultural structures whose alteration would otherwise send devastating ripples across a cultural system.¹⁶ Even if core cultural norms do not budge in an individual in which they are already entrenched, in today’s world of pluralistic societies with modes of mass communication, new generations may acquire cultural elements not only vertically from their parents but also (and especially) from their peers in adjacent cultures, allowing even the core norms of a population to be transformed over time.

Surplus Moral Norms as the Result of Special Interest Adaptations

There is a tendency to think of organisms either as harmonious wholes or else as mere vehicles through which genes ensure their representation in the next generation. Both of these conceptions are wrong. The fallacious “gene’s eye” perspective has held a strong sway over the public understanding of evolution, thanks to the effective popularization of evolution by the likes of Richard Dawkins. Over the last few decades, however, it has become clear that the genetic level is only one level of the biological hierarchy at which natural selection can act. Multilevel selection theory explains, with the aid of rigorous modeling, how evolution can act on multiple levels simultaneously and how this multilevel selection process can produce adaptations at one level of organization that are detrimental to the stability of another level.¹⁷

¹⁶ Ibid.

¹⁷ See David Sloan Wilson and Elliott Sober (1994), “Reintroducing Group Selection to the Human Behavioral Sciences,” *Behavioral and Brain Sciences* 17(4): 585–654.

The biological world is comprised of individuals nested within other individuals: genes aggregate to form chromosomes, cells aggregate to form multicellular organisms, multicellular organisms aggregate to form colonies, and so on. Importantly, the necessary conditions for selection—heritable variation causally connected to differential survival and reproduction—can be met at any of these nested levels simultaneously. So, contrary to the gene’s eye view, genes are not the only objects of selection. Because selection can work simultaneously in different directions across levels of organization, a trait that is adaptive at one level can be deleterious at other levels. Indeed, the key to understanding the formation and maintenance of higher levels of biological organization—such as the multicellular organism—is to understand how cooperation is achieved and evolutionarily “selfish” tendencies mitigated among its lower-level parts. In the case of paradigmatic biological individuals comprised of lower-level individuals—such as multicellular organisms comprised of cells or colonies comprised of organisms—there are mechanisms in place to ensure that lower-level adaptations deleterious to higher-level survival and reproduction will be selected against.¹⁸ For example, the division of labor between germ and somatic cells prevents any particular cell line from “going it alone”; likewise, cancerous cell lines are targeted by the immune system in multicellular organisms, and eusocial insect nest-mates attack workers that attempt to reproduce.

Human societies do not resemble paradigmatic individuals, however, in part because they have far more limited means of addressing interlevel replication conflicts. This is especially true of large, complex post-Neolithic societies, in which adaptations of lower-level components (such as elite castes) can emerge despite their deleterious consequences for larger human collectives.

¹⁸ Richard Michod, “Evolutionary Transitions in Individuality,” in B. Calcott and K. Sterelny (eds.), *Major Transitions in Evolution Revisited* (MIT Press, 2011, pp. 169–197).

Thus, norms that benefit an elite subset of individuals or a privileged class within a society can persist despite the fact that they are deleterious for most individuals within the society and even for the society as a whole. We will call such cultural adaptations *special interest adaptations*. In the long haul, special interest adaptations may undermine cultural group stability and thus be selected against; in the short term, however, there is historically little by way of social mechanisms to stop the evolution of special interest adaptations in hierarchically complex, multilayered human societies, which have historically lacked adequate enforcement mechanisms at the group level to guard against them. The rule of law and constitutional democracy are very recent cultural innovations that, in effect, are designed to check special interest adaptations and place limits on state and elite class power.

