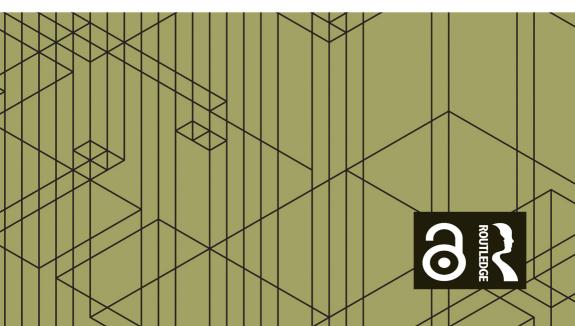# MORAL TELEOLOGY

A THEORY OF PROGRESS

Hanno Sauer

Moral Teleology

This book develops a unified theory of moral progress. The author argues that there are mechanisms in place that consistently drive societies towards moral improvement and that a sophisticated, naturalistically respectable form of teleology can be defended.

The book's main aim is to flesh out the process of moral progress in more detail, and to show how, when the right mechanisms and institutions of moral progress are matched together, they create pressure for the desired types of moral gains to manifest. The first part of the book deals with two issues: the conceptual one about what moral progress is, and the broadly empirical one whether it is possible. It shows that cultural evolution successfully explains the origins of modern forms of morally welcome change. The second part argues that there is logical space for a moderate, scientifically credible form of teleology, and that the converse case for moral decline is weak. It addresses the types, drivers, and institutions of moral progress that allow for the storage, transmission, and cumulative improvement of our normative infrastructure over time. Finally, the third part demonstrates why moral progress cannot be accounted for in metaethically realist terms.

Moral Teleology will be of interest to researchers and advanced students working in ethics, moral epistemology, and moral psychology.

Hanno Sauer is Associate Professor of Philosophy at Utrecht University, the Netherlands. He is the author of *Moral Judgments as Educated Intuitions* (2017), *Debunking Arguments in Ethics* (2018), and *Moral Thinking, Fast and Slow* (Routledge, 2018).

Routledge Studies in Ethics and Moral Theory

For more information about this series, please visit: www.routledge.com/
Routledge-Studies-in-Ethics-and-Moral-Theory/book-series/SE0423

Moral Teleology

A Theory of Progress

Hanno Sauer

Contents

Introduction

In his 1841 essay on *Compensation*, Ralph Waldo Emerson voices his frustration with the idea that there is no reward for being good—at least not for now. He opens with a story about a sermon he once attended, in which the preacher urged his congregation to be patient: for in this world, the "wicked are successful", and the "good are miserable". Emerson objects: "The fallacy lay in the immense concession that the bad are successful; that justice is not done now". Against this attitude of otherworldly resignation and sinister despair, he insists that the universe has its way of securing balance:

> Things refuse to be mismanaged for long. *Res nolunt diu male administrari*. Though no checks to a new evil appear, the checks exist, and will appear. If the government is cruel, the governor's life is not safe. If you tax too high, the revenue will yield nothing. If you make the criminal code sanguinary, juries will not convict. If the law is too mild, private vengeance comes in. If the government is a terrific democracy, the pressure is resisted by an over-charge of energy in the citizen, and life glows with a fiercer flame. The true life and satisfactions of man seem to elude the utmost rigors or felicities of condition and to establish themselves with great indifference under all varieties of circumstances.

Res nolunt diu male administrari: things refuse to be mismanaged for long. This book is an elaboration on this slogan. It aims to defend moral teleology, and to show that history is not morally indifferent: social change is biased in favor of moral improvement.

The story I have to offer is one of cautious optimism. While there are no *guarantees* that the arc of history bends towards justice, there are also no guarantees that the wicked will prevail. Indeed, there is reason to believe that there are mechanisms in place that more or less reliably push societies in an ameliorative direction. If we harness the power of those mechanisms, and stop sabotaging their force, a morally better world—and another one, and another one—is within reach.

While declinist anxieties are now as fashionable as ever—already 4,000 years ago, cranky Assyrians complained about "kids these days" and the impending social collapse ushered in by their unruly progeny (Protzko & Schooler, 2019)—I have little patience for pessimism, at least in its unadulterated form. In a contemporary review, Thomas Mann accused Oswald Spengler's *The Decline of the West* of a "malicious apodicticity and hostility towards the future, disguised as scientific mercilessness", and derided it for its "schoolmasterly lack of sympathy". I am inclined to agree; things refuse to be mismanaged for long, and even though there are (and always will be) episodes of decline and reaction, the overall trajectory of history points in a more agreeable direction.

This book will make the case for a *naturalistic moral teleology*. That social change is biased in favor of moral improvement is an idea virtually no one wants to be associated with. Most intellectuals, I suspect, would rather be unmasked as perverts than Panglossians. Meanwhile, the issue of moral progress is currently attracting increasing philosophical attention. In recent years, philosophers and social scientists have written about the nature of moral progress, the cognitive underpinnings of moral learning, the psychological obstacles to progressive change, or more specific moral transformations that have occurred over the past decades and centuries.

The book will intervene in those debates and develop a unified theory of moral progress. In order to do so, the following main issues will have to be addressed:

- What is moral progress?
- Is moral progress possible?
- What would it mean for moral progress to have a certain direction?
- What is the available empirical evidence for teleological moral progress?
- Which types of moral progress are there?
- How can moral progress be promoted?
- What causes progress to occur?
- How can moral progress be stabilized?
- Can there be moral progress without objective moral facts?

The upcoming chapters are supposed to answer these questions.

The main features of my account of moral progress are that it is:

- Non-inclusivist: the expanding circle of moral concern is not the main type of moral progress. It is merely one type of moral progress among many, and often merely a welcome by-product of moral fundamental moral gains.
- Externalist/Non-psychologistic: moral progress happens in supraindividual institutions, rather than individual minds.

- Anti-realist: for moral progress to be possible and/or intelligible, we do not need to assume the existence of mind-independent moral facts.
- Teleological: social change is headed towards moral improvement.

There is no unique end point of ultimate moral perfection that all societies are guaranteed to converge upon in virtue of a set of iron laws governing the trajectory of world history. Rather, the gist of the version of moral teleology I will develop and defend is that there are mechanisms in place that reliably make societies gravitate towards moral betterment. The core of this account is cultural evolutionary: the filters that regulate the transmission of our moral infrastructure from one generational learning cohort to the next are morally non-neutral. Human beings, in virtue of their cultural nature, have the ability to engage in progressive moral learning. Mechanisms of moral change bring about various types of moral progress which are then stored in external social institutions. Over time, a society's moral capital accumulates.

The drivers of moral improvement range from new moral insights over social movements, technological innovations, and cultural change to increases in group size. When the force of these mechanisms is channeled into the right social institutions, they engender the types of moral progress that we want to see happening. For instance, the circle of moral status is driven towards expansion when societies literally *grow*. Ultimately, more individuals are admitted to the benefits of cooperation on more equitable terms. Reduced discrimination and outgroup bias—the instances of moral progress we want to see—are the result. Or, to give another example, people increasingly experiment with new forms of living together, thereby demonstrating in practice the viability of alternatives to the system of unnecessary, oppressive, or harmful norms they have come to see as natural and inevitable. Or, social movements demand equality, and democratic empowerment ends up delivering the participatory rights whose time has come. Or, new scientific insights debunk commonly held views about alleged group differences, and obsolete prejudices wither away.

These remarks remain tentative. The main aim of the book will be to flesh out this process in more detail, and to show how, when the right mechanisms and institutions of moral progress are matched together, they create pressure for the desired types of moral gains to manifest.

At the very least, I wish to advance our understanding of *how* we should think about moral progress, when we do. Even if my teleological argument ultimately doesn't convince everyone, I will be satisfied if the key distinctions that I develop, and the logical landscape of our thinking about moral progress they illuminate, significantly improve our understanding of the tasks that any successful theory of moral progress must accomplish, and the nuances it must heed to do so successfully.

This book will reanimate the case for moral teleology. I will argue that there are good reasons for taking teleological accounts of moral progress much more seriously than we usually do. This aim is a modest one; to reanimate something means that it will be alive, not that it *does well*. I will explain what it means for there to be mechanisms in place that consistently drive societies towards moral improvement and show that a sophisticated, naturalistically respectable form of teleology can be defended. I call it *teleology 2.0*.

There are three parts and eight chapters in total. The first part (Moral Progress: The Fundamentals) paves the way for my defense of moral teleology. It deals with two basic issues: the conceptual one about what moral progress is, and the broadly empirical one about whether it is possible. The main result of the first chapter will be that we should rely on a wide notion of moral progress that includes both improvements in people's moral norms, values, and attitudes as well as morally desirable social improvements. Chapter 2 will show that our evolved psychology poses no meaningful constraints on the feasibility of moral progress. This is not, however, due to the fact that recent moral gains in favor of progressive morality are not amenable to evolutionary explanations. Rather, cultural evolution successfully explains the origins of modern forms of non-parochial prosociality and other forms of morally welcome change.

