Journal of Ethics and Social Philosophy
VOL. 29, NO. 2 · JANUARY 2025
https://doi.org/10.26556/jesp.v29i2.3316

IT'S ONLY NATURAL!

MORAL PROGRESS THROUGH DENATURALIZATION

Charlie Blunden

MORAL PROGRESS occurs when things change for the better, morally speaking. Questions of moral progress have recently been receiving increasing interest from philosophers.[1] But how does moral progress happen? This question concerns the *causality of moral progress.*[2] In this paper, I seek to advance the discussion on a potential cause of moral progress that I will refer to as *denaturalization*.[3]

Denaturalization has been investigated by several philosophers in the moral progress literature, most notably Nigel Pleasants, Julia Hermann, Dale Jamieson, and Elizabeth Anderson.[4] The idea is that moral progress can be facilitated by people coming to have a more accurate understanding of the extent to which their institutions are natural or necessary. Proponents of denaturalization as a cause of moral progress argue that progressive moral change is often blocked by a false understanding on behalf of relevant social actors that their current institutional setup is in some way "natural and indispensable."[5] These beliefs

1 For an overview, see Sauer et al., "Moral Progress."

2 Extant theories include that moral progress is caused by greater knowledge of the moral facts (see Huemer, "A Liberal Realist Answer to Debunking Skeptics"); by adaptively plastic psychological mechanisms that respond to increased material security (see Buchanan and Powell, *The Evolution of Moral Progress*, ch. 6); or by the exercise of moral consistency reasoning under favorable social conditions (see Kumar and Campbell, *A Better Ape*).

3 I borrow the term *denaturalization* from Jaeggi, *Critique of Forms of Life*, 8, though I make no claim to be using the term in her sense. Rather, I am using the term to refer to a proposed cause of moral progress discussed by several philosophers in the moral progress literature, described below.

4 See Pleasants, "The Structure of Moral Revolutions" and "Moral Argument Is Not Enough"; Anderson, "Social Movements, Experiments in Living, and Moral Progress"; Jamieson, "Slavery, Carbon, and Moral Progress"; and Hermann, "The Dynamics of Moral Progress."

5 Hermann, "The Dynamics of Moral Progress," 305. See also Jamieson, "Slavery, Carbon, and Moral Progress," 177–80; Pleasants, "Moral Argument Is Not Enough," 166; and Anderson, "Social Movements, Experiments in Living, and Moral Progress," 16.

can often be a significant impediment to changes away from an unjust status quo, and undermining them can be a significant cause of moral progress, as the unjust status quo is then left with no "veneer of naturalization" to hide behind.[6]

The paradigm case that denaturalization is meant to explain is the successful abolitionist movement in nineteenth-century Britain, and I will explore this case in more depth in the first section. Denaturalization has also been implicated in other past or potential instances of moral progress. Hermann points out that appeals to naturalness have played a role in defending practices of discrimination against homosexuality and the oppression of women, which may imply that, to the extent that these practices have been undermined, denaturalization has played a role.[7] Proponents of denaturalization have also suggested that it may have a role to play in moving away from a carbon-intensive economy or in challenging the view that "there is no plausible alternative to wage labor and the market economy" so that an alternative and morally preferable economic system, if one is indeed possible, can be adopted.[8]

The current literature on denaturalization as an explanation of moral progress contains some vagueness about what denaturalization is and how it works, which makes it difficult to work out: what exactly denaturalization is; what empirical presuppositions need to be correct for denaturalization to be a psychologically realistic account of how moral progress happens; and whether and under what conditions denaturalization might lead to moral progress. Thus, my main aim is to develop, using the existing literature as a guide, a more detailed and explicit account of what denaturalization is and how it might work so that the aforementioned points of unclarity can be made clearer.

This paper has four sections. In the first section, I specify denaturalization by clarifying the different interpretations one could have of claims that a given practice or institution is natural or necessary. I argue that the interpretation most compatible with the existing literature is that claims of naturalness or necessity are claims about the *costs* of getting rid of existing institutions and moving to an alternative. In the second and third sections, I develop what I call a *costs account of denaturalization*. In the second section, I explicate a general framework, using recent advances in philosophical understandings of conventionality, which enables us to understand claims of naturalness and necessity as

6 Hermann, "The Dynamics of Moral Progress," 307; and Jamieson, "Slavery, Carbon, and Moral Progress," 180.

7 Hermann, "The Dynamics of Moral Progress," 307.

8 On the potential role of denaturalization in moving away from a carbon-intensive economy, see Jamieson, "Slavery, Carbon, and Moral Progress," 177–78. On its potential role in overcoming the notion that there is no alternative to wage labor and a market economy, see Pleasants, "Moral Argument Is Not Enough," 176–77.

claims about the costs of abandoning status quo institutions and to understand how these claims can be mistaken in degrees. In the third section, I present a brief case for the psychological realism of this account of denaturalization. I suggest that the costs account has some claim to being psychologically realistic, while also highlighting the limits of this claim and outlining the kinds of empirical evidence that proponents of denaturalization need for a convincing account of the psychological realism of denaturalization as a cause of moral progress. Fourth, with the more detailed costs account of denaturalization in hand, I investigate whether and under what conditions denaturalization can lead to moral progress.

1. DISAMBIGUATING DENATURALIZATION

In this section, I will introduce the idea of denaturalization as it has previously been discussed in the literature, clarify some possible interpretations of denaturalization, and make explicit which interpretation I am adopting. To introduce denaturalization and clarify the interpretations of it that one could hold, I will first consider in greater depth the paradigm example of denaturalization: British abolitionism in the nineteenth century.[9]

Historically, slavery was widely seen as a natural practice without alternative. As the historian Seymour Drescher documents, for most of recorded human history, slavery has been a ubiquitous institution, viewed as "part of the natural order," and the presence of slavery was so taken for granted that its existence "set limits on how a social order could be imagined."[10] Even by the time of the eighteenth century, estimates put the number of unfree laborers (enslaved persons, serfs, and people otherwise in bondage) at 95 percent of the global population.[11] People throughout history have recognized that enslaved people suffer greatly. Bernard Williams observes that, in ancient Greece, people who were slaveowners or otherwise benefited from slavery nonetheless "granted that [slavery] was intensely unpleasant for the slaves."[12] In the same vein, Thomas Haskell emphasizes that "the suffering of slaves had long

9 I am focusing on the case of British abolition because this is the case most commonly discussed by proponents of denaturalization. In doing so, I am not claiming that abolitionist movements in other countries were less important or less instrumental in eventually ending legalized slavery worldwide. Thanks to an anonymous reviewer for pointing out this potential unclarity.

10 Drescher, *Abolition*, ix. The ubiquity of slavery is also made apparent in Holslag, *A Political History of the World*, especially 540, 551, 555–56.

11 Drescher, *The Mighty Experiment*, 14.

12 Williams, *Shame and Necessity*, 109.

been recognized" before the eighteenth century, but this recognition had not previously led to "active opposition to the institution of slavery."[13] In addition, articulated arguments against slavery go back at least to the time of Aristotle.[14] Thus, prior to abolition, the suffering of enslaved people was recognized, and arguments that slavery was immoral had long been articulated, but these factors did not lead to any sustained efforts to abolish slavery.

Why was this the case? Proponents of denaturalization argue that people often thought that slavery was a necessary economic institution without which it was impossible to produce a social surplus and that this perception made abolishing slavery an unacceptable idea.[15] Bolstering this claim is the observation that moral arguments *in favor* of slavery (often referring to the purported moral responsibility of slave owners and/or the racial inferiority of enslaved people) were quite uncommon until the mid-eighteenth century.[16] Pleasants argues that this lack of positive justifications for slavery until very late in the institution's history is indicative of the fact that for the majority of that history, it was simply taken for granted: for most of its existence, slavery was seen as a "natural, necessary, and inevitable feature of the social world."[17]

In the eighteenth century, wage labor became increasingly widespread. This provided a salient alternative institution to slavery: after all, it was obvious that a substantial social surplus could be produced via the institution of wage labor. This "cracked" the "veneer of naturalization" that had previously attached to the institution of slavery.[18] Prior to the British abolition of slavery in 1833, specific experiments with wage labor had been trialed in former slave plantations in Barbados in the 1780s and 1790s; in Trinidad in 1806 and subsequently in 1812–15 when American former enslaved persons settled there; in Sierra Leone from 1792 onwards; and most notably, in Venezuela in the 1830s, where the number of enslaved persons had been drastically reduced due to legislated freedom

13 Haskell, "Convention and Hegemonic Interest in the Debate over Antislavery," 848.

14 Cambiano, "Aristotle and the Anonymous Opponents of Slavery."

15 Anderson, "Social Movements, Experiments in Living, and Moral Progress," 14–15; Williams, *Shame and Necessity*, 111–13, 124–25; Pleasants, "Moral Argument Is Not Enough" and "The Structure of Moral Revolutions"; and Hermann, "The Dynamics of Moral Progress."