An example of a deleterious set of norms generated and perpetuated as special interest adaptations concerns the profoundly incompetent, ineffective, and unjust criminal justice systems of many developing countries that have gained independence from colonial rule. The norms that underpin criminal justice institutions in many postcolonial developing countries were originally designed to protect the property and power of colonial rulers and their elite allies at the expense of the general population—in essence, they served as mechanisms of popular suppression. After independence, rather than reforming these norms, many postcolonial regimes preserved and benefited from them, with elite groups coming to occupy the powerful positions held by their former colonial rulers.¹⁹

Traits that were originally selected for performing one fitness-enhancing function may come to perform a new function, including a special interest function that enhances the fitness of a subset of a collective at the expense of other individuals. The complex of

norms that constitute the Indian caste system, for example, may have reduced the risk of biological parasites that aboriginal peoples of the subcontinent posed to their Vedic conquerors.²⁰ But later, when the conquered and the conquerors came to comprise one larger society, these same norms may have functioned to consolidate the power of the conquerors' descendants through their control over the state apparatus (coercion) and religious authority (ideology), by preventing the dilution of power through intermarriage with descendants of the conquered and by reserving valued social positions for themselves.²¹

Surplus Moral Norms as the Result of Failures of Collective Action

As discussed above, the evolution of paradigmatic evolutionary individuals, such as organisms and colonies, hinges on the evolution of effective mechanisms for regulating reproductive conflicts among lower-level units. In particular, it requires mechanisms that control the ability of lower-level units to act in their own evolutionary "self-interest," preventing them from "defecting" or "free-riding" in ways that undermine cooperation at the higher level. Thus, the formation of evolutionary individuals requires that the evolutionary process find solutions to difficult collective action problems, and the evidence suggests that human societies have only limited resources for solving them.

In human cultural evolution, collective action problems can not only undermine cooperation that is beneficial for all but also sustain harmful cooperative structures that prevent defection in ways that leave everybody worse off, including elites that originally benefited from those cooperative arrangements. In other

²⁰ W. H. McNeil, *Plagues and Peoples* (New York: Anchor, 1998).

²¹ It is worth noting that while in the past socioeconomic and political advantages may have been conducive to individual reproductive fitness, this is no longer true in many societies, where the better off tend to have lower rates of reproduction.

¹⁹ Gary Haugen and Victor Boutros, *The Locust Effect: Why the End of Poverty Requires the End of Violence* (Oxford University Press, 2015), pp. 171–186.

words, some surplus moral norms can persist due to failures of collective action even when they confer no reproductive advantage on the group or any of its members. Dueling, footbinding, and female genital mutilation may be examples. Dueling may have originated through a combination of sexual selection (in an earlier environment in which females put a premium on physical courage in males) and selection for relatively constrained forms of violence (where conflicts are settled on the field of honor by the actions of two individuals, as opposed to the continuing, uncontrollable conflict of intergenerational blood feuds between groups). But even when these original functions became otiose due to cultural innovations that provided less physically destructive outlets for competition among males—and even though most people, including most participants in the practice, recognized how destructive it was—dueling persisted. Similarly, careful investigations of female mutilation norms, such as footbinding and genital cutting, show that these practices arose initially as special interest adaptations (in particular, as paternity confidence measures for wealthy elites) and then spread to the general population, where they were bolstered by false empirical beliefs (e.g., about their health benefits).²²

Why do such destructive and apparently maladaptive norms persist, even when they fail to benefit or confer a fitness advantage on anybody? A central explanation for their persistence is that abolishing them requires solving difficult collective action problems. For instance, even if each potential duelist believes the practice to be irrational and even immoral, any defecting individual will face debilitating social stigma or, in the case of footbinding and genital mutilation, severely reduced marital prospects. Similarly, as Kim Sterelny has suggested, even if the initial victims of female genital mutilation enjoyed an advantage in the mate selection market (given the cultural context

of a patriarchal, deeply sexist society in which female chastity was inordinately valued), once the practice became widespread, the advantage evaporated: if virtually every woman has mutilated genitals, having them confers no comparative advantage.²³ In other words, undergoing female genital mutilation came to be a matter of horrific costs with no offsetting fitness benefits for the individuals undergoing it. The practice persists, Sterelny argues, because abolishing it, as with the case of dueling, requires solving a difficult collective action problem. The first defectors from the practice will suffer a prohibitive reproductive penalty because they will be viewed as inappropriate mates in societies in which unmarried females have grim economic prospects, and they will be subject to moral condemnation, stigmatization, and intimidation.