The second part (Teleology 2.0) forms the core of the book. In Chapter 3, I discuss what a teleological account of moral change would amount to, and speculate about why it encounters such strong resistance. I argue that there is logical space for a moderate, scientifically credible form of teleology that such a version of the teleological thesis—that social change is patterned by morally ameliorative biases—may well be defensible and sketch some reasons for why we should take the teleological option more seriously than it is commonly done. In Chapter 4, I argue that the converse case for moral decline is weak and that we have no reason to think that our assessments of moral progress are particularly likely to be epistemically unreliable. I discuss the best currently available empirical evidence showing that moral progress is actually occurring in the real world and develop the outlines of an explanation for why we are headed for further moral improvements in the future. The following three chapters are about the *types* (Chapter 5), *drivers* (Chapter 6), and *institutions* (Chapter 7) of moral progress. I will develop a systematic typology of various dimensions of moral progress, including some that currently tend to receive insufficient attention in the extant literature. I identify the main mechanisms engendering progressive social change, and argue that social institutions function as an extended moral mind that allows for the storage, transmission, and cumulative improvement of our normative infrastructure over time. When the right mechanisms and institutions are paired up, there will be pressure for the desired types of moral progress to occur. This means that moral progress has a weakly teleological direction.

Finally, the third part (The Metaethics of Moral Progress) addresses the issue of whether moral progress can only be accounted for in metaethically realist terms (Chapter 8). I will show why this is not the case. The main motivation for thinking that moral progress entails the existence of mind-independent moral facts is that such facts are required to explain moral convergence. Over time, individuals and cultures progress from an initial state of moral disagreement towards a shared set of values. This dynamic is best explained in terms of an increasing acknowledgment of objective moral truths. Or so it seems: I will reject this argument not because convergence can be better explained non-realistically, but because there is no initial state of moral disagreement for societies to gradually converge away from. There is strong empirical evidence for universal cross-cultural and diachronic moral agreement. There is no convergence for moral realism to explain.

Along the way, I will argue that in order for us to be able to talk about moral progress, it is not required that we have already reached the end of history, nor that there is one. For there to be moral progress, it is not required that we now know what the goals of human development are, nor that we ever will. Moral norms and values, on my view, are like headlights: you can only see so far, but if you rely on them, you can make the whole trip.[1]

Note

1. Quote adapted from E. L. Doctorow.

1 The Shape of Things to Come

What Is Moral Progress?

Introduction

What is moral progress? Over the course of this book, I will try to clarify what the teleological thesis—that social change is biased in favor of moral improvement—amounts to, provide an explanation of why it encounters so much resistance, offer some evidence for its truth, and sketch an explanation for why it may be true. In this first chapter, I want to take a step back and explain which concept of moral progress I will rely upon, and formulate some desiderata for a successful theory of moral progress. Along the way, I will introduce some of the major themes of this book, and explain some key distinctions that will matter for my discussion of the types, mechanisms, and institutions of moral progress in subsequent chapters. This is supposed to lay the groundwork for defending the idea that society gravitates towards moral improvement. My main task here won't be so much to provide answers to the various issues I address but to highlight various key distinctions which must be kept track of if one wants to provide such answers in the end.

There are ten sections. In section 1.1, I will briefly look into existing definitions of moral progress, and argue that the core concept of moral progress is actually quite simple and straightforward. I will nevertheless abstain from defining "moral progress", because definitions aren't very helpful for my purposes here. I will then introduce some basic distinctions between local and global (section 1.3), individual and social (section 1.4), moral and social progress (section 1.5) and discuss the extent to which for something to count as genuinely moral progress, and it must be something that was brought about by human agency (section 1.6). Finally, I discuss some apparent "paradoxes" of moral progress according to which moral progress often unfolds with a "one step backward, two steps forward" dynamic. This can be difficult to appreciate because the strictness of moral standards evolves more quickly than the pace of actual moral gains can keep up with, and many forms of progress involve at least some regressive aspects.

1.1   The Concept of Moral Progress

The point of conceptual analysis, as I take it, is to settle on a subject matter—nothing less, nothing more. Concepts are not natural kinds. They are tools, manufactured by us, to shed light on an issue of interest. I will not get caught up playing the counterexample game, which is boring (to me) and useless (to everyone else). Chiseling definitions is not what I am here for—I leave possible worlds to themselves, and focus on the actual world instead. The point is not to end up with a watertight definition of a term, uncovering and articulating its supposedly fixed inferential content. The point is to have a starting point of investigation and to say what that starting point is.

My approach is perhaps best described in terms of the distinction between characterizing something and defining it. When someone asks us what another person is like, we are being asked to characterize that person. Defining the person—say, "she is the one world-class tennis player who is also a gifted calligraphist"—is unhelpful. When we characterize something, we do try to pick it out, but by saying something about it that would be useful and/or interesting to know to the inquirer. This is what I intend to do here.

Ethica ordine geometrica demonstrata is not the way. Let's start from uncontroversial examples instead: the abolition of slavery and the emancipation of women first come to mind. (A quick caveat: phrases such as "the emancipation of women" seem to suggest that the development referred to here has been thorough and complete, as if all women everywhere had been fully emancipated and freed from discrimination, hardship, and vulnerability. It goes without saying that this is not so, and that talking about "gay rights" and the like is compatible with conceding that the implementation of civil rights for historically disadvantaged groups has been very much imperfect and is an ongoing pursuit.) The list of uncontroversial examples of moral progress also includes the disappearance and near total ostracism of tremendously harmful inanities such as foot-binding or dueling (Appiah, 2011).

Some examples, however, are somewhat more controversial. The recognition of animal rights, for instance, is not universally acknowledged as change for the better (though perhaps it should be). Biomedical enhancement is even more divisive (Persson & Savulescu, 2019), and is considered by many as an assault on human integrity rather than a promising fix for some of our more unwieldy dispositions.

These are some of the vivid ones. But we can also come up with more mundane examples of slightly more fine-grained social developments for the better. Change.org is a website specifically designed for rallying people around various progressive causes. A quick glance on a random day

informs me that people have posted petitions to ban fireworks on New Year's Eve, introduce car-free Sundays, tame political lobbyism, or reduce value-added taxes to 0% for second-hand shops. It is hard to identify anything that unifies these developments, besides the fact that they all aim to make the world a (morally) better place.

This, indeed, is the concept of moral progress suggested by Jamieson (2002), who holds that moral progress occurs when "a subsequent state of affairs is [morally] better than a preceding one, or when right acts become increasingly prevalent" (20). I believe that this "definition" of the concept of moral progress is essentially correct, even though it amounts to little more than what, in a different context, Kant once called the "explanation of its name" (KrV 52). It is nevertheless important to emphasize it, because many authors seem to conflate the *concept* of moral progress—things improve, morally speaking—with substantive *criteria* for moral progress.

In her landmark paper on the issue, for instance, Ruth Macklin (1977) claims that her "analysis" of the concept of moral progress yields the following two principles:

> (1) The principle of humaneness: One culture, society, or historical era exhibits a higher degree of moral progress than another if the first shows more sensitivity to (less tolerance of) the pain and suffering of human beings than does the second, as expressed in the laws, customs, institutions, and practices of the respective societies or eras. (2) The principle of humanity: One culture, society, or historical era exhibits a higher degree of moral progress than another if the first shows more recognition of the inherent dignity, the basic autonomy, or the intrinsic worth of human beings than does the second, as expressed in the laws, customs, institutions, and practices of the respective societies or eras.
>
> (371f.)

I emphatically agree that these things do indeed constitute moral progress. I do not agree, however, that to deny this betrays a *conceptually confused* mind. The reduction of gratuitous suffering and the recognition of human dignity are arguably two of the worthiest causes there are, but that this is so cannot be settled conceptually. Macklin's account of moral progress conflates substantive criteria of progress with what "moral progress" means.

Elsewhere, Jamieson (2017) toys with the thought that moral progress is really about the structure of reasons that people appreciate as morally relevant and adopt as motivationally effective: "One audacious idea is that moral progress consists in the increasing dominance of objective, impersonal, or agent-neutral reasons for action over subjective, personal, or agent-relative reasons" (174). It is not entirely clear what "dominance" means here, but the basic thought seems to be that forms of egoism and

tribalism are gradually replaced with more universal ethical outlooks. Again: I agree with this claim, but it is not helpful to disguise moral propositions that clearly have substantive normative content as conceptual truths. Normative reasoning, not conceptual analysis, is where claims regarding the universality of moral progress must be adjudicated.