16 Brown, *Moral Capital*, 35–36, 52; and Jamieson, "Slavery, Carbon, and Moral Progress."

17 Pleasants, "Moral Argument Is Not Enough," 166; see also 165n4. I will explore further in section 4 how instances of denaturalization can lead to the emergence of ideological justifications for continued injustice.

18 Hermann, "The Dynamics of Moral Progress," 307; and Jamieson, "Slavery, Carbon, and Moral Progress," 180.

at birth, and agricultural output had been flourishing.[19] These instances of wage labor replacing slave labor were appealed to in parliamentary debates on whether or not to abolish slavery in the British Empire. Proponents touted the proposed Slavery Abolition Act as a "mighty experiment" in free labor that would have morally weighty consequences for as yet unborn subjects of the British Empire and for the "welfare of millions of slaves in foreign colonies."[20] Opponents disagreed, calling it "a procedure with disproportionate social risks—a 'mere,' 'hasty,' or 'dangerous' experiment."[21] More generally, British abolitionists, though often respected members of the bourgeoisie (and thus deeply involved in the wage labor system), were often "denounced as quixotic knights-errant, as pious charlatans all too happy to ruin the empire with costly and disastrous experiments in social engineering."[22] The Slavery Abolition Act was passed in 1833, although enslaved people in the British Empire were not in fact freed until 1838 when campaigns to end the transitionary apprenticeships that continued to bind former enslaved persons to their former masters were successful.[23] For proponents of denaturalization, the morally transformative abolition of slavery came about, at least in significant part, because the emergence of widespread wage labor denaturalized the institution of slavery and thus enabled moral criticism of slavery to become effective and led to the abolition of the practice.[24]

Before moving on to consider how we might understand the notion of naturalness and necessity, I will consider a reasonable response to this historical narrative of British abolition: Why does it focus so much on the perceptions and actions of slaveholders and others who benefitted from or tolerated slavery rather than focusing on the perceptions and actions of enslaved people? After all, it is plausible that enslaved people have always known that slavery is wrong and have always been motivated to overthrow the institution. The issue is that due to their position of extreme disadvantage relative to their enslavers,

19 Drescher, *The Mighty Experiment*, 91–94, 108–20.

20 Drescher, *The Mighty Experiment*, 123; and Anderson, "Social Movements, Experiments in Living, and Moral Progress," 17–18.

21 Drescher, *The Mighty Experiment*, 124.

22 On abolitionists often being members of the bourgeoisie, see Haskell, "Capitalism and the Origins of the Humanitarian Sensibility," 341–46. See also Davis, *The Problem of Slavery in the Age of Revolution*, 81–82. On the denouncements that they were subject to, see Brown, *Moral Capital*, 10.

23 Drescher, *Abolition*, 264.

24 Anderson, "Social Movements, Experiments in Living, and Moral Progress," 15–24; Pleasants, "Moral Argument Is Not Enough," 175–76; and Hermann, "The Dynamics of Moral Progress," 306–7.

enslaved persons have almost never successfully overthrown slavery through their own actions—with the very notable exception of the Haitian Revolution.[25] For instance, around the time of the Slavery Abolition Act being passed in Britain, the "Baptist War" erupted in Jamaica. It was the largest slave rebellion in the history of the British Empire, involving one-fifth of the population of enslaved people on the island (nearly sixty thousand people). However, this uprising lasted only eleven days, from December 25, 1831, to January 5, 1832, due to the limited power of enslaved people to resist heavily armed colonial militias.[26] As such, an explanation for the abolition of slavery, in the British case and likely in other cases besides, must extend beyond the agency of enslaved people to include the agency of the people who were not enslaved.

The notion that slavery for most of its history was seen as a "natural, necessary, and inevitable feature of the social world" is a complex one. For one thing, naturalness, necessity, and inevitability are not identical concepts. To provide a more detailed model of denaturalization, it is necessary to disambiguate what proponents of the mechanism have in mind when they claim that a certain practice or institution such as slavery was seen as a "natural, necessary, and inevitable feature of the social world." To disambiguate naturalness, I will propose three distinct interpretations of what could be meant when someone claims that a practice or institution is natural or necessary in order to defend the idea that it should not be changed. In doing so, I am offering a rational reconstruction of the different meanings that one could draw upon in defending the claim that some practice or institution is natural, in order to see which of these interpretations best fits existing discussions of denaturalization. Naturally, what people have in mind when they claim that a practice or institution is natural may be ill defined, confused, or inchoate, and so their claim may not fit neatly into any of the three categories described below. However, if such claims were to be better defined, made less confused, and clarified, then, I claim, they would fall into one of the following categories:

> *Impossibility*: To say that a practice is natural or necessary is to claim that it cannot be changed. This type of necessity can be understood easily in other domains. For instance, given our current understanding of the terms and current level of technology, it is impossible for a piglet to mature into a cow. If it is claimed that a practice or institution is natural or necessary in this sense of the term, then it follows from the principle

25 James, *The Black Jacobins*, ix; Drescher, *Abolition*, 174; and Popkin, *A Concise History of the Haitian Revolution*.

26 Drescher, *The Mighty Experiment*, 121, and *Abolition*, 260–64.

that ought implies can that one ought not to try to change that practice or institution.

Costs: To say that a practice is natural or necessary is to claim that attempts to change that practice will come with perhaps unbearably high costs. It could be that the practice or institution is functionally necessary to secure some desirable outcome or that there are not any viable alternatives for fulfilling this function, and thus attempting to change this practice or institution will lead to costs in the form of the desirable outcome not being achieved. It could also be the case that changing the practice will come with transition costs that are deemed too high.

Natural Is Good: To say that a practice is natural or necessary is to say that it is good. For instance, according to certain traditional Aristotelean views, finding out that the function of human sexual organs is to facilitate reproduction directly implies that the ethical purpose of human sexual activity is reproduction. With regard to slavery, David Brion Davis claims that "for the [ancient] Greeks (as for Saint Augustine and other early Christian theologians) physical bondage was part of the cosmic hierarchy, of the divine scheme for ordering and governing the forces of evil and rebellion."[27] More generally, cosmologies in hierarchical agricultural societies have often emphasized the divinely or cosmically ordained nature of hierarchical social institutions, such that challenging these institutions would be against the natural order of things and thus wrong.[28] These are examples of natural-is-good-type explanations for why practices or institutions are natural or necessary and thus should not be changed.

Which of these three interpretations do proponents of denaturalization have in mind? I argue that of these three interpretations, the costs interpretation is the best fit. For instance, when discussing the views that people have historically held about slavery, philosophers tend to emphasize the indispensable social role that slavery was thought to play in producing a social surplus. The idea is that people in slaveholding societies believed that, as a matter of functional necessity, without forced labor people would voluntarily work only enough to secure their own subsistence, and therefore there would be no social surplus. Without a social surplus, all forms of manufacturing that require investment,

27 Davis, *The Problem of Slavery in the Age of Revolution*, 42. But see Williams, *Shame and Necessity*, ch. 5 for a perspective that attributes this cosmological view mainly to Aristotle rather than to ancient Greek society at large.