Effectively counteracting special interest adaptations, such as those discussed above, may involve solving a collective action problem not only for the society as a whole (through, e.g., anti-infibulation or anti-footbinding pledge societies) but also for the disadvantaged subset of society in particular. A powerful elite or privileged contingent has at its disposal impressive resources for blocking the collective action necessary for the masses to emancipate themselves from surplus norms that favor special interests. These resources include coercion, or attaching material costs to noncompliance with such norms, as well as ideologies that obscure the fact that the norms are nothing more than instruments of class domination. Just as mechanisms for reducing somatic mutation rates sustain cooperation in organisms, so too do ideologies function to reduce rates of cultural “mutation” that could destabilize societal arrangements that benefit all or, in some cases, that benefit primarily an elite caste. In essence, ideologies can act as immune systems, blocking invading cultural variants that could destabilize existing institutional structures and undermine social

²² Gerry Mackie (1996), “Ending Footbinding and Infibulation: A Convention Account,” *American Sociological Review* 61(6): 999–1017.

²³ Kim Sterelny (2007), “SNAFUS: An Evolutionary Perspective,” *Biological Theory* 2: 317–328.

cohesion, whether this is to the benefit of all or only a subset of society.

Furthermore, participation in a revolution involves a cost to the individual participant, but whether the revolution will succeed depends upon whether enough people participate. Even though emancipation would be best for all members of the oppressed class, it may be rational for each oppressed member to refrain from participating in the revolution. Indeed, the tendency of the worse off to rationalize special interest norms—that is, to buy into ideologies that preserve the status quo—may in fact be a mechanism for avoiding the costs of challenging prevailing special interest norms, given the likelihood that such challenges will not succeed due to problems of collective action and given that failure could have disastrous and potentially fatal consequences. Thus, ideologies may function both as special interest adaptations and as adaptations that enable subjugated groups to cope with special interest adaptations.

Surplus Moral Norms as the Result of the Dynamics of Cultural Transmission

Finally, some harmful surplus moral norms may proliferate and be sustained in a society simply due to the intrinsic dynamics of cultural transmission, which allow maladaptive variants to spread rapidly in a cultural population. Unlike the clean lines of vertical descent exemplified by genetic transmission, cultural variants can be acquired from and transmitted to any member of a population within a single lifetime; this allows cultural variants to spread much more rapidly than genetic variants, but it also makes cultural transmission uniquely susceptible to the spread of maladaptive variants. As discussed in Chapter 5, Richerson and Boyd's modeling work has shown that cultural copying biases—such as tendencies to copy cultural variants that are common, to emulate prestigious individuals, and to identify transparently successful strategies—can allow for cumulative cultural adaptation.

But these are far from fail-safe heuristics as destructive norms are often adopted by prestigious individuals—consider, for example, celebrity “anti-vaxxers” (opponents of childhood vaccination) and more generally the tendency of experts to opine outside of their proper domain of expertise. Furthermore, futile or harmful norms are often mistaken for successful ones, particularly in cases of complex causation, such as epidemiology and disease.²⁴

In sum, seemingly arbitrary, irrational, or bigoted moral norms may be sustained in a society even though they are deleterious in modern selective environments, even though they are adaptive for only small subsets of human populations (such as powerful elites), even though they actively harm large segments of society, and even though they no longer (or never did) confer a fitness advantage on anyone. It follows that the fact that a norm is maintained in a society does not, therefore, provide persuasive or even *prima facie* evidence that the norm has a salutary function. This simple fact, as we shall soon see, has momentous implications for traditional conservative thinking, which takes the longevity of social practices and institutions as evidence of their “wisdom.”

How Can One Reliably Identify Surplus Moral Constraints?