We could also try to come up with what Evans (2017) refers to as a "working definition" of moral progress. Instances of moral progress are, he admits, easy enough to come by. But the underlying nature of these instances remains riddled with controversy. So why not make do with a "proxy property" for the time being? This will allow us to track occurrences of moral progress without having a rock-solid definition of the concept in our hands. Evans suggests population welfare as such a proxy, a proposal I am quite sympathetic to, and to which I will return later on. I prefer a good proxy definition over a bad intentionally adequate one; however, I don't see the need for a proxy in the first place. The disagreements Evans talks about again concern substantive questions regarding what moral progress consists in. They have nothing to do with what the phrase "moral progress" means. We know perfectly well what it means, namely that things improve, morally speaking.

1.2   But Is It Progress?

It may seem that the prime motivation for wanting a concept of moral progress, is to be able to identify a proposed change or an anticipated development as progress or regress in advance. Soon we may have self-driving cars, artificial wombs, or a social credit system. Soon, we may do away with our concepts of race and gender altogether. But are these things *really* progressive, or are they more properly described as things getting worse?

However, this question—but is X progress?—is itself regressive, and normatively stacks the deck against most proposed forms of social change, many of which have the *potential* to be progress, a potential we should often tap into. For one thing, even though it is correct that we never truly know beforehand whether something is progress, we also do not know that it is *not* progress. What, then, is gained by asking this question? For another thing, new technologies or ways of living together are actually biased in favor of being progressive, because of all the logically possible things humanity could be working on bringing about, we are only working on a tiny selection, namely the ones that probably have a real shot at turning out to be good ideas. It's not like doctors, biologists, and medical engineers are working on all sorts of new wombs—some artificial, some installed on the back, and some with automatic self-destruction after six months—such that we are actually worried about exactly which one of

these womb innovations on the horizon is the progressive one. Rather, there is a real problem here: pregnancy is cumbersome at best, dangerous at worst, and as far as burdens go, its distribution is quintessentially sexist. Artificial wombs may turn out to be a terrible idea, but this real problem is what they are responding to. Moreover, in asking "is it progress?", we are implicitly activating all kinds of unreliable biases. The question whether something is progress invites us to imagine all the ways in which things could go wrong, instead of all the ways in which they could go right. Humans are already quite bad at properly assessing the costs *and benefits* of any given development, and it doesn't help to double down on this recalcitrant tendency. Finally, and perhaps most importantly, the *But is this progress?* question implicitly suggests that once we go down a certain route, and decide to pursue a certain development, we will be stuck with it forever. That is quite evidently not true. Rather, we can experiment with new norms or technologies and, if they work to our satisfaction, scale them up later. We just don't need to know with certainty in advance whether something is progress.

How, then, should we characterize the concept of moral progress? A useful schema for understanding moral progress starts from its logical structure:

> A statement of moral progress is about a change toward a morally better state and typically has the following logical structure: *x* has made (moral) progress regarding *y* in relation to *z*, where *x* is the *subject* or *maker* of (moral) progress, *y* the *matter* of (moral) progress or what the progress consists in, and *z* the *dimension* of comparison, that is, the relata between which the comparison is made.
>
> (Egonsson, 2013)

Statements about moral progress presuppose something about the *thing* that undergoes the improvement, the *way* in which the respective improvement is made, and a *focal point* such as a particular group, or a time period, in which the improvement is said to have occurred. For instance, *over the past 50 years, US society has become less racist.*

Of course, when it comes to the dimension of progress, there is an ambiguity in that it is not always clear whether the progressive gains that have been achieved are due to new and improved values or whether what has improved is the level of compliance with preexisting values. That *US society has become less racist over the past 50 years* can mean that it shed or reduced its racist norms (or adopted new, non-racist ones), or that it started to adhere more consistently to its already existing anti-racist values. If Moody-Adams (1999) is right, as I suspect her to be, the latter case is far more common. The essence of moral progress rarely lies in

the discovery of novel principles and values (I will return to this point in greater detail in Chapter 8), and more often in acquiring a deeper understanding of the values we already regard as important. Perhaps more often still, moral stagnation simply comes down to a "dearth of incentives to scrutinize social practices" (183).

1.3   Local and Global

A first interim result is that moral progress occurs when things improve, morally speaking, where this requires us to specify a subject of progress, a reference point for comparison, and a dimension in which the progress is made, the latter being ambiguous between improvements in values and improvements in levels of compliance.

But we can say more. Many authors insist that judgments about progress can only be made locally (see, for instance, Moody-Adams, 1999, p. 169), but it is unclear whether this reluctance stems from genuine theoretical reasons or from a more general metaphysical anxiety over sweeping claims about where the history of humankind is headed. I don't share this anxiety. Global progress is local progress, everywhere.

The main advantage of judgments about local progress is that here the epistemic burdens to make good on any claims about progress are more easily shouldered:

> Judgments of local progress are facilitated because it is possible to focus on a single function, or concordant set of functions, and examine how completely (thoroughly, speedily, etc.) they are discharged. Assessments of global progress are difficult because multiple functions come into play.
>
> (Kitcher, 2011, p. 242)

In contrast, judgments of global progress would require us not just to compare two different levels of solving the same problem within one group but to make comparable, in terms of better and worse, mechanisms of solving entirely *distinct* problems in *separate* societies.

The distinction between local and global progress also becomes relevant for the issue of teleology I am after. Notice that, when it comes to the question of whether society more or less reliably gravitates towards moral improvement, a teleological account of social change can seem more, rather than less, plausible at the global level. There is simply no guarantee that this or that society will embark upon a trajectory of moral progress. On a global level, that is for humanity as a whole, the goal-directedness of moral progress may become more plausible, with various different groups spearheading the peloton at different times.

1.4   Individual and Collective

In talking about moral progress, we typically talk about what happens at the level of groups of different sizes and configurations. An issue that comes up less frequently is whether whatever gains have been made at the collective level also sediment into corresponding gains at the individual level: "Individual moral progress refers to the moral-psychological development of an individual; when an individual develops in a desirable direction in the moral domain (so undergoes durable positive change) this constitutes moral progress" (Schinkel & de Ruyter, 2017, p. 124).

When societies progress, must the individuals inhabiting it progress as well? Can collective moral progress be achieved even when most individuals stagnate, or indeed when the moral quality of their cognitive-motivational dispositions regresses? Can individuals—precociously, as it were—progress beyond what a collective has thus far achieved? Is collective moral progress driven by individual progress, or is individual moral development by collective ethical gains? Is there any such order of priority in the first place?

In one sense, the idea that there should be noticeable individual moral progress seems questionable. Consider the often vast transformations many societies have undergone over the past decades or centuries. How many of these, and to what degree, can plausibly be attributed to changes in individual people's psychology which likely remained unaltered over the course of such relatively recent and swift developments? At the same time, it seems that collective moral progress couldn't be durable if the involved individuals didn't follow suit at least to some extent. Social improvements must somehow be transmitted to the building blocks of society, that is, to individual minds.

1.5   Wide and Narrow

Perhaps most importantly for the purposes of my argument, there is a distinction between two senses of "moral progress":

(1) Narrow: There is progress in people's moral capacities or concepts. Here, morality *is the thing that undergoes the improvement*.
(2) Wide: There is some other form of social change or development which is *welcome* from the moral point of view, without morality itself necessarily being the thing that improves.

Changes in infant mortality, health, or happiness are *morally desirable social improvements*, but some hesitate to refer to them as improvements *of morality*. They are, it seems, moral improvements only in the following way: every sound normative theory (and morally competent mind) must

welcome and approve of these developments as morally desirable. They are not about moral progress *per se*.

Let's file this away (for now) as the distinction between *wide* and *narrow* progress. In this section, I will elaborate on this distinction in somewhat more detail and explain why I want to reject it, or at least deemphasize its importance. Over the course of this book, I will typically use the simpler term moral progress, by which I mean to use a notion that encompasses both sections 1.1 and 1.2, moral progress in the narrow sense *and* morally welcome social change in the wider sense.

What is the difference this distinction is trying to capture? Perhaps it can be explained in terms of the distinction between the wrong and the bad. Tsunamis are bad, but they are not *wrong*, presumably because they do not involve any human agency. High infant mortality is, provided that nothing can be done about it (which it often can), merely very *bad*, but not *morally* wrong. It makes little sense to declare tsunamis impermissible, however devastating they may be. Wide moral progress, then, reduces the bad; narrow moral progress mitigates the wrong.