28 Acemoglu and Johnson, *Power and Progress*, 121.

as well as the social roles of magistrates, clergy, educators, writers, artists, and scientists, could not be sustained. In combination, these claims amounted to the belief that slavery was necessary to sustain civilization.[29] Pleasants seems to hold this interpretation. He rejects the impossibility interpretation, most clearly in his discussion of the work of Michelle Moody-Adams. Moody-Adams attributes the impossibility interpretation to people who claim that perceptions of naturalness and necessity have upheld unjust social practices and institutions. She then argues that such claims must be bogus because it is not possible for competent language users to truly think that any of their social practices are necessary, because their ability to negate statements implies their ability to imagine social states in which any particular practice does not exist.[30] Pleasants (in my view rightly) responds that this is an implausibly strong interpretation of what it means to interpret some social practice as necessary, because it implies that any member of slaveholding society should have been willing to "give up slavery even if they believed that doing so would severely diminish the quality and viability of their society's way of life."[31] For Pleasants, claims about the necessity or naturalness of a practice amount to claims that there is no plausible alternative to the practice that is readily available and would not destabilize the social order and leave people "much worse off."[32] This is another example of what I have labeled the costs interpretation. As such, it seems that proponents of denaturalization claim that in the case of British abolitionism, denaturalization occurred because the alternative institution of wage labor enabled people (both those in positions of power and those in the broader public sphere who campaigned against slavery) to make their judgments about the costs of abandoning slavery more accurate: this cracked the veneer of naturalization.

For the rest of this paper, I will therefore adopt the costs interpretation as the understanding of what it means to claim that a practice is natural, necessary, or indispensable. However, before proceeding, a little more should be said about the natural-is-good interpretation. While I believe that costs and natural-is-good are conceptually distinct senses of naturalness, this does not mean that, on a psychological level, they are separate. It could well be that beliefs about an institution or practice being inevitable or very costly to abandon in

29 Anderson, "Social Movements, Experiments in Living, and Moral Progress," 16–17. Anderson is not claiming (and neither am I) that *if* this belief about the functionality of slavery was epistemically justified *then* the practice itself would be morally justified. Rather the claim is that this belief about the functionality of slavery had an effect on people's willingness to consider abandoning the practice.

30 Moody-Adams, *Fieldwork in Familiar Places*, 100.

31 Pleasants, "Moral Argument Is Not Enough," 169.

32 Pleasants, "Moral Argument Is Not Enough," 169.

a descriptive sense can foster beliefs about that institution or practice being morally good.[33] In that case, in order to fully understand historical instances of denaturalization, we need, in addition to a costs perspective, an account of how natural-is-good beliefs operate and how they can be overcome. Due to space constraints, I will focus only on an understanding of denaturalization that uses the costs interpretation, but this is not because I think that this is the only interpretation worthy of investigation.

As it stands so far, the idea that moral progress can be facilitated by people coming to have more accurate beliefs about the costs of abandoning their institutions is an intriguing one. However, this notion is currently vague. Exactly how should we understand these "costs"? How can we understand institutions being compared in terms of the benefits they provide and hence the costs of abandoning one to move to the other? And, given that the costs interpretation is a rational reconstruction of naturalness claims, is it psychologically realistic to think that people have something like these kinds of judgments about the costs of abandoning their institutions? In the following two sections, I will offer answers to these questions and, in doing so, develop a more detailed account of denaturalization.

2. Denaturalization as Improving Costs Judgments

Given the interpretation of naturalness settled on in the previous section, denaturalization occurs when an individual or group has some judgment, perhaps inchoate, about costs such that they believe getting rid of an institution will come with high costs, and then these judgments are rendered more accurate. This then facilitates a change away from that institution to a morally preferable one. Going forward, I will make use of the idea of a *costs judgment.* This is a judgment about the costs of moving from a status quo institution or practice to an alternative institution or practice. Naturally, much more needs to be said about how these costs of moving from one institution to another are to be understood. In this section, I will attempt to provide a more precise understanding of costs. I will argue that we can understand what costs judgments attempt to track using resources from the philosophy of conventionality.

33 See Jost, "A Quarter Century of System Justification Theory"; and Jost et al., "The Future of System Justification Theory." However, in section 4 below I will also explore the possibility of the opposite relationship obtaining, such that when an institution or social practice is denaturalized, this will incentivize people who benefit from that institution or social practice to produce moral justifications in its favor.

David Lewis analyzes conventions as equilibria in repeated *coordination games*.[34] Consider the following game in which the two players would like to coordinate their actions:

		Player 2	
		A	*B*
Player 1	*A*	1, 1	0, 0
	B	0, 0	1, 1

FIGURE 1 Simple coordination game

The game has two players (1 and 2) and two strategies (*A* and *B*) that yield certain payoffs. It is standard to interpret payoffs as representing preference rankings expressed in terms of *utility* in the rational choice sense of the term.[35]

In this game, the players are able to coordinate if they both choose the same strategy: if they either both play *A* or both play *B*. If the players coordinate, for example, by both playing *A*, then they have reached what Lewis refers to as a *proper coordination equilibrium*. In such a situation, neither player can improve their own payoff by unilaterally switching strategies, and neither player can improve the payoff for the other player by unilaterally switching strategies. Settling on *A*/*A* as a strategy is a convention because it is arbitrary: the players would have been just as well-off if they played *B*/*B* instead. However, if the game is played repeatedly, then once the *A*/*A* pattern emerges, it is a stable equilibrium because it is a proper coordination equilibrium: each player has a strong incentive to keep playing *A* because they cannot benefit themself or the other player by unilaterally switching to *B*.

Institutions can also be illuminated using this theoretical apparatus. Institutions can be modeled as sets of (often formalized) norms that, along with incentives and expectations, coordinate people's actions and thus stabilize patterns of behavior. Because of this stabilizing function, institutions can be understood as the (conventional) equilibria of repeated coordination games, as in figure 1.[36] Of course, a model of any actually existing institution would be vastly more complex than figure 1, involving many more players and many more possible outcomes.

34 Lewis, *Convention*.

35 Guala, *Understanding Institutions*, 21–22; and Gaus, *On Philosophy, Politics, and Economics*, ch. 2. See also Kogelmann, "What We Choose, What We Prefer," for a recent and sophisticated account of how to understand preference rankings.

36 Guala, *Understanding Institutions*, ch. 2.

Why think that looking at conventions can give us insight into how costs judgments can be modeled and that this insight might help us think more clearly about denaturalization? The institution of slavery was clearly conventional in the sense that it was a coordination equilibrium that could have been (and now is) otherwise. A pertinent question is how to understand the institution of slavery using the kind of model sketched above. If we model the institution of slavery as an equilibrium in a complicated coordination game, then are enslaved people counted as players in this game? Francesco Guala argues that slavery, seen as an equilibrium to a coordination game that includes enslaved people, can be seen as generally beneficial in the technical and circumscribed sense that the real alternative to being subject to the institution of slavery for many people throughout history has been being killed.[37] This claim seems to assume that enslavement is preferable to being killed according to the utility function of enslaved people. However, it is not clear that this claim is plausible. For one thing, Guala's characterization of alternatives may be inaccurate: in some contexts, the alternative to enslavement may not have been death or the risk of death but rather (the risk of) severe punishment. For another, even in cases where (the risk of) death was the alternative to enslavement, we have plenty of evidence that the demand for liberty from enslavement sometimes motivated enslaved people to take up arms against their enslavers in the face of fearsome odds of death, which suggests that the arrangement was not always beneficial even in Guala's circumscribed sense.[38]

When trying to use this understanding of institutions as the equilibria of repeated coordination games to understand the costs judgments of people who accepted slavery, I think it makes most sense to think of enslaved people as not being players in the game. The costs of abandoning slavery are thought to be costs for people who are not enslaved, and it is these perceived costs that affect the views and actions of people who directly benefit from or tolerate the institution of slavery. However, when considering whether denaturalization is always or generally morally progressive in section 4, this issue of who is included in the set of people whose costs judgments become more accurate will be very important.

I have now described a view of institutions according to which they can be modeled as the equilibria of repeated coordination games. These equilibria are conventional when they are arbitrary, and they are arbitrary when alternative coordination equilibria are possible. If we link this account of institutions to the description of denaturalization given in section 1, then we can say proponents

37 Guala, *Understanding Institutions*, 4–5.

38 James, *The Black Jacobins*; and Popkin, *A Concise History of the Haitian Revolution*.

of denaturalization hold that many past people had a false view of the institution of slavery, according to which it was, in some sense, not conventional: it was rather a "natural, necessary, and inevitable feature of the social world." Understanding more about what conventions are can help us understand the way in which past persons were mistaken, and this can help us understand how denaturalization, understood through the lens of the costs interpretation, might operate by correcting these mistakes.