This chapter began with a list of relatively uncontroversial cases of surplus moral constraints and hence proper targets for de-moralization. De-moralization, however, can go awry: people can and often have come to regard as morally permissible behaviors that are in fact morally wrong. For example, in the thrall of Nazi ideology, many ordinary Germans came to believe that behavior

²⁴ On the unique susceptibility of cultural transmission to deleterious variants and how cultural copying biases partially overcome these susceptibilities, see Robert Boyd and Peter Richerson, “Norms and Bounded Rationality,” in Gerd Gigerenzer and Reinhard Selten (eds.), *Bounded Rationality: The Adaptive Toolbox* (MIT Press, 2002, pp. 281–296).

they previously viewed as immoral when directed at any person was permissible—or even obligatory—when the target was a Jew. Similarly, as discussed in Chapter 7, eugenic propaganda in the United States and elsewhere convinced many people that policies of compulsory sterilization that would otherwise have been rejected as immoral were not only permissible but obligatory, given the false assumption that the human gene pool and even civilization itself was imperiled by the rampant proliferation of “defective germplasm.”²⁵ How can one reliably ascertain when abandonment of a moral norm and the constraints it entails is a case of proper de-moralization—and hence of moral progress—and when it is not?

One might think that the solution to the problem is simple: any moral norm is likely to be a surplus moral constraint, and hence a proper target for de-moralization, if no sound justification can be given for retaining it, whereas a positive moral justification can be given for abandoning it. Indeed, if one can give a plausible evolutionary explanation of an apparently destructive moral norm along the lines sketched earlier, then this may give one a particularly good reason to doubt its validity, *if no reasoned justifications for the norm are forthcoming*. Assuming that one knows how to identify sound moral justifications, what more is needed? If one adopts a broadly liberal perspective, then justifications must appeal ultimately to the freedom and welfare of individuals, and brute appeals to religious authority or tradition do not suffice. If compliance with some supposed moral norm exacts significant human costs and there is no justification for it in terms of its contribution to individual well-being and freedom, then isn't one justified in thinking that it is a surplus moral constraint?

Unfortunately, things are not so simple, as generations of conservatives have emphasized. Whether an accepted moral norm

ought to be disregarded as being a case of improper moralization cannot be determined unless we can reliably ascertain its role within a complex web of norms, institutions, and social practices. The justification of moral norms must be holistic or, more precisely, ecological; but given the limitations of our knowledge about the social wholes within which norms operate, we are at risk of failing to see the true value of certain norms. For example, some moral norms may fit the “monkey drill” analogy discussed above: compliance with them may produce no particular substantive good, but they may nonetheless be valuable because they cultivate and sustain the disposition to follow those moral norms that are important for human flourishing. Other norms may in fact contribute to some substantive good or to the preservation of valuable liberties but in complex ways that are not likely to be captured by widely understandable—and, to that extent, simple—moral justifications.

Consider, for example, a set of norms concerning sexual morality that includes a prohibition on unmarried women bearing children and that requires stigmatization of those who violate the norm. Compliance with this norm inflicts serious psychological, social, and economic costs on unmarried women who have children. Yet the norm may in fact be beneficial overall and in the long run, for the class of women as a whole and perhaps even for disadvantaged women, in an environment in which social support for unmarried mothers is lacking and in which marriage is unattainable for many disadvantaged women (or, if attainable, does not constitute an economic improvement). In such an unjust social order, a norm that imposes severe costs on unmarried women who bear children might, depending upon the factual particularities, make moral sense, provided that we give significant weight to the well-being and opportunity of women over the long run. And if that is so, then concluding that it is a surplus moral constraint—and striving to abolish it—might not be morally progressive, all things considered. Whether the norm is justifiable will depend upon complex moral reasoning that includes

²⁵ Allen Buchanan (2007), “Institutions, Beliefs and Ethics: Eugenics as a Case Study,” *Journal of Political Philosophy* 15(1): 22–45.

among its premises highly disputable empirical predictions about the consequences of adhering to it or abandoning it.