There are other ways of drawing the distinction. Kitcher (2017), while decidedly anti-teleological in his ambitions, argues,

> [S]ocieties make progress when they change in ways that improve the lives of their members. When the sciences make socially embedded progress, they provide benefits for everyone in the circumambient society (sometimes as large as the whole of humanity.) Hence, there is a mode of social progress.
>
> (53)

Here we see again at the very least fuzzy boundaries between whatever social and moral progress are supposed to be. It is not obvious why it should not count as moral progress when significant benefits are provided to large numbers of people through deliberate cooperative efforts such as science and technological innovation.

A third way in which this distinction could be drawn is in terms of human competence. Must moral progress, in order to count as genuinely moral, be *brought about* by humans? Buchanan and Powell argue that it does. Considering the example of a hypothetical decline in disease rates, they hold:

> [H]ow such changes came about is arguably relevant to whether they are instances of moral progress properly described. Suppose that the great decline in the incidence of deadly infectious diseases had not come about, even in part, by deliberate efforts undertaken in the recognition that it is morally good or mandatory to reduce preventable human

> suffering and death. Suppose further, that this decline did not involve the exercise of any human motivational capacities, moral or otherwise. Suppose instead that the reduction occurred as a result of events utterly beyond human control—such as a naturally occurring environmental change that wiped out many infectious agents. Under these conditions, the reduction in the incidence of deadly infectious diseases would have undoubtedly been an improvement from a moral point of view, but it would be strange to call it an instance of moral progress.
>
> <div align="right">(Buchanan & Powell, 2018, p. 49)</div>

Suppose, they could argue, an item of great value simply fell from the sky, like mana, allowing a group of people to significantly improve their level of well-being. It would indeed be odd to classify this as moral progress.

Buchanan and Powell insist that an instance of progress should count as moral in a robust sense only if it is progress of the narrow kind. So, for instance, the eradication of a disease would not count as robust moral progress per se, only, perhaps, as morally welcome social progress. But this distinction, while being on to something real, is largely artificial. Indeed, Buchanan and Powell do not really explain why they consider this point important at all. This is remarkable, since their claim that only narrow moral progress is moral progress, period, plays little or no role in their pluralist account of moral progress. Methodologically speaking, they are not committed to denying that wide moral progress could constitute another form of moral progress next to inclusivist shifts or the elimination of surplus moral norms.

To appreciate the counterintuitive implications of their point, consider the fact that Buchanan and Powell even hesitate to classify declining rates of violence as morally progressive:

> For the great change that Elias and Pinker document appears to have occurred without improvements in or through the exercise of human moral capacities—that is, improvements in moral concepts, motivations, or virtues; in moral reasoning; in moral emotions; or in the ability to discern valid moral norms. Instead, it resulted from the introduction of institutionalized incentives that aligned self-interested action with valid moral norms—institutional changes that do not appear to have been morally motivated.
>
> <div align="right">(2018, p. 50)</div>

For one thing, when institutional changes make people reliably behave more virtuously, this seems to me like a *paradigmatic* instance of moral progress, even if those changes were not *only* morally motivated. For another thing, the claim that institutionally secured drops in violent crime such as murder

or rape were not at least partly morally motivated seems deeply implausible anyway. But suppose they were, would the conceptual point still stand that in that case, this trend should not be described as moral progress? Buchanan and Powell think so, but never really explain why:

> We think that the third, weakest understanding of moral progress [changes that are merely desirable from a moral point of view, H. S.] ought to be rejected because we believe it is important to distinguish between changes that are merely desirable from a moral point of view and changes that are morally progressive in some stronger sense.
>
> (51f.)

But we are given no reason for this, other than that they happen to find it important. Notice, moreover, that one can agree with their claim that one should hold on to this distinction if and because it highlights an interesting difference. This does not mean, however, that that distinction must consist in a difference between moral and *non-moral* progress. We can keep making the distinction, but *within* the concept of moral progress. My suggestion would be to mark this as the distinction between *wide* and *narrow* moral progress.

   Here are six reasons for why a wider notion of moral progress is methodologically preferable. One reason for doubting that this distinction can be sharply drawn as one between genuine moral and non-moral social progress is that in almost all cases in the real world, even improvements that appear to be merely wide "social" progress will be in some sense due to human achievement rather than happenstance. Even diseases and poverty don't usually reduce themselves. Even the death toll of natural disasters is for the most part dependent on human cooperative accomplishments. Consider deaths by lightning:

> And what about the very archetype of an act of God? The projectile that Zeus hurled down from Olympus? The standard idiom for an unpredictable date with death? The literal bolt from the blue? . . . [T]hanks for urbanization and to advances in weather prediction, safety, education, medical treatment, and electrical systems, there has been a *thirty-seven-fold* decline since the turn of the 20th century in the chance that an American will be killed by a bolt of lightning.
>
> (Pinker, 2018, p. 189, emphasis H. S.)

At first, it seems that for this to happen, no improvement in people's moral code or behavior was required. But that is not so: it took various forms of technological and institutional progress to lead to circumstances which could then, partly by design and partly as a byproduct, lead to a reduction

in deaths by lightning strikes. This is a form of (wide) moral progress—morally desirable changes, brought about directly or indirectly by improvements in human cooperation and knowledge.

Second, wide and narrow progress cannot be meaningfully disentangled in real life: moral progress without social progress is empty. We would think that improvements in people's morality—their moral codes or beliefs—without any tangible social pay-off would be hollow and worthless, even eerie and perverse. For instance, we would not recognize it as moral progress if everyone came to agree that slavery is abhorrent, with no discernible effect on the practice of slavery or the well-being and rights of the enslaved and their descendants. In fact, there would often be something particularly morally obscene about mere narrow moral progress without corresponding social gains.

Third, social improvements are almost always accompanied by, or indeed followed by, corresponding moral progress. This is an important causal claim. A decline in infant mortality brought about by medical advances will almost always bring with it a lower tolerance towards the death of children, such that a development like this improves our moral sensitivity as well as securing greater protection for the most vulnerable. This is moral progress: the last time I checked, not wanting to see children die premature and avoidable deaths was generally considered a virtue. Even if there were some natural, entirely non-moral process that brings about some benefit, there will still be corresponding moral ramifications. Say a slight tilt in the trajectory of the earth leads to fewer tornadoes, and thus fewer tornado deaths. This, it seems, has nothing to do with morality.[1] But still, this change will likely decrease our tolerance towards tornado deaths, so even in such conceptually fringy cases, wide and narrow moral progress are stably linked.

Fourth, "wide" social gains are almost always given direction and orientation by moral values: wide progress without narrow progress is blind. Moral values specify which social gains we want to realize, and how to prioritize them. Social and moral progress cannot be separated because one wouldn't occur, or not in this way, or not as soon, without the other. This is not to suggest that there may not be some rare and outlandish cases where a social gain can be achieved with almost no inherently moral contribution. Perhaps soon the torture and killing of billions of animals in factory farms will have ended because we have figured out how to artificially simulate the taste of meat in some other way (think again of so-called "clean" burgers). In such a case, an enormous social and moral gain—the end of animal suffering inflicted by factory farming—could be brought about by largely morally neutral motivations.

But despite the clear economic interests at stake here, the emergence of clean meat is undeniably morally motivated as well. Still, it does not seem implausible to suggest that many people would happily go back to killing

animals as they used to had lab-grown meat not been invented. However, even in such special and marginal cases, we find that (a) the invention of the technology was probably at least partly, and often largely, driven by the moral desire of some people to reduce animal suffering, (b) the *de facto* end of animal suffering will over time likely lead to greatly reduced tolerance towards animal suffering, thereby further entrenching the moral gain and making it less reversible, and (c) all of this led to the extremely morally desirable development of ending factory farming. So even such exceptional cases of an almost complete separation between wide and narrow moral progress, it seems to me, deserve to be described as genuine moral progress.

Medical advances that reduce disease rates don't make and distribute themselves: they are almost always themselves moral achievements made possible by improved human cooperation in the development of technology, the sharing of knowledge, and the distribution of benefits. One of the most important progressive gains comes from the reduction of global poverty over the past decades (Rosling, 2018). However, people generally weren't bothered all that much by the phenomenon of poverty—that is, until it started to decline. Since then, measurable indicators of "poverty awareness" (such as prevalence of the word "poverty" in written records) have been on the rise.[2]

Fifth, the claim that social progress and moral progress are in an important sense distinct is hard to maintain in light of the best available examples of progress. The evidence suggests that more and more people are lifted out of extreme poverty; that equality rates are up, and that racism and discrimination are down; it suggests that democracy is on the rise, and that international peace and cooperation have been improving (despite recent setbacks). If these things shouldn't count as moral progress, I don't know what should.