Recent work by Mandy Simons, Kevin Zollman, and Cailin O'Connor provides this understanding by giving more insight into the notion of conventionality.[39] They suggest that the arbitrariness of a convention is not a binary matter. Instead, it can vary depending on three factors:

1. *Payoffs:* Some conventions have higher payoffs than others.[40]
2. *Stability:* Some conventions are more stable than other conventions in that they can tolerate a greater amount of deviance (people failing to play the conventional strategy) before the convention collapses to be replaced by another.
3. *Likelihood of Emergence:* Some conventions are more likely to emerge than others, either because there are only a small number of possible conventions or because some convention is more attractive to players due to higher payoffs, shared cultural norms, or cognitive biases.

To understand the first factor, consider the following game.[41]

		Player 2	
		A	*B*
Player 1	*A*	1, 1	0, 0
	B	0, 0	*x*, *x*

FIGURE 2 Coordination game in which *B*/*B* is the preferable equilibrium if $x > 1$.

Let us assume that $x = 100$. In this case, both *A*/*A* and *B*/*B* are proper coordination equilibria as defined above, and so they would both be candidates to be conventions on Lewis's account. However, if the players were to settle on

39 Simons and Zollman, "Natural Conventions and Indirect Speech Acts"; and O'Connor, *The Origins of Unfairness* and "Measuring Conventionality."

40 Another way of putting this is that some conventions are Pareto-superior to others. See Simons and Zollman, "Natural Conventions and Indirect Speech Acts," 7.

41 O'Connor, "Measuring Conventionality," 582.

the *B*/*B* equilibrium, then although their choice is arbitrary in that it *could* have been otherwise (*A*/*A* is also a proper coordination equilibrium), the explanation of why the players settled on *B*/*B* will likely involve an appeal to the much higher payoffs of *B*/*B*. Thus, while *B*/*B* is arbitrary in some sense, there is also a strong functional explanation available for why *B*/*B* might come to dominate as a strategy over *A*/*A*. Furthermore, if players who were playing *B*/*B* were asked to move from that equilibrium to *A*/*A*, then they could truly claim that that transition would come with very large costs because (again, assuming that $x = 100$) *A*/*A* has such low payoffs relative to *B*/*B*. Here we can see how a claim about payoffs can be related to a costs judgment.

Figure 2 can also help us understand the second factor, stability. If $x = 100$, then *A*/*A* will be a relatively unstable equilibrium. Why is this? Because if a population is playing *A*/*A*, then it will take only a relatively small percentage of the population defecting to playing *B*/*B* for the *A*/*A* equilibrium to collapse.[42] Regarding the third factor, there are several things that affect the likelihood of a convention emerging. For one thing, the likelihood of a given practice emerging depends on how many proper coordination equilibria exist with regard to that practice. For instance, imagine that figure 1 represents two possible conventions: driving on the left-hand side of the road and driving on the right-hand side of the road. Both conventions are proper coordination equilibria. Driving on the left-hand side of the road is arbitrary, but it is not *that* arbitrary because there is only one other proper coordination equilibrium: driving on the right-hand side. However, if we are dealing with a coordination game in which there are many different proper coordination equilibria (assuming, for now, that these equilibria have equivalent payoffs), then any given equilibrium will be more arbitrary simply because there are more possible alternative equilibria. Thus, we might say that the more proper coordination equilibria there are in a coordination game, the more arbitrary the emergence of any particular equilibrium is because there are more ways that this convention could have been otherwise.[43] The payoffs of a convention can also influence its likelihood of emerging, particularly due to the fact that a convention with higher payoffs is more likely to be adopted and more likely to spread from one social group to another.[44] Lastly, the likelihood of a convention emerging can be affected by perceptual, cognitive, or cultural biases that make a particular convention more salient for the relevant population.[45]

42 Simons and Zollman, "Natural Conventions and Indirect Speech Acts," 7–9; and O'Connor, "Measuring Conventionality," 584.

43 O'Connor, "Measuring Conventionality," 582.

44 Cohen, "Cultural Variation," 464; Henrich, *The WEIRDest People in the World*, 88–99; and O'Connor, "Measuring Conventionality," 584.

45 Guala, *Understanding Institutions*, 14–16.

We now have three factors that can influence the degree to which a practice or institution is conventional. How can this understanding of conventions inform our understanding of costs judgments? We can think of a costs judgment as the claim that changing an existing institution will result in drastically lower payoffs and/or that alternative institutions will be unstable and so unable to coordinate people's behavior in order to deliver acceptable payoffs. Thus, the first factor, payoffs, is directly relevant to the accuracy of a costs judgment: if a status quo institution provides the highest possible payoffs, and abandoning it will result in very low payoffs, then one can have an accurate costs judgment that abandoning that institution would come with heavy costs. This is a more abstract and precise way of articulating the kind of belief that Anderson, Jamieson, Hermann, and Pleasants attribute to people who thought that slavery was a natural, necessary, or indispensable institution: although, of course, in this case, the costs judgment was inaccurate. Stability is also relevant, because if an alternative institution is highly liable to defection and thus highly unstable, then this instability might result in significant costs when the institution collapses. This would make the alternative institution undesirable in terms of payoffs, relative to the status quo institution. The relevance of the third factor, the likelihood of emergence, is less clear. It seems relevant for costs judgments than an institution is likely to emerge because it has high payoffs, but this is just an indirect way of talking about the first factor. However, it does not seem directly relevant to assessing the costs of moving away from a given institution or practice that it is a convention that was highly likely to emerge due to the shared cognitive biases or cultural norms of the population that has that practice or institution. This would be relevant to a costs judgment only if these same cognitive biases or cultural norms mean that there would be costs involved in transitioning away from said institution. However, that the status quo institution is supported by shared cognitive biases or cultural norms may be very relevant for explaining why groups may be reticent to move away from the status quo, as will be explored further in section 3.

This model from the philosophy of conventionality gives us a clearer way of thinking about the features of practices and institutions that costs judgments attempt to track—namely, their payoffs and stability. If we have this understanding of costs judgments, then denaturalization would function by making them more accurate. Therefore, one important empirical assumption made by the account of denaturalization that I have developed is that people have judgments that, in some way, attempt to track the payoffs of their own institutions and social practices relative to alternatives. Fully developing an account of what these judgments are and how they attempt to track payoffs is too large a task to attempt in a paper of this length, although I will make a limited case for the

psychological realism of this account of denaturalization in section 3. For now, my point is that for the costs account of denaturalization sketched above to be a plausible causal mechanism of moral progress, we need a satisfactory account of what such payoff-tracking judgments are and how they work. Alternatively, we need to develop an alternative costs account of denaturalization to the one developed here that can explain what the relevant costs are and does not need an account of payoff-tracking judgments, or to develop an account of denaturalization that does not adopt the costs interpretation of naturalness and necessity but rather some other interpretation.

Assuming some psychological account of how people's costs judgments track the payoffs of institutions and social practices, how might costs judgments be made more accurate? According to proponents of denaturalization, exposure to existing alternative institutions can make costs judgments more accurate. Exposure to these alternative institutions can provide information about the payoffs and stability of alternatives, which can denaturalize the status quo institution by making it clear that abandoning this institution will not lead to unbearably high costs in terms of loss of payoffs. Once costs judgments are rendered more accurate, moral considerations can then play more of a role in motivating people to change their institutions. One implication is that the ability of people to improve the accuracy of their costs judgments is bounded by the actual alternative institutions that exist: without actual alternatives, one cannot assess the relative payoffs of alternatives to the status quo. On this account, people who tolerated or supported slavery before the emergence of widespread wage labor had an inaccurate costs judgment to the effect that a social surplus was not possible without slavery (which we now know is possible), but surveying existing alternative institutions at the time would not have provided the kind of information needed to update this costs judgment. Thus, this model of denaturalization implies that there are great benefits to engaging in institutional experimentation because such experiments in living are the only way to provide the evidence about payoffs and stability of alternative institutions that are vital to improve the accuracy of costs judgments and to potentially achieve denaturalization.[46]

46 On the value of institutional experimentation, see Anderson, "Social Movements, Experiments in Living, and Moral Progress"; Müller, "Large-Scale Social Experiments in Experimental Ethics"; and Robson, "The Rationality of Political Experimentation." Naturally, engaging in such experimentation may have diminishing returns, and the costs account of denaturalization says nothing about the opportunity costs of engaging in institutional experimentation. Nonetheless, the costs account does imply that there are strong *pro tanto* reasons to engage in institutional experimentation.