An Evoconservative Challenge to De-moralization

Conservatives might attempt to bolster these pre-scientific reflections on the difficulty of knowing the unintended consequences of de-moralization by appealing to evolutionary theory, much as they did in arguing against the prospect of inclusivist moralities (see Chapter 4). In particular, they could make two evoconservative assertions—and then conclude, too quickly, we shall argue, that societies should err in the direction of adherence to the moral status quo. The first assertion is that if a moral norm has persisted over a long period of time, then this is in itself a good reason to believe it is beneficial. This assertion reposes on the two-pronged assumption that cultural selection acts as an optimizing force, or at least as a force that tends to produce group-beneficial moral norms, and that group-beneficial moral norms tend to be non-surplus moral norms properly conceived. As the above discussion suggests, however, this evoconservative assumption turns on a vulgarized, inaccurate view of cultural evolution. Namely, it wrongly assumes that maladaptive social practices will be winnowed out in reasonably short order in favor of adaptive ones, and it mistakenly assumes that adaptive moral norms will tend to be valid or beneficial from the standpoint of morality—and hence, it fails to prop up the conservative assertion that the longevity of a practice is evidence of its salutary nature. Recall that to say that a norm is an adaptation is to make a purely backward-looking assertion; it tells us nothing whatsoever about the present effects of the norm.

As we have seen, highly destructive and immoral social practices can persist for long periods of time because they confer a fitness advantage (i.e., they are under stabilizing selection), because they are the result of the consistent misfire of adaptive propensities,

because they are evolutionary hangovers, because they are entrenched and thus refractory to modification due to their contingent causal connections to other adaptive cultural norms, because they serve the special interests of powerful elites at the expense of other segments of the population, because abolishing them requires solving difficult collective action problems, and because cultural transmission is highly susceptible to the spread of maladaptive variants.

The second idea that conservatives typically invoke to support their claim that judgments about surplus moral constraints are so unreliable as to warrant a strong bias for the moral status quo relies on a metaphor: the notion that society is a “seamless web” or, more hysterically, a “house of cards.” Snipping one apparently insignificant fiber may unravel the whole thing (alternatively, making what one thinks is a minor adjustment in the position of one card may cause the whole edifice to collapse). The operative notion here is simple: there are dense interconnections among moral norms and the social practices that support them, and given how little we know about the particulars of these dense connections, it is hubristic—and morally irresponsible—to abandon a norm simply because we cannot produce a convincing justification for it.

The idea that moral culture is like a seamless web is closely, if implicitly, allied to the evoconservative assertion that cultural evolution tends to produce optimal configurations of the variegated components of moral systems, taking advantage of subtle causal interconnections and managing complex trade-offs of which human would-be social engineers are incurably unaware. Yet there is an unacknowledged tension here between the ability of cultural evolutionary processes to sculpt adaptations, on the one hand, and the aptness of the seamless web metaphor, on the other. As we have argued in a very different context, the seamless web and house of cards metaphors greatly exaggerate the core conservative insight that we ought to take seriously the risk of unintended bad consequences when we “tinker” with complex

biological and social systems.²⁶ Societies, like individual organisms, are not seamless webs, and they are certainly not like houses of cards, because neither type of entity is a plausible evolutionary product of natural or cultural selection. For an entity to evolve through natural or cultural selection, some features of that entity must be able to change without altering (in a countervailing way) other crucial characteristics. Biological theorist Richard Lewontin refers to this as the “quasi-independence” criterion for adaptation²⁷—that to be shaped by selection, traits must be capable of modification without disrupting other important components of the system. Likewise, Lewontin stressed what we might call the “incrementality” condition for adaptation: that if an evolving system is to be capable of achieving an adaptive match to some ecological design problem, then small changes in certain features of that system cannot have large ramifications for the overarching shape and ecological position of the system—since if this were not the case, then it is unlikely that selection could ever push a lineage up an adaptive peak.