Finally, a narrow concept of moral progress of the kind envisioned by Buchanan and Powell makes it at least somewhat harder to properly identify cases of moral *regress*. For instance, those who want to resist overly optimistic narratives of continuous socio-moral improvement frequently insist on emphasizing the offsetting costs of other developments that allegedly undermine the "Panglossian" narrative. The most frequently invoked example here is climate change: progress *on net* is hard to defend if we all get roasted or drown.

But notice that climate change is precisely the sort of trend that, on the narrow conception of moral progress, we couldn't recognize as genuine moral regress anymore. Climate change largely concerns unwanted side-effects of human action. For a chain of events to count as genuine progress, Buchanan and Powell require that it be instigated by the "exercise of or improvement in human moral capacities" (2018, p. 52). But climate change's dire effects are not, like some heinous calamity orchestrated by a

James Bond-like villain, due to the "exercise of or deterioration in human moral capacities". Climate change is the unanticipated consequence of otherwise perfectly morally innocuous human activity, like driving to work, or owning a refrigerator, or operating a factory. The general lesson here is that in coming up with a concept of moral progress, one should bear in mind that a narrower conception of the good cases yields a narrower conception of the bad cases as well. The criteria for what counts as progress and regress ought to be symmetrical.

1.6   Moral Regress

What makes people most uncomfortable with all this talk about progress and social change for the better is the suspicion that there is a sleight of hand somewhere, that one is being tricked into believing that everything is fine. Strictly speaking, this fear is unwarranted: saying that things are better now than they used to be in no way entails that everything is perfect now, or that there isn't any more room for more improvement. Psychologically (rather than logically) speaking, however, the worry makes sense. Narratives of progress can crowd out a proper sensitivity for whatever is lost, or not yet gained. Progress in a global sense tallies *net* improvements, but net gains refer to the balance of gains over losses. Those losses constitute pockets of regress that a theory of moral progress must find a place for.

Moral regress is a possibility that pluralistic theories of progress in particular must reckon with (Dixon, 2005). Only those who support a monist metric of social improvements are at least somewhat safe from problematically overlooking parallel moral regress, because if there is only one normative standard by which to assess ethical gains—increases in aggregate welfare, say—then either there has been progress on that one metric or there hasn't been. With multiple measures of progress, the situation is more complicated: perhaps a society has managed to improve along one axis, but not without deteriorating along another.

Let's not ignore regress, then. For instance, if moral progress has led to increases in well-being which are in large part due to enhanced cooperation among a greater number of people and the unleashing of unprecedented productive forces required to sustain such growth, climate change—and our apparent inability to handle it at scale—becomes an obvious candidate for corresponding regress. The erosion of local community ties will be a problem for modern societies whose members are increasingly mobile, experimental, and unfettered by regional customs and traditions (Putnam, 2000). The flourishing cultural diversity supplied by regional idiosyncrasies in clothing, food, architecture, or temperament may start to wither away as well.

(Of course, such local ties often have a darker, oppressive side as well; ascriptive community membership is being replaced with other forms of

community, which are less local, but need not be any less meaningful. And if Joseph Henrich [2020] is right, the loss of family community ties and the move towards smaller families and greater individualization produced enormous benefits on the whole.) Extreme poverty may be down, but the absolute poorest of the poor often remain shut out of those progressive gains. Finally, human–animal relationships are perhaps the most striking example for a discrepancy between improvements in people's moral attitudes on the one hand, and our actual treatment of animals on the other. The idea that animals could deserve moral consideration would have baffled people at almost all times in the past and in almost all places even today. But even though nowadays, millions of people take the possibility of animal rights seriously and have adopted a vegetarian or vegan diet as a consequence, society as a whole subjects pigs, chickens, or cows to unspeakable horrors on an unfathomable scale (Huemer, 2019).

1.7   One Step Back, Two Steps Forward

In many cases, moral progress has a "one step backward, two steps forward" trajectory. It would be immense moral progress for factory farming to end. But the only reason it can be ended at all is because it was introduced in the first place. This shows that even when unambiguous moral progress occurs between t_1 and t_2, how much "net" progress has been made between t_0 and t_2 can nevertheless remain an open question—the gains achieved by any given episode of progress may be eaten up by the fact that it was just about reversing earlier regress.

The emergence of large, peaceful, and (aspirationally) egalitarian modern societies came with a steep price. This price fell particularly unevenly on the early generations of people who participated in the emergence of statehood in the first place. We must imagine the hunter-gatherer happy: a merry band of quite healthy, quite cheerful, quite egalitarian people who—if, indeed, they made it past infancy—could enjoy a relatively long life with surprisingly little toil, surprisingly much leisure, and a surprisingly strong degree of control over how their lives go (Widerquist & McCall, 2015). For the vast majority of people, then, the emergence of states was catastrophic. As the size of groups and their capacity for warfare increased, individual people became increasingly more miserable; most had to work long hours under excruciating conditions, forcibly kicking up the lion's share of their harvest to a tiny ruling class and their religious jingoists.

Was it worth it? It remains hard to tell. The emergence of states was almost always accompanied by coercive labor and a deterioration of quality of life for everyone but the ruling caste. (The moral history of the Neolithic revolution has a strong libertarian favor.) It seems difficult to imagine, however, how human society could ever have engaged in any

significant development at all without the gruesome and exploitative path through early statehood.

Why on earth did people join those states at all? In many cases, the answer is: brute force. Primitive forms of sedentary living and the inception of agriculture made the formation of early states possible. Jared Diamond (1987) famously called this transition "the worst mistake in the history of the human race". Financing the whole thing was made possible through coercive taxation. Accordingly, early states ran on difficult to hide but easy to count grain (Scott, 2017). The resulting bottleneck—people who steadfastly refused to join this new arrangement were often exterminated—likely accelerated humanity's journey towards self-domestication (Hare, 2017). Only the prudent, servile, and industrious could just about tolerate their new fate of schlepping giant boulders around to build oversized sepultural memorials for narcissistic and delusional tyrants obsessed with their afterlife. It seems that only recently, some societies have managed to catch up with pre-Neolithic levels of well-being.

1.8   Regress for All!

Should a theory of moral progress permit certain genuine normative paradoxes? In the previous two sections, I have discussed cases in which progressive change is accompanied by parallel regress as well as cases in which a progressive gain does not count as net progress because it has merely reversed earlier deterioration.

It may also be possible for some social changes to count as both progressive *and regressive* at the same time. Easterbrook (2018, p. 155ff.) notes that technological progress in weapons systems made warfare more precise. This means, among other things, that weapons such as bombs and missiles became safer in virtue of becoming even deadlier—an ambivalent outcome if there ever was one. In other cases, access to institutions can become more inclusive (progress), even though the spread of the institution itself may further entrench objectionable values and practices. For instance, one may believe that traditional marriage is, on the one hand, a regressive institution, yet also welcome the fact that blatantly discriminatory forms of exclusion from this institution for gay couples are gradually being removed (Chambers, 2017).

Perhaps most paradoxically, some argue that morality itself is an obstacle to moral progress (Kovacheff et al., 2018). Moral judgments can help regulate anti-social or harmful behavior, but it can also amplify polarization along the lines of socially constructed group membership, undermining civil political discourse; it can make people resist seemingly morally discordant scientific facts, and it can stifle the acceptance of technological improvements, such as vaccinations or nuclear power, whose spread would increase general welfare.

1.9   Imperfect Allies

In thinking about the mechanisms of moral progress, we should be prepared to find that the mechanisms driving the social changes we deem noblest may themselves not be especially pretty. Moral progress can be brought about by morally neutral or even questionable processes. From the perspective of animal rights, the reduction of animal suffering is one of the paramount goals. Factory farming and the meat industry are arguably the leading causes of anthropogenic animal suffering, and most animal rights advocates would thus like to see them end. To the extent that artificial, lab-grown meat contributes to this goal, activists should welcome the proliferation of clean meat as something extremely morally desirable.

But there is nothing inherently morally worthy about this proliferation: it is driven by the very same, presumably not very ethically impressive, desire for the taste of animal flesh that motivated the spread of the meat industry in the first place; and it is created by the very same financial interests that incentivized the meat industry to cater to this taste at all. People see a tangy burger; the clean meat industry sees dollar bills. There is nothing morally noble about this neutralization of our baser instincts, and yet such morally extraneous developments can be a source of tremendous moral progress.