There is a complication worth noting here. Suppose that a given group forms the judgment, perhaps based on some small-scale institutional experiments, that moving to an alternative and more just institution will not be prohibitively costly for them, and this judgment makes moral criticism of the status quo institution more effective and facilitates a transition to a new institution. However, it then transpires that this costs judgment was wrong. Moving to this new institution, while it is *ex hypothesi* more just, has much lower payoffs for them than the status quo institution, such that the institutional change is perceived to be prohibitively costly. In this case, what cracked the veneer of naturalization of the status quo institution was not that the group in question came to have more accurate beliefs about the costs of moving to an alternative institution but rather that they *believed* that moving to the alternative institution would not have prohibitive costs.[47] Anderson points out that in the case of British abolitionism, a group of British elites extrapolated their judgments about the payoffs of abolishing slavery based on small-scale experiments in abolition (as described in section 1), but for at least some of these people, their expectations of increased productivity in the lucrative British sugar colonies of the Caribbean following abolition (better payoffs from the new institution as compared to the old) were disappointed.[48] In other words, their belief about improved payoffs from moving to an alternative institution was false, but this belief still facilitated a transition to a more just institution. So do more accurate costs judgments really matter for facilitating institutional change, or is what matters simply that people who would otherwise resist those changes come to believe that those changes will not be prohibitively costly for them, even if they are wrong?

I believe that more accurate costs judgments are in fact important if durable institutional change is to be obtained. If people have mistakenly optimistic judgments about the costs of moving to alternative institutions as described above, then while this may facilitate institutional change, it is also likely to lead to backlash once it becomes clear that the new institution has prohibitively high costs. I submit that institutional change is likely to be more durable if people's projections of the costs of moving to alternative institutions are at least relatively accurate, so that it is true that the more just institutions are not prohibitively unstable and do not deliver unacceptably low payoffs. Returning to the example of British abolitionism, this was by and large the case. Despite the mistaken beliefs described by Anderson of some British elites regarding the relative productivity of wage labor versus that of slave labor, Pleasants makes clear that the "abandonment of slavery for the newly emerging paradigm of freely

47 I thank an anonymous reviewer for drawing my attention to this kind of case.

48 Anderson, "Social Movements, Experiments in Living, and Moral Progress," 18–20.

contracted wage labour served the medium- to long-term economic interest of the liberators spectacularly well."[49] In the medium to long term, wage labor as an alternative institution to slavery did not deliver unacceptably low payoffs, and the fact that this *was* the case (as opposed to people merely projecting, wrongly, that it *would* be the case) can reasonably be thought to have played a role in ensuring that the morally progressive transition from slavery to wage labor has been sustained.

When forming costs judgments about moving from a status quo institution to a more just institution based on small-scale institutional experiments, we must recognize that our costs judgments are always going to be projections, and we will not know whether our costs judgments have truly become more accurate except in hindsight. However, if I am correct about the importance of more accurate costs judgments, then this implies that great attention should be given to the potential pitfalls of extrapolating incorrect predictions from small-scale experiments with alternative institutions because if our costs judgments only appear to have become more accurate rather than really becoming so, then this could facilitate unstable moral progress and dangerous backlashes.

I have now explicated an account of denaturalization, the costs account, that is more detailed than the descriptions of denaturalization thus far offered in the literature. My account is explicit about the interpretation of naturalness being used, shows how this kind of naturalness can be understood using resources from the philosophy of conventionality, and shows how people can be mistaken about the naturalness of their institutions in degrees. However, in Popperian fashion, making the hypothesis that denaturalization is a causal mechanism of moral progress more detailed and specific does not necessarily make it more convincing; instead, it brings into sharp relief the various points of criticism that can be leveled against the account. I see this as an entirely good thing, if one's aim is to advance our knowledge about this proposed mechanism of moral progress. In the following section, I will add more detail to the account by making a brief case for its psychological realism.

3. The Psychological Realism of Denaturalization

While the costs interpretation of denaturalization is a rational reconstruction, it is nonetheless the case that denaturalization is meant to at least partially explain real processes of historical change. For this to be plausible, it must be the case that the costs interpretation is rooted in some real psychological mechanisms that explain people's behavior. What needs to be established in order to believe

49 Pleasants, "The Structure of Moral Revolutions," 591.

that the account of denaturalization offered above is psychologically realistic? We would need to establish that people have psychological states that are similar to what I have been calling costs judgments—judgments that attempt to track the payoffs and stability of their institutions and practices relative to available alternatives.[50] Further, we should have evidence that people's costs judgments can become more accurate through being exposed to alternative institutions and practices: this would be evidence that experiments in living can provide correctives to inaccurate costs judgments, thus denaturalizing status quo institutions.

In this section, I will provide some evidence for the psychological realism of my account of denaturalization, with the proviso that more evidence would need to be provided to truly vindicate the account. Nonetheless, this section provides a sampling of the kind of evidence needed to support an account of denaturalization like the one outlined in section 2 or any similar account that takes the costs interpretation of naturalness described in section 1.

Firstly, do humans actually keep track of the payoffs and stability of their institutions relative to alternatives? Evidence from anthropology and cultural evolutionary theory suggests that they do. One source of evidence is research on *subjective selection*. Subjective selection refers to the selective retention of beliefs, practices, and other cultural variants that people subjectively evaluate as being useful, especially for fulfilling their goals.[51] In addition to explaining how people selectively retain or reject things like hunting practices and tools, subjective selection also affects the selective retention of rules and norms that are perceived to satisfy the interests of those who are in positions to build, maintain, and enforce rules and norms.[52] As a mechanism of cultural change, subjective selection requires that people have psychological states that track the subjective costs and benefits of different beliefs and practices. These psychological states are similar to those that I have described as costs judgments.

Another source of evidence comes from research on intergroup competition. Joseph Henrich describes how cultural evolution can give rise to packages of prosocial norms and institutions through a process of intergroup competition.[53] There are numerous ways in which competition between groups with

50 That people have these kinds of psychological states is an important presupposition of the costs account of denaturalization. If, instead, people typically do not make such assessments of status quo institutions, then this would count against the costs interpretation.

51 Singh, "Subjective Selection and the Evolution of Complex Culture," 266.

52 Singh, "Subjective Selection and the Evolution of Complex Culture," 267, 272–73; and Singh et al., "Self-Interest and the Design of Rules."

53 By 'prosocial', Henrich means norms and institutions that lead to success in intergroup competition, for instance by fostering cooperation or internal harmony within the in-group. See Henrich, *The Secret of Our Success*, 169.

different norms and institutions can lead to the spread of more prosocial norms and institutions, but two in particular are relevant for the purposes of this paper: *differential migration* and *prestige-based group transmission*.

Differential migration describes the process in which individuals preferentially migrate to more successful groups whose norms and institutions create "greater internal harmony, cooperation, and economic production."[54] Of course, greater internal harmony, higher levels of cooperation, and greater economic production are all things that contribute to higher payoffs and greater stability in the senses described in the previous section.[55] This suggests that people who are migrating preferentially to more successful groups have judgments that, at least to a large extent, track the payoffs and stability of the institutions and practices of the group that they migrate to relative to the institutions and practices of their original group. These judgments appear to approximate costs judgments.

Prestige-based group transmission occurs when individuals in one group preferentially attend to and copy the social norms of other, more successful, groups.[56] Where the individuals in the copying group also have the ability to legislate norm and institution change for their entire group, this can also result in an entire group adopting the norms and institutions of a more successful group. Henrich offers the example of a community in New Guinea called Ilahita who in the late nineteenth century copied a package of rituals, religious beliefs, norms, and institutions (collectively called the *Tambaran*) from a militarily successful group called the *Abelam*, whose expansion was a potential threat to Ilahita. The Tambaran was already being adhered to by the Abelam, and it was thought by Ilahita's elders that the Tambaran was the source of the Abelam's success. By copying the Tambaran and making some felicitous errors in how they copied it, Ilahita ended up not only matching but surpassing the military might, level of cooperation, and scale of the Abelam.[57] Prestige-based group transmission suggests that people within groups have judgments about the relative payoffs (often in terms of military might or level of cooperation) of their institutions and practices and the institutions and practices of other groups, and where the institutions or practices of other groups are superior, people are sometimes motivated to copy them.[58] These judgments also appear to approximate costs judgments.