Since we know that adaptation exists at both individual and cultural levels, quasi-independence and incrementality must obtain at these levels as well.²⁸ It stands to reason that if cultural systems were as fragile as the seamless web and house of cards metaphors suggest, they would not be resilient enough to survive and adapt to changes in the environment, including competition from other societies and individuals—and thus, contra the evoconservative, there would be no scientific reason to think that moral systems

²⁶ Allen Buchanan, *Better Than Human* (Oxford: Oxford University Press, 2011); Russell Powell and Allen Buchanan (2011), “Breaking Evolution’s Chains: The Prospect of Deliberate Genetic Modification in Humans,” *Journal of Medicine and Philosophy* 36(1): 6–27.

²⁷ Lewontin, “Adaptation,” supra note 5.

²⁸ See Robert Brandon, “Evolutionary Modules: Conceptual Analyses and Empirical Hypotheses,” in Werner Callebaut and Diego Rasskin-Gutman (eds.), *Modularity: Understanding the Development and Evolution of Natural Complex Systems* (MIT Press 2005), pp. 51–60.

are optimally or even beneficially configured. The phenomenon of cultural adaptation is not compatible with the hyper-dense developmental connectedness that conservative metaphors imply.

If one insists on textile or architectural metaphors, then it would be better to say that individuals and societies are like seamed webs or complexly modular buildings. Any entity that is subject to selective shaping is likely to feature a good deal of modularity and redundancy. Modules are functional units that have denser connections among their own constituents than between themselves and other functional units. Modularity is conducive to adaptability because it allows for incremental (intramodular) changes that do not result in catastrophic disruptions to the larger system, with the boundaries between modules akin to seams in a web. Likewise, functional redundancy is conducive to adaptability because it allows for changes that undercut a function in one system or organ to occur without complete loss of that function. For example, many genetic innovations in evolution are made possible by gene duplication that initially results in functional genetic redundancy, which in turn frees up one of the duplicates to assume a novel evolutionary function. If cultures are robustly evolving systems capable of achieving adaptation, then it is likely that they too have resources for functional redundancy that permit lower-risk evolutionary tinkering and innovation.

The inaptness of evoconservative metaphors is further confirmed by the brilliant modeling work of Boyd and Richerson, which shows that individual norm compliance is much more developmentally autonomous than the conservatives’ favorite metaphors suggest. Almost any norm, including one that requires abandonment of a pre-existing widely accepted norm, can enjoy robust compliance if there is effective punishment for non-compliance.²⁹ Norms are thus not as densely interconnected as

²⁹ Robert Boyd and Peter Richerson (1992), “Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups,” *Ethology and Sociobiology* 13: 171–195.

evoconservatives assume; and if this is so, then the risk of improper de-moralization is not as great as they assert. In short, conservatives are fond of saying that we don't know enough about society to disregard long-standing moral norms, and evoconservatives might appeal to evolutionary theory to bolster this assertion; but this would be to misunderstand the implications of biocultural evolutionary theory and to ignore the fact that we now have a great deal of evidence that societies are not like seamless webs or houses or cards.³⁰ None of this is to deny that the risk of unforeseen bad consequences is a serious problem for the reliable identification of proper targets of de-moralization. The point, rather, is that the evoconservative stance on this problem, much like its view on the bounds of moral inclusivity, is unduly pessimistic and unsupported (and, in fact, contradicted) by current evolutionary theory.

Contained Experiments in De-moralization

Although the evoconservative line on de-moralization overstates the risks of moral reform, any theory of moral progress that takes seriously the need for emancipation from surplus moral constraints must develop a plausible strategy for managing the risks of bad unintended consequences of de-moralization. "Managing" is the right term here because it would be both

³⁰ Of course, enforcement only works if it is employed. It might be the case that a norm N1 could be abandoned without bad consequences, including the undermining of a valid norm N2, but only if another norm N3 were enforced. Suppose, however, that the fact that the enforcement of N3 is necessary to prevent the abandonment of N1 from causing damage to N2 is not known and a consequence N3 is not enforced. This possibility lends support to a moderately conservative thesis with which the authors agree, namely, that anyone proposing or welcoming the abandonment of a norm ought to take seriously the risk of unintended bad consequences of doing so. It does not support the assumption of extremely dense interconnections among norms suggested by the seamless web metaphor.

unfeasible—and undesirable if feasible—to reduce the risks of bad unintended consequences to zero. Here, as elsewhere, risk reduction is not costless, and the marginal costs of risk reduction are likely to rise at some point within the feasible set. Instead of eliminating risk, the goal is to achieve cost-effective risk management, where costs are construed quite broadly.