A similar dynamic can be seen at play in Appiah's (2011) account of how moral revolutions happen. The practices he focuses on—dueling, foot-binding, or slavery—are arguably among the most pernicious and harmful customs human societies have ever come up with. Appiah claims that a sense of "honor" often helped bring about the sometimes incredibly swift unraveling of the corresponding norms and values. However, not only is it difficult to establish any real causal links between the moral revolutions and the upsets of honor he identifies. People living at the time of the revolutions he mentions sometimes expressed their dissatisfaction with the status quo in terms of honor, but this doesn't suffice to show that the vocabulary of honor played a huge role in causally bringing those revolutions to fruition. Second, in most cases, what Appiah ends up describing is that a certain group of people—usually a select privileged few, as foot-binding, slavery, or duels were inevitably "privileges" of the wealthy—abandoned a practice not primarily because the practice was deemed dishonorable, but when and because it was made available to the unwashed *hoi polloi*, segregation from which had to be vigorously maintained. Honor revolutions are rarely about morally noble forms of honor and dignity, and almost always about defending class distinctions. This does not mean, however, that those revolutions weren't desirable, or don't deserve to be classified as moral progress. Progressives should not be prissy about the often dubious causes of social change: if we want to make the world a better place, we will have to work with some imperfect allies.

1.10   The Princess and the Pea

What further complicates the picture is that moral progress has an at least twofold structure. Progress does not merely bring about changes for the better. It also sharpens the tools by which we assess whether things have improved or not. The resulting problem is that these two developments can accelerate at a different pace. If things get better, but the standards by which we judge such improvements become keener even more quickly, it can seem that we are on a path to regress even though we actually aren't. You don't get uglier just because you got a new camera with a higher resolution.

Recent empirical evidence suggests that this "princess and the pea"—character of moral progress is a real phenomenon. Levari et al. (2018) dub it "prevalence-induced concept change": perception is famously sensitive to contextual factors, such as when shades of grey appear brighter or darker depending on their surrounding colors. It seems that this can happen with more complex notions as well: in a series of studies, Levari et al. examined how the extension of (proto)normative concepts such as "threatening" or "unethical" can widen as the prevalence of threatening or unethical items decreases. When subjects were supposed to identity faces or acts in terms of the respective concepts in a series of 1,000 trials, they started applying these concepts more often (in the final 200 trials) as their prevalence went down over the whole series (see also Haslam, 2016).

In one sense, of course, this is a very good thing. Other things being equal, our moral standards are *supposed to* become stricter, and our tolerance of wrongful behavior is supposed to go down. The problem, if there is one, is that stricter standards can blind us to actual improvements. As long as our norms become less forgiving while maintaining a healthy appreciation of the gains that have in fact been made, nothing is amiss. (Moreover, I wish to emphasize that in the aforementioned study, concept creep concerned only the grey area. At the extremes, where items were either obviously unethical/threatening or undeniably ethical/non-threatening, little or no change occurred.)

Moral progress can consist in objective improvements to society, or in rising standards of moral optimality, or both. But there is no guarantee that the two will evolve in tandem, and there is no mechanism to keep them in perfect equilibrium. So in some cases, it can happen that there are objective moral improvements that aren't identified as such because people's standards of moral evaluation haven't kept up. Or, and perhaps more frequently, the evolution of moral standards and their stringency can accelerate without the social developments being able to keep the same pace. In such cases, there may be actual moral improvements that aren't properly recognized because the mercilessness of our increasingly surgical moral standards finds them wanting.

Notes

1. Thanks to Victor Kumar for helpful discussions on this point.
2. https://voxeu.org/article/poverty-enlightenment-awareness-poverty-over-three-centuries?fbclid=IwAR0d8_eC586C1GZ5dhUhPHbdqXNYFgMbRpD5Fqkyo Ol3Hmonvsn_4wQReaA

2 Butchering Benevolence
Is Moral Progress Possible?

Introduction

My aim in this book is to explore the prospects of moral teleology. So far, I have addressed some of the more conceptual issues surrounding the notion of moral progress I want to rely on. Ultimately, I will focus on how to understand the teleological thesis, and why one should take it more seriously than commonly assumed. In subsequent chapters, I will explain which types of moral progress can be distinguished, which mechanisms are driving it, and which institutions people come up with to store morally progressive gains. Finally, I will address the metaethical question of whether progress must be progress towards an objective moral truth. But before I get into all that, let me pause for a moment and consider whether moral progress is *possible* in the first place, or whether it faces obstacles we simply don't know how to overcome.

When it comes to furnishing an explanation of the building blocks of our moral psychology, evolutionary theory holds a lot of promise. At the same time, there is an inherent tension between an evolutionary perspective, which looks for the distal causes of deeply entrenched features of cognition and behavior, and the possibility of moral progress, which concerns how human thinking and action can be changed for the better. Evolution is about how the way we used to be informs how we are *today*; progress is about how we can be *tomorrow*. The Pleistocene environment we were selected for and the environment we inhabit today are utterly different from each other. The environment we will inhabit, if advocates of moral progress get their way, will be much more different still. For evolved beings like us, genuine moral progress may be out of reach.

Allen Buchanan and Rachell Powell (2018) claim, first, that standard evolutionary explanations cannot account for inclusivist shifts that expand the circle of moral concern beyond strategically relevant cooperators. This is supposed to show that *evolutionary conservatism*—the view that our inherited psychology imposes significant feasibility constraints on how

much inclusivist moral progress can be achieved—is unjustified. Second, they hold that inclusivist gains can be sustained, and exclusivist tendencies curbed, under certain favorable socio-economic conditions.

In this chapter, I am interested in such feasibility challenges to moral progress (Hermann, 2017). I will argue that Buchanan and Powell concede too much to the evolutionary conservative, because their second point shows that evolutionary conservatives are right about the first: inclusivist shifts are unrealistic where moral progress matters most, namely under harsh social, political, and economic conditions. I suggest two ways around this problem. One is to focus on different *types* of moral progress that are underwritten by psychological capacities which are not subject to comparable evolutionary constraints. The other is to look beyond possible extensions of our *psychological* capacities, but for *institutional* support that renders them irrelevant. We need to bypass, rather than further stretch, the constraints of our evolved psychology. These two strategies, I argue, complement each other, and together offer a more promising response to the challenge to moral progress posed by evolutionary conservatism than the one suggested by Buchanan and Powell.

There are essentially three main ways of countering the evolutionary challenge to the feasibility of moral progress. First, one could grant that our psychology did inherit significant constraints, but reject the conservative implications of this claim. Second, one could grant the basic evolutionary story regarding our psychological limits, but start looking for other capacities which can secure the desired moral gains without being subject to the same evolved constraints. Third, one can try to reject the very idea that a selectionist account of our moral psychology succeeds in establishing significant limits on our moral attitudes at all. I will explore all three options here and see which holds the most promise.

This chapter has five sections. In section 2.1, I briefly outline the basic thrust of the challenge from evolutionary conservatism. In section 2.2, I offer a diagnosis of the central problem with Buchanan and Powell's progressivist response. Section 2.3 develops my first objection, according to which a more promising reply to the evolutionary conservative draws on moral attitudes that can secure the desired inclusivist gains without being subject to comparable psychological constraints. In section 2.4, I sketch an institutionalist solution to the overall problem that is based on the idea that in many cases, smart institutional kludges allow us to economize on moral motivation in a way that bypasses issues of psychological feasibility altogether. Section 2.5 situates the argument developed here in a wider context of cumulative moral learning and its evolutionary role. Finally, I argue that even if there were significant evolved constraints on our psychology, moral progress would remain feasible (5).

2.1 The Limits of Concern

In his *Practical Ethics*, Peter Singer writes:

> Those who put forward the first version of the objection often make observations about human nature. They point out that we all are much more concerned about our own interests, and those of our immediate family, than we are about the interests of strangers. That is, they may add, because we have evolved from a natural process in which those with a high degree of concern for their own interests, or the interests of their offspring and kin, tended to leave more descendants in future generations than those who were not so concerned with their own interests or those of their kin.
>
> (Singer, 2011a, p. 211)

Adam Smith also famously noted the limitations of our capacity for empathy and compassion:

> Let us suppose that the great empire of China, with all its myriads of inhabitants, was suddenly swallowed up by an earthquake, and let us consider how a man of humanity in Europe, who had no sort of connection with that part of the world, would be affected upon receiving intelligence of this dreadful calamity. He would, I imagine, first of all, express very strongly his sorrow for the misfortune of that unhappy people, he would make many melancholy reflections upon the precariousness of human life, and the vanity of all the labours of man, which could thus be annihilated in a moment. He would too, perhaps, if he was a man of speculation, enter into many reasonings concerning the effects which this disaster might produce upon the commerce of Europe, and the trade and business of the world in general. And when all this fine philosophy was over, when all these humane sentiments had been once fairly expressed, he would pursue his business or his pleasure, take his repose or his diversion, with the same ease and tranquillity, as if no such accident had happened. The most frivolous disaster which could befall himself would occasion a more real disturbance. If he was to lose his little finger to-morrow, he would not sleep to-night; but, provided he never saw them, he will snore with the most profound security over the ruin of a hundred millions of his brethren, and the destruction of that immense multitude seems plainly an object less interesting to him, than this paltry misfortune of his own.[1]
>
> (Smith, 2004 [1759], p. 157)

That there are limits to our moral concern can be empirically corroborated. Recent studies suggest that we suffer from steeply declining marginal empathy (Västfjäll et al., 2014, 2015). We can muster concern for one

(identifiable) individual. Two—not so much. (This is called "compassion fade".) This is exacerbated by the fact that the impossibility of helping everyone makes us less likely to believe that we can help *anyone*. (This phenomenon is called "pseudoinefficacy".)