54 Henrich, *The Secret of Our Success*, 168.

55 Heath, "The Benefits of Cooperation."

56 Henrich, *The Secret of Our Success*, 168.

57 Henrich, *The WEIRDest People in the World*, 88–99.

58 One crucial caveat about prestige-based group transmission is that the link between the practices and institutions of other groups and the desirable higher payoffs of these practices and institutions is often causally opaque: it is not clear which practices or institutions

Additionally, we need to explain how costs judgments can be made more accurate through exposure to alternative institutions. In part, this explanation is provided by the above account of forms of intergroup competition in which people acquire information about the payoffs of alternative institutions. However, denaturalization is meant to work by correcting inaccurate costs judgments. What factors could make costs judgments inaccurate, such that denaturalization can then act to make them more accurate? Firstly, people could simply lack knowledge about other possible institutions that have equivalent or higher payoffs than their status quo institutions. Secondly, the cultural evolutionary framework referred to above may support the idea the people have something like costs judgments, but it also suggests that humans have a norm psychology that makes social norms and institutions difficult to change because people are often intrinsically motivated to follow the norms that they grew up with and to punish norm violations. Punishment can then render systems of norms stable against shocks, including deliberate attempts to change such systems.[59] To the extent that people's intrinsic motivation to follow their status quo norms and their motivation to punish norm violations can bias their perception of the costs of changing their status quo norms, practices, or institutions, these factors could contribute to explaining why costs judgments can be inaccurate.

Thirdly, people could underestimate the payoffs of moving to an alternative practice or institution and thus overestimate the costs of moving from the status quo to the alternative. This possibility is suggested by the phenomenon of *loss aversion*, in which the risks of loss associated with changing away from the status quo can weigh much more heavily in people's minds than the prospective gains associated with change—a particularly important error when it comes to making accurate costs judgments.[60] Loss aversion has recently been challenged on a number of grounds: that much of the evidence for loss aversion has been overinterpreted because there are other interpretations of these results that do not support the existence of loss aversion, and that whether or not losses are weighed more heavily depends on the context of choice.[61] But

are causally responsible for the perceived success. As a result, when people choose to copy the practices or institutions of other groups, they tend to copy quite indiscriminately, adopting many such practices and institutions rather than adopting only the ones that contribute to the higher payoffs in a targeted way (Henrich, *The WEIRDest People in the World*, 97).

59 Kelly and Davis, "Social Norms and Human Normative Psychology," 63–64; Henrich, *The Secret of Our Success*, ch. 9; and Boyd and Richerson, "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups."

60 For classic descriptions of loss aversion, see Kahneman and Tversky, "Choices, Values, and Frames"; and Kahneman et al., "Anomalies."

61 Gal and Rucker, "The Loss of Loss Aversion"; and Yechiam, "Acceptable Losses."

more recently, high-powered studies have demonstrated that loss aversion is a robust phenomenon, even when dealing with small losses, but also that loss aversion has moderators: there are some features of decision makers that can attenuate loss aversion.[62] More educated decision makers are less prone to loss aversion than less educated ones, older decision makers are more prone to loss aversion than younger ones, and people with more experience and knowledge about the decision domain in question are less prone to loss aversion than those with less experience. This last moderator in particular suggests that experience with relevant alternatives can aid in making costs judgments more accurate by mitigating loss aversion, which bolsters the case that institutional experimentation can contribute to denaturalization.

The evidence presented in this section makes a preliminary case for the psychological realism of my account of denaturalization by arguing that people have psychological states that approximate costs judgments; that there are psychological factors, including how human norm psychology works and our vulnerability to loss aversion, that can explain why costs judgments can be inaccurate; and that exposure to alternative institutions can make costs judgments more accurate. Given the brevity of this presentation of evidence, we of course cannot say conclusively whether the account is psychologically realistic. However, this section nonetheless gives an indication of the kind of evidence that would be needed to demonstrate that an account of denaturalization (especially one based on some version of the costs interpretation) is realistic. Future accounts of denaturalization should try to provide similar and ideally more advanced evidence for their psychological realism.

4. Denaturalization and Moral Progress

So far, I have analyzed denaturalization as it has been proposed in the literature; argued that denaturalization works by making costs judgments more accurate; provided a model of how we can understand what costs judgments aim to track; and provided evidence that my account of denaturalization possesses a degree of psychological realism. Taken together, this gives us an account of denaturalization that is more detailed and specific in its claims than previous discussions of denaturalization in the literature. I hope that this account can be critically assessed and improved upon in future philosophical work.

In this last section, I will assume that the costs account of denaturalization is correct in order to situate denaturalization as a cause of moral progress within

62 Ruggeri et al., "Replicating Patterns of Prospect Theory for Decision Under Risk"; and Mrkva et al., "Moderating Loss Aversion." In Ruggeri et al., $n = 4{,}098$ participants from nineteen countries; and in Mrvka et al., $n = 17{,}720$ across five unique samples.

the broader moral progress literature and attempt to answer a key question: Will denaturalization always or even generally lead to moral progress? After all, it could instead be a mechanism of moral change with a random moral valence or, worse, be generally biased in favor of morally regressive social change.

Before getting started, let us briefly consider the question of what it means for something to be morally progressive. Firstly, I will assume that certain cases are canonical examples of moral progress that are beyond reasonable doubt—including the abolition of slavery, gains in gender equality, and increasing recognition of the moral acceptability of same-sex relationships.[63] Secondly, I will assume that all human beings have equal moral status. Given this moral standard, social changes that result in this belief being more widely held and, correspondingly, result in people being treated equally regardless of group membership will count as moral progress.[64]

If denaturalization was a contributing cause of the British abolition of slavery, then it is hard to doubt that it was morally progressive in that specific case. However, in general, whether denaturalization will lead to moral progress depends on a number of factors. Firstly, recall that denaturalization works by making costs judgments more accurate so that a switch to an alternative institution is no longer (falsely) thought to have unacceptably high costs. With this false belief removed, moral criticism of the status quo institution can then be more effective in mobilizing change. According to this story, denaturalization alone is not sufficient for moral progress. Justified moral beliefs or values are also necessary to motivate the change away from the status quo institution and towards the morally preferable one. Thus, denaturalization can facilitate moral progress when inaccurate costs judgments that are contributing to the inefficacy of justified moral criticism are removed, but this justified moral criticism is still necessary for denaturalization to facilitate progress.

Secondly, assuming that people have justified moral beliefs or values, whether denaturalization can facilitate progress depends on the actual payoffs of alternative institutions relative to the status quo. If we imagine that in fact there were no alternatives to the institution of slavery for producing a social surplus, then if people who benefitted from or tolerated slavery came to have more accurate costs judgments, this would not facilitate progress. Rather, it

63 Buchanan and Powell, *The Evolution of Moral Progress*, 47–48, 241; Buchanan, *Our Moral Fate*, xiii; Kitcher, *Moral Progress*, 13; and Kumar and Campbell, *A Better Ape*, 181.