It may turn out that, generally speaking, people are better at making reliable *retrospective* judgments as to whether the abandonment of a norm is a case of proper de-moralization for two reasons. First, in retrospect (at least if enough time has elapsed) we may have reason to conclude that abandoning the norm did not in fact have serious unintended bad consequences. Here it is worth noting that there are many cases where conservatives have predicted dire consequences of de-moralization that have not materialized—for example, that if same-sex marriage is permitted, the institution of marriage will be damaged, or that if interracial marriage is permitted, it will lead to the degeneration of the "white race" or to social chaos, or that if consensual homosexual acts are decriminalized, fundamental values will be eroded and the social fabric will unravel. Or consider the extremely pessimistic, if not hysterical, predictions of the social and psychological consequences of allowing in vitro fertilization when it first became available in the 1970s.

Second, if sufficient time lapses after the abandonment of what was previously thought to be a valid norm and if we believe that during the interval the cause of social justice has advanced or at least not been significantly retarded, then we may conclude that the overall effects of abandoning the norm, over the long run, have not been bad. Fortunately, the abandonment of a norm often comes gradually, in stages, as when physician-assisted suicide first becomes permissible only under certain highly constrained circumstances, when medical use is first allowed as an exception to the prohibition on using marijuana, when gay marriage is legalized in certain jurisdictions, or when alternative reproductive practices (such as germline modification) are permitted under

limited conditions. In such cases, there will be time to determine whether complete abandonment of a prohibition is advisable or not, and the costs of norm modification will be reversible and contained.

Such "moral experimentation," as it might be called, is a crucial complement to moral reasoning in assessing the justificatory value of existing norms because, as evoconservatives rightfully note, the intricacies of cultural casual relations may elude even our best moral reasoning and social modeling. Thus, even if, despite the evolutionary considerations adduced earlier, we take the existence of norm N to be *prima facie* evidence that N serves some valuable social function that would be vitiated if N were altered, we can conduct controlled "experiments" in norm modification that allow us to assess the unintended consequences that are likely to flow from N's alteration.

This is not to say, however, that contemporaneous or prospective judgments about surplus moral constraints are never justified. In some cases, the human costs of continued compliance with a supposed moral norm are so horrific, the benefits so arbitrarily skewed toward one group in society, and the lack of a justification so patent that we may rightly conclude that adherence to the norm is a case of unnecessary, self-inflicted curtailment of liberty. Several of the instances of de-moralization listed at the beginning of this essay seem to us to satisfy these criteria.

In addition to the problem of predicting the consequences of abandoning a given norm, there is the even more difficult problem of *evaluating* the predicted consequences. This evaluation may, in some cases, turn on highly disputed issues of distributive justice—perhaps issues that no current theory may be capable of resolving satisfactorily. De-moralization may be beneficial for some, perhaps many, but quite harmful to others. In other words, an account of proper de-moralization must ultimately take a stand on some of the most fundamental and disputed issues concerning distributive justice.

In conclusion, any theory that aims to explicate the importance of increased liberty or welfare for moral progress ought to take the phenomenon of de-moralization seriously and must develop an account of the conditions for making reliable judgments about surplus moral constraints. Such a theory should make use of knowledge gained from controlled moral experimentation, as well as current research in evolutionary theory, to better understand not only the reliability of moral judgments but also the origins and persistence of invalid moral norms and to provide practical guidance as to how emancipation from these norms can be achieved.