It is worth mentioning here that, even if this were true, which features the ingroup/outgroup distinction latches onto is largely arbitrary (Heath, 2014a). There is no reason why our exclusivist tendencies cannot be channeled towards morally arbitrary but *harmless* features. The boundaries of group membership could be drawn, for instance, not on the basis of race or ethnicity, but in terms of which inherently arbitrary group one identifies with, such that no morally and politically relevant forms of discrimination and disadvantage remain attached to these group divisions. Some people may support Manchester United, some may root for Liverpool. But it isn't particularly likely that these forms of group membership are tied to significant differences in wealth or social status. If we also grant, for the sake of the argument, that it is not ideal when social institutions encounter individual minds whose psychology is rigged against them, this gives us a powerful argument for thinking that certain proposed progressive goals ought to be abandoned, or at least reconsidered (FitzPatrick, 2019).

2.2   From Evolution to Conservatism

Evolutionary conservatism is the view that certain progressive social developments are at odds with our evolutionarily inherited psychology. This fact, in turn, is supposed to have non-trivial morally and politically conservative implications.

According to the dominant account of moral progress, ethical improvements consist in an "expanding circle" (Singer, 2011b) of moral concern. Buchanan and Powell describe this as a series of "inclusivist shifts" (2018, p. 55). Not long ago, (full) membership in the moral community was restricted to an astonishingly narrow group of people. Over time, the moral franchise was extended from, say, wealthy men of the respectively dominant ethnic group to a wider selection of people. Sex and gender, race and ethnicity, wealth and education, religion and nationality, health and even species membership came to be recognized as essentially morally arbitrary distinctions that carry no independent moral weight. This process has frequently been imperfect, or hasn't happened at all (yet?) in many places. But across a wide variety of social contexts, the expanding circle of inclusivism has been an important aspect of the trajectory of moral development.

Evolutionary conservatives think that these shifts, however desirable they may seem in theory, are simply not feasible. More precisely, they argue that at the very least, this dynamic of inclusion cannot be continued indefinitely, and will ultimately run up against decisive psychological

obstacles that we owe to our ancestors' evolutionary trajectory. In its stronger forms, evolutionary conservatism is the view that we have recently reached the limits of progressivism or, stronger still, have already begun to *overstretch* the carrying capacity of our moral psychology, such that exclusivist relapses may be impending.

Their reasoning goes, roughly, as follows: in order to be sustainable, progressive social developments crucially depend on a certain psychological environment. For racism and speciesism to become a thing of the past, people need to be able to have moral concern, and extend moral consideration, to beings beyond their nearest and dearest, kin and kith. Otherwise, the aforementioned progressive developments will remain unstable.

The problem, as conservatives are quick to point out, is that our ability to care about others at all has an evolutionary rationale. In small, closely genetically related tribal groups, certain cooperative dispositions such as kin altruism or reciprocal altruism can be adaptive. Internally cooperative groups will enjoy a selective edge in intergroup competition for scarce resources in the environment of evolutionary adaptedness. Individuals who are disposed to care for their offspring and engage in reciprocal chains of sharing and helping increase the comparative frequency of their genes in the next generation. Unfortunately, this also means that our capacity to care for others is, and must be, limited. Natural selection operates on genes, and will tend to favor cooperative strategies that promote the proliferation of copies of alleles. This seems difficult to accomplish through cooperation that goes beyond relatives or reciprocators (Bowles & Gintis, 2011).

Evolutionary pressures have equipped our minds with a recalcitrant tendency to carve up the world in terms of an in- and an out-group. Benevolence exists, but *universal* benevolence is evolutionarily unstable. The very dispositions that make us cooperative also make us tribalistic. Buchanan and Powell do not deny this. Since progressive moral developments require psychological support, evolutionary conservatives claim that this imposes substantial limits to how much inclusivist progress can be achieved. Progressives may fantasize about a cosmopolitan *kumbaya* world with open borders and full-blown animal rights. But a world without *us* and *them*, conservatives suggest, is problematically utopian. It is an ideal that should be given up.

The biological evolutionary challenge isn't the only feasibility challenge to moral progress. Cultural evolution may pose its own obstacles to the moral development of society. Here, the challenge would be based on the idea that because human beings are deeply cultural institutions, there will often be no reason to think that we understand the workings of culture well enough to be competent social engineers:

> We cannot help but wonder what sustained attention to the details of this [of cultural evolution, H. S.] might reveal that could usefully inform

philosophers' understanding of moral progress and moral decay. For example, some initial work that takes this tack raises questions about what the kind of cultural evolutionary perspective defended by Henrich implies for traditional conservatism and the feasibility of intentionally guided moral progress. . . . Given how complicated, numerous, and intricately intertwined all of the different components are, attempts to initiate changes to the social order or the moral codes that govern it, the conservative argument goes, are more likely to go awry in unforeseen ways than produce improvement or genuine progress. Henrich's emphasis on the ability of the process of cultural evolution to design products, including complex social institutions, whose virtues and functions we individually and collectively often fail to completely understand appears to give strength to the conservative view.

(Kelly & Hoburg, 2017, p. 837f.)

This is an essentially Burkean challenge according to which we do not know with enough depth and accuracy how and why institutions work, which makes radical overhaul a risky endeavor. As with biological evolution, the conservative challenge from cultural evolution has a kernel of truth to it: there may be no limits to moral progress per se, but limits to how swift and radical progressive change should occur. It should be noted that we are probably biased in favor for the status quo, underestimate the potential benefits of change, overemphasizing their potential costs, and generally think that social change occurs too rapidly even when it doesn't. Also, the challenge from cultural evolution is actually not an argument in favor of conservatism at all. If anything, it is an argument for moral progress and reform, combined with a note of caution that competent social design is very difficult to pull off, and should probably be done in a piecemeal and step-by-step fashion.

It is true that culture is too intricate to understand, much less design, well (Kling, 2016). Should we therefore leave it alone to let it sort things out by itself? Not quite: the point is well taken, but it does not necessarily license conservative conclusions in any normatively significant sense of "conservative": first, conservatives want to deliberately design the trajectory of culture as well, typically by recommending a return to the past or a preservation of the present. This is precisely the kind of radical social engineering that conservatism itself would recommend *against*. A return to the past may seem like a return to the tried-and-true that isn't liable to the same kind of skepticism about social engineering. But this is not the case: returning to something after a period of time which created entirely different circumstances does not amount to a return to the same thing. One cannot step into the same culture twice. Moreover, the past that conservatives imagine is largely fictional, or at the very least shot through with nostalgia.

Acknowledging the complexity of cultural evolution thus gives the conservative no distinct advantage. Rather, the progressive recommendation is to not smother the engines of cultural innovation and to let social evolution run its course, rather than to freeze it in a homely status quo. Burkean premises do not mandate politically conservative conclusions.

2.3   A Conservative Advantage?

What is Buchanan and Powell's solution to the conservative feasibility challenge? They argue that the reality of the "inclusivist anomaly" shows that our psychology is not, or at least not thoroughly, rigged against progressive shifts. Modern forms of "subject-centered" (Buchanan & Powell, 2018, p. 48) morality show that our morality can evolve against the biological grain, since it grants moral status on the basis of the mere fact that someone is a certain kind of entity such as a person or a sentient being, rather than being a strategically relevant (potential) cooperator. The latter feature threatens to exclude infants, the severely disabled, or future people, and may exclude most (if not all) non-human animals (Buchanan & Powell, 2018, p. 57f.).

I will argue that the claim that such moral outlooks are not amenable to evolutionary explanation is misleading, or at any rate unnecessary. We do not need our onboard psychological resources to support progressive change. We need cleverly designed institutions that harness the power of our evolved dispositions for counterintuitive cooperative arrangements.