64 Buchanan and Powell, *The Evolution of Moral Progress*, 11–18. Questions can certainly be asked about how the standards for moral progress are justified. However, for the purposes of exploring how the denaturalization mechanism relates to the overall philosophy of moral progress, I will rely on these moral standards, which are already widely accepted in the moral progress literature.

would entrench the belief that slavery could not be abandoned without high costs. In such a case, it would better facilitate moral progress if such people came to have even more inaccurate costs judgments so that they falsely believed that alternative institutions had comparable or higher payoffs to their status quo slave institutions (though, as mentioned in section 2, such moral progress based on inaccurate costs judgments would likely be unstable). Victor Kumar and Richmond Campbell argue, paraphrasing Stephen Colbert, that "reality has an inherent progressive bias" such that when people come to have more accurate beliefs about the world around them, they tend to modify their moral norms and values in the direction of inclusion, equality, and progress.[65] For denaturalization to be reliably progressive, it must be the case that this is by and large true, so that coming to have more accurate costs judgments about the relative payoffs of unjust status quo institutions and relatively more just alternative institutions has the effect of making the status quo seem less natural, inevitable, and necessary rather than entrenching this impression. Whether this *is* largely true is a difficult question to answer: it seems like something that rather needs to be considered on a case-by-case basis. Nonetheless, it seems to be the case that whether denaturalization can facilitate progress is largely hostage to whether the facts are such that there really are more just and roughly equivalent payoff institutions. These facts in turn are influenced by factors such as:

- Which institutions happen to be available as actual alternatives, which may largely be a matter of historical happenstance.
- What the other institutions and social norms of the people who are making costs judgments are. This is important because the payoffs of any given institution or practice depends to some extent on the culture (which includes the other institutions, practices, beliefs, and social norms) of the people who will be adopting them. Because of this, there is a certain path dependency whereby some institutions that might be highly effective for one group may be much less effective for another.[66]
- What kind of technologies are available, as technologies can also alter the payoffs of different social norms and institutions.[67]

These factors, at least, are important for working out whether, given justified moral values and beliefs, denaturalization can facilitate moral progress.

Thirdly, let us return to a point briefly made in section 2 about who is in the group from whose perspective costs judgments are being made. When we

65 Kumar and Campbell, *A Better Ape*, 195.

66 Henrich, *The WEIRDest People in the World*, 98, 476–78.

67 Hopster et al., "Pistols, Pills, Pork and Ploughs," 21–22.

consider the story of British abolitionism endorsed by proponents of denaturalization, the people whose costs judgments mattered were the antislavery campaigners and the political elites in Britain, because these were the people whose beliefs were causally efficacious in legislating the end of legal slavery. In this situation, it is fortunate that *that* rather limited group updated their costs judgments to believe that they would not experience unbearably low payoffs if they switched from their unjust status quo institution. But it is easy to imagine cases in which switching from an unjust status quo institution to a more just alternative institution will lead to higher or equivalent payoffs for the majority of people affected by the status quo institution but will lower the payoffs of the group who have decision-making power to effect that switch. In this case, updating the costs judgments of that group would not facilitate moral progress because updated costs judgments, even if they showed that an unjust institution could be abandoned without significantly lowering payoffs for the majority of people affected by the institution, would not be likely to result in any institutional change. Thus, it seems that denaturalization is more likely to facilitate moral progress the more inclusive the group that gains more accurate costs judgments is and the more inclusive the decision-making procedures to secure institutional change are. So, broadly speaking, we should expect denaturalization to work better in a context of inclusive morality, where many people's interests and moral status are equally respected, and inclusive institutions, in which many people whose interests are affected by those institutions have decision-making power within them or, at the limit, have an ability to influence those with decision-making power (as was the case with petitioners during the campaigns for abolition in Britain).[68]

However, I think there is also an interesting feedback loop between the inclusivity of social norms and institutions and the effectiveness of denaturalization as a mechanism of moral progress. British abolitionism led to an expansion of the moral circle and a gain in moral inclusivity through the recognition of a basic level of moral status and securing a basic level of legal status for formerly enslaved persons, but this gain in inclusivity was driven by a non-inclusive group that was numerically dominated by non-enslaved people.[69] If

68 On the importance of equality of moral status and respect, see Buchanan and Powell, *The Evolution of Moral Progress*, 62–64; Buchanan, *Our Moral Fate*, 23–24; and Kumar and Campbell, *A Better Ape*, 184–86. On inclusive institutions, see Acemoglu and Robinson, *Why Nations Fail*, 79–83. And on the position of petitioners in the British abolition movement, see Anderson, "Social Movements, Experiments in Living, and Moral Progress," 10–15.

69 Kumar and Campbell, *A Better Ape*, 203–7; and Buchanan and Powell, *The Evolution of Moral Progress*, 57, 212–14.

denaturalization can lead to gains in inclusivity, and gains in inclusivity can then increase the likelihood that further denaturalization will lead to moral progress, then denaturalization as a mechanism of moral progress and gains in moral inclusivity as a form of moral progress may form a positive feedback loop.

Fourthly, a less morally welcome feedback loop is that successful instances of denaturalization may give rise to ideologically motivated justifications for the moral rightness of an unjust practice. Imagine that within a slaveholding society, the group of people who are either slaveholders or who tolerate slavery come to see that moving to an alternative and more just institution will not result in prohibitively high costs—for example, it will still be possible to produce a social surplus without slavery. This denaturalization will then make moral criticism of slavery more effective. Even if this is the case, it is still going to be the case that some within the group will lose substantial benefits that they currently enjoy if slavery is abolished. Supposing that slavery has been denaturalized such that it is no longer plausible that it is a natural and necessary institution (according to the costs understanding of this claim), these people will no longer be able to make uncontested claims about the naturalness of slavery as an institution without alternatives. But this does not mean that this group will no longer have an interest in slavery continuing. Rather, it means that they need to produce justifications in favor of maintaining slavery. Indeed, as described in section 1, some historians have argued that explicit moral justifications for slavery emerged only late in the history of the institution—around the time that slavery was being denaturalized by the emergence of wage labor as an alternative institution. It is plausible that many instances of denaturalization will leave some members of the group that undergoes that denaturalization with strong interests in maintaining the status quo institution and thus with strong interests in producing moral justifications for the denaturalized status quo institution. These moral justifications will be ideological in the sense that they are epistemically distorted, in this case by the self-interest of the members of the group producing them.[70] Such ideologically distorted purported moral justifications for unjust institutions may commonly emerge in the wake of morally progressive denaturalization.[71]

To sum up, it seems that denaturalization is not a mechanism that is guaranteed to facilitate moral progress. Whether denaturalization will lead to moral progress depends on the factors enumerated above: whether there are justified moral beliefs and values that will correctly identify unjust status quo

70 Barrett, "Ideology Critique and Game Theory," 714n1.

71 Thanks to an anonymous reviewer for prompting me to discuss this phenomenon in greater detail.

institutions and push for their removal after they are denaturalized; whether it is in fact the case that there are more just alternative institutions with equivalent or higher payoffs available; how inclusive the group whose costs judgments are rendered more accurate is and how many members of that group have access to decision-making power to change the unjust status quo; and whether and to what extent particular social groups are able to produce successful ideological moral justifications of the unjust status quo in the face of denaturalization.

5. CONCLUSION

Moral progress, to the extent that it occurs, is likely to evade simple monocausal explanations.[72] In that spirit, this paper can be taken as an investigation into one of the many mechanisms that have been proposed to explain past instances of moral progress and that could potentially lead to future moral progress.

I have articulated a more detailed understanding of denaturalization than has thus far been offered in the literature, so that the mechanism can be critically assessed on empirical and philosophical grounds. I have argued that denaturalization works by improving our costs judgments and that these judgments are accurate to the extent that they track the relative payoffs and stability of different institutions. I have also provided evidence for the psychological realism of this account of denaturalization, both to bolster the case for my account and to show what kind of empirical evidence would be required to make the case that denaturalization is psychologically realistic. I hope that this developed account can be critically assessed by other philosophers interested in the mechanisms of moral change and moral progress and that it can encourage the development of further accounts of denaturalization—understood as improving costs judgments, understood as a mechanism that corrects false beliefs that fit into the natural-is-good interpretation outlined in section 1, or understood in some other way. Finally, with a more detailed account of denaturalization in hand, I have investigated its potential to facilitate moral progress and laid out the factors that affect whether denaturalization is progressive after all.[73]

Utrecht University
c.t.blunden@uu.nl

72 Eriksen, "The Dynamics of Moral Revolutions." For an account of some of the difficulties that are faced by accounts of what causes moral progress, see also Rehren and Blunden, "Let's Not Get Ahead of Ourselves."

73 Many thanks to Joel Anderson, Joseph Heath, Benedict Lane, Paul Rehren, and Hanno Sauer for discussing these ideas with me and providing criticism and feedback. I would like to extend special thanks to Chiara Cecconi for inviting me to present a draft of this paper

REFERENCES

Acemoglu, Daron, and Simon Johnson. *Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity*. Basic Books, 2023.

Acemoglu, Daron, and James A. Robinson. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Currency, 2012.

Anderson, Elizabeth. "Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain's Abolition of Slavery." Lindley Lecture, University of Kansas, 2014.

Barrett, Jacob. "Ideology Critique and Game Theory." *Canadian Journal of Philosophy* 52, no. 7 (2022): 714–28.

Boyd, Robert, and Peter J. Richerson. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology* 13, no. 3 (1992): 171–95.

Brown, Christopher Leslie. *Moral Capital: Foundations of British Abolitionism*. University of North Carolina Press, 2006.

Buchanan, Allen. *Our Moral Fate: Evolution and the Escape from Tribalism*. Massachusetts Institute of Technology Press, 2020.

Buchanan, Allen, and Rachell Powell. *The Evolution of Moral Progress: A Biocultural Theory*. Oxford University Press, 2018.

Cambiano, Giuseppe. "Aristotle and the Anonymous Opponents of Slavery." *Slavery and Abolition* 8, no. 1 (1987): 22–41.

Cohen, Dov. "Cultural Variation: Considerations and Implications." *Psychological Bulletin* 127, no. 4 (2001): 451–71.

Davis, David Brion. *The Problem of Slavery in the Age of Revolution, 1770–1823*. 2nd ed. Oxford University Press, 1999.

Drescher, Seymour. *Abolition: A History of Slavery and Antislavery*. Cambridge University Press, 2009.

———. *The Mighty Experiment: Free Labor versus Slavery in British Emancipation*. Oxford University Press, 2002.

Eriksen, Cecilie. "The Dynamics of Moral Revolutions: Prelude to Future Investigations and Interventions." *Ethical Theory and Moral Practice* 22, no. 3 (2019): 779–92.

---

to the History of Philosophy Colloquium at Utrecht University and to the participants at this colloquium for their engagement and their robust and incisive criticism, which were useful for the further development of the ideas in this paper. Thanks also to audiences at the IV GECOPOL Geneva Graduate Conference in Political Philosophy and the SOPHIA 2023 Conference. Thanks to the European Research Council (grant number 851043) for funding my research. Lastly, I would like to thank my two anonymous reviewers, the editorial team, and the copyeditor at the *Journal of Ethics and Social Philosophy*.

Gal, David, and Derek D. Rucker. "The Loss of Loss Aversion: Will It Loom Larger Than Its Gain?" *Journal of Consumer Psychology* 28, no. 3 (2018): 497–516.

Gaus, Gerald F. *On Philosophy, Politics, and Economics*. Thomson Wadsworth, 2007.

Guala, Francesco. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton University Press, 2016.

Haskell, Thomas L. "Capitalism and the Origins of the Humanitarian Sensibility: Part 1." *American Historical Review* 90, no. 2 (1985): 339–61.

———. "Convention and Hegemonic Interest in the Debate over Antislavery: A Reply to Davis and Ashworth." *American Historical Review* 92, no. 4 (1987): 829–78.

Heath, Joseph. "The Benefits of Cooperation." *Philosophy and Public Affairs* 34, no. 4 (2006): 313–51.

Henrich, Joseph. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press, 2016.

———. *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux, 2020.

Hermann, Julia. "The Dynamics of Moral Progress." *Ratio* 32, no. 4 (2019): 300–11.

Holslag, Jonathan. *A Political History of the World: Three Thousand Years of War and Peace*. Pelican, 2018.

Hopster, Jeroen, Chirag Arora, Charlie Blunden, Cecilie Eriksen, Lily Frank, Julia Hermann, Michael Klenk, Elizabeth O'Neill, and Stephen Steinert. "Pistols, Pills, Pork and Ploughs: The Structure of Technomoral Revolutions." *Inquiry* (2022): 1–33.

Huemer, Michael. "A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism." *Philosophical Studies* 173, no. 7 (2016): 1983–2010.

Jaeggi, Rahel. *Critique of Forms of Life*. Translated by Ciaran P. Cronin. Belknap Press, 2018.

James, C. L. R. *The Black Jacobins: Toussaint L'Ouverture and the San Domingo Revolution*. 2nd rev. ed. Vintage Books, 1989.

Jamieson, Dale. "Slavery, Carbon, and Moral Progress." *Ethical Theory and Moral Practice* 20, no. 1 (2017): 169–83.

Jost, John T. "A Quarter Century of System Justification Theory: Questions, Answers, Criticisms, and Societal Applications." *British Journal of Social Psychology* 58, no. 2 (2019): 263–314.

Jost, John T., Vivienne Badaan, Shahrzad Goudarzi, Mark Hoffarth, and Mao Mogami. "The Future of System Justification Theory." *British Journal of*

Social Psychology 58, no. 2 (2019): 382–92.
Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5, no. 1 (1991): 193–206.
Kahneman, Daniel, and Amos Tversky. "Choices, Values, and Frames." *American Psychologist* 39, no. 4 (1984): 341–50.
Kelly, Daniel, and Taylor Davis. "Social Norms and Human Normative Psychology." *Social Philosophy and Policy* 35, no. 1 (2018): 54–76.
Kitcher, Philip. *Moral Progress*. Edited by Jan-Christoph Heilinger. Oxford University Press, 2021.
Kogelmann, Brian. "What We Choose, What We Prefer." *Synthese* 195, no. 7 (2018): 3221–40.
Kumar, Victor, and Richmond Campbell. *A Better Ape: The Evolution of the Moral Mind and How It Made Us Human*. Oxford University Press, 2022.
Lewis, David. *Convention: A Philosophical Study*. Harvard University Press, 1969.
Moody-Adams, Michele M. *Fieldwork in Familiar Places: Morality, Culture, and Philosophy*. Harvard University Press, 1997.
Mrkva, Kellen, Eric J. Johnson, Simon Gächter, and Andreas Herrmann. "Moderating Loss Aversion: Loss Aversion Has Moderators, but Reports of Its Death Are Greatly Exaggerated." *Journal of Consumer Psychology* 30, no. 3 (2020): 407–28.
Müller, Julian F. "Large-Scale Social Experiments in Experimental Ethics." In *Experimental Ethics: Towards an Empirical Moral Philosophy*, edited by Christoph Luetge, Hannes Rusch, and Matthias Uhl. Palgrave Macmillan, 2014.
O'Connor, Cailin. "Measuring Conventionality." *Australasian Journal of Philosophy* 99, no. 3 (2021): 579–96.
———. *The Origins of Unfairness: Social Categories and Cultural Evolution*. Oxford University Press, 2019.
Pleasants, Nigel. "Moral Argument Is Not Enough: The Persistence of Slavery and the Emergence of Abolition." *Philosophical Topics* 38, no. 1 (2010): 159–80.
———. "The Structure of Moral Revolutions." *Social Theory and Practice* 44, no. 4 (2018): 567–92.
Popkin, Jeremy D. *A Concise History of the Haitian Revolution*. Blackwell, 2012.
Rehren, Paul, and Charlie Blunden. "Let's Not Get Ahead of Ourselves: We Have No Idea if Moral Reasoning Causes Moral Progress." *Philosophical Explorations* 27, no. 3 (2024): 351–69.
Robson, Gregory. "The Rationality of Political Experimentation." *Politics,*

Philosophy and Economics 20, no. 1 (2021): 67–98.
Ruggeri, Kai, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D. Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, et al. "Replicating Patterns of Prospect Theory for Decision Under Risk." *Nature Human Behaviour* 4, no. 6 (2020): 622–33.
Sauer, Hanno, Charlie Blunden, Cecilie Eriksen, and Paul Rehren. "Moral Progress: Recent Developments." *Philosophy Compass* 16, no. 10 (2021): e12769.
Simons, Mandy, and Kevin J. S. Zollman. "Natural Conventions and Indirect Speech Acts." *Philosophers Imprint* 19, no. 9 (2019): 1–26.
Singh, Manvir. "Subjective Selection and the Evolution of Complex Culture." *Evolutionary Anthropology* 31, no. 6 (2022): 266–80.
Singh, Manvir, Richard Wrangham, and Luke Glowacki. "Self-Interest and the Design of Rules." *Human Nature* 28, no. 4 (2017): 457–80.
Williams, Bernard. *Shame and Necessity*. University of California Press, 1993.
Yechiam, Eldad. "Acceptable Losses: The Debatable Origins of Loss Aversion." *Psychological Research* 83, no. 7 (2019): 1327–39.