Buchanan and Powell do not just argue that there is a certain fact—the existence of inclusivist moral developments that lead people to recognize the moral status of strategically irrelevant entities such as animals or the severely disabled—that contradicts evolutionary conservatism. They also develop a positive account of the conditions under which inclusivist moral progress is possible. Here, their core claim is that exclusivist tendencies are *adaptively plastic traits*. Under favorable conditions, these traits need not be expressed. It is only when conditions are harsh, or when enough people come to believe that they are harsh, that exclusivism will manifest, creating a hostile environment for a more generously drawn moral circle. (Actually it seems that the second condition is sufficient, because when conditions are objectively harsh but people somehow don't believe that they are, exclusivism may not manifest either. It's about what people *believe* to be the case.)

However, remember that evolutionary conservatives argue that inclusivist shifts are psychologically infeasible or at least so psychologically unrealistic that they threaten the sustainability of progressive change. And note that achieving progressive shifts in the direction of extended cooperation and less uncooperative behavior (such as violent conflict between groups) is by far most urgent in societies that suffer from unstable cooperation and

lots of violent conflict due to the morally arbitrary constrictions of the moral circle prevalent in them. These two claims together entail that even on Buchanan and Powell's optimistic story, evolutionary conservatives are proven right where it matters most: places ridden with conflict and poverty are also most likely to struggle with exclusivist tendencies towards minorities, from discrimination to genocide. These places are thus where the need for inclusivist shifts is greatest. But due to socially, economically, and politically harsh conditions, inclusivist moral progress *is* psychologically infeasible in those places, which is to say where it matters most. For most intents and purposes, evolutionary conservatism is correct.

Consider, for the purpose of illustration, that dense urban environments can foster liberal attitudes (Campbell, 2017). Evolutionary conservatism thus seems strictly speaking false. Stable progressive societies are possible. The problem is that in order to get to those dense urban environments in the first place, many of the most pressing problems of intertribal conflict must already be sorted out. Large, prosperous cities are *facilitated* by liberal attitudes as much as they promote them. If Buchanan and Powell's admission that inclusivist attitudes are an evolutionary "luxury good" (Buchanan & Powell, 2018, p. 188) is correct, evolutionary conservatives are proven right about the fact that expansions of the moral circle are psychologically unrealistic in the very contexts where they are most needed.

Let me emphasize that the selectionist explanation of altruistic attitudes sketched earlier does not mandate conservative conclusions. Joshua Greene (2013), for instance, argues that we evolved a certain type of cognitive machinery to deal with an evolutionary "tragedy of the commons" and the fact that free-riding and selfishness remain individually rational for everyone, thereby undermining cooperative relations for mutual gain. But the solution that was selected for then ends up trading one problem— the tragedy of the commons—for another, the "tragedy of commonsense morality" (Greene, 2013, p. 1ff.). This second problem is a direct result of the first: for morality to evolve at all, it had to remain restricted to group members, thereby replacing the "me vs. us" structure of the former tragedy with the "us vs. them" structure of the latter. At first glance, this looks a lot like the evolutionary conservative's story as reconstructed by Buchanan and Powell: our evolution has locked our minds in the trappings of ingroup/outgroup thinking, and there is nothing we can do about it. Given that there are generally good reasons to refrain from pursuing moral (or non-moral) ideals which our psychology makes it close to impossible to reach, we should refrain from trying to overcome our parochialism and learn how to live with or even embrace it. According to Greene, however, we can still search for a common normative currency that can constitute a "metamorality" which will allow us to mitigate intergroup conflict and reconcile competing intragroup moralities. Moral progress is feasible if we

emphasize our common moral ground. We can buy into the evolutionary account of how parochialism evolved without buying into its conservative implications.

The goal of moral progress is to eliminate the wrongs caused by discrimination, conflict, and poverty. But if inclusivist attitudes can only thrive under conditions of economic prosperity and social stability, expansions of the moral circle will be out of reach in most places, and remain psychologically feasible only where the cooperative conditions they are supposed to result in already obtain. Now, it needs to be emphasized that the fact that progressive attitudes both require and facilitate favorable social conditions does not entail that there is no way out of this apparent vicious circle, as long as there are ways to get the progressive ball rolling without relying on those very attitudes. In the following, I will argue that various forms of institutional support can indeed perform this function of unleashing moral progress.

2.4   The Wrong Kind of Progress

In what follows, I wish to suggest ways around this problem. For one thing, empathy and altruism may be particularly unsuitable for achieving moral progress (Bloom, 2017). For another, it may be that we should not expect our individual moral psychology to play an important role in promoting or maintaining moral progress to begin with.

The good news is that the problem with Buchanan and Powell's naturalistic account of moral progress is homegrown. They are overly focused on the wrong kind of moral progress, and the wrong kind of mechanism to support it. This is at least somewhat surprising, since they clearly appreciate the general importance of such other types of progress and the importance of institutional support. What they fail to see is their potential when it comes to countering the challenge from evolutionary conservatism. In the following two sections, I will elaborate on these points.

The main task is to find a form of moral progress that satisfies three conditions:

 (i) it is able to deliver the same kinds of moral gains as the expanding circle of moral concern is (supposedly) able to while
 (ii) not being subject to comparable evolved psychological constraints and
(iii) not being an evolutionary "luxury good".

Here, it is worth mentioning again that the expanding circle of moral concern is only one type of moral progress among many. I will have more to say about this in Chapter 5. Other types include processes of proper demoralization, proper moralization, or indeed contractions of the moral circle (Evans, 2017; Summers, 2017; Arruda, 2017;

Rønnow-Rasmussen, 2017). Life insurance used to be considered morally repugnant, and so were early forms of vaccination, or flushing a toilet on screen. Nowadays, these things have become demoralized, and rightly so. Conversely, workplace sexism used to be seen as a morally neutral fact of life. It constitutes progress that it is now considered morally problematic.

Buchanan and Powell (2017) are of course aware of this. However, in developing their naturalistic theory of moral progress, they nevertheless decide to focus on inclusivist moral gains and the psychological capacities such as altruism or empathy that underwrite them. This is a mistake, because it deprives them of a plausible and normatively satisfying response to the challenge from evolutionary conservatism. Empathy and altruism *are* limited. Fortunately, there are other emotions beyond empathy, and other forms of moral progress beyond expansions in the moral circle, that can deliver the same moral goods of moving us towards more inclusive and less discriminatory practices. I will use disgust and demoralization as my main examples, but the point generalizes to other attitudes as well.

Dynamics of demoralization can replace expansions of the moral circle. It is misleading to suggest that progressive developments towards inclusivism share an elective affinity with empathy and altruism. Exclusivism is the historical default, but it doesn't result from a lack of empathy as much as revulsion and aggression towards members of other groups. The reason sexist oppression prevails in, say, India isn't that powerful men don't care about or don't particularly like women, or that they don't consider them to be members of the moral community. It is that they mistreat women on the basis of dehumanizing attitudes of disgust, in combination with having the power to put those attitudes into practice.[2] This makes redirecting or mitigating (moral) disgust towards marginalized people at least as promising and potentially powerful as expanding empathy and altruistic dispositions towards a larger group of people.

A frequently overlooked advantage of disgust is that the mandates of empathy and other inclusive attitudes frequently remain within the realm of what's praiseworthy rather than obligatory. "Imperfect duties" of aid or assistance require judgment; the demands of disgust, on the other hand, yield claims of impermissibility and obligatoriness that admit of much less leeway. What is disgusting *must* not be done; what is helpful *may* (or may not) be (Kahan, 1999). Moreover, disgust is motivationally powerful. Empathy usually is not (Prinz, 2011; Bloom, 2017) or, more precisely, empathy is often surprisingly motivationally weak even when something of great moral significance is at stake (think of global poverty). Its motivational strength remains biased in favor of one's ingroup of friends and family. Indeed, altruistic tendencies and disgust have complementary virtues and vices: empathy may be inherently benign, but is weak and rigid;

disgust may be inherently problematic, but it is forceful and malleable. Empathy mainly picks up on distress cues, which are processed in a biased fashion. What people are disgusted by, on the other hand, is highly culturally variable (beyond a core of markers for pathogen and toxicity vectors).

Now, the problem with disgust is that it seems like a paradigmatically *non*-progressive emotion, and thus an awkward bedfellow for those who want to accelerate moral progress. There is extensive evidence linking disgust to concerns about bodily purity and generally conservative political views (Graham et al., 2009). It is also often thought to be particularly unreliable. Due to its origins as a safeguard against contamination, it follows a hypersensitive "better safe than sorry" logic that yields an excessive number of false positives (Kelly, 2011). When disgust is coopted to police social norms, it can have all kinds of pernicious effects (Nussbaum, 2009).

In order to enlist disgust for progressive causes, one needs to get people to be disgusted by the right things and, perhaps even more importantly, prevent them from being disgusted by the wrong things. This is where the notions of demoralization and (proper) moralization come into play. Demoralization happens when individuals and/or groups get rid of unjustified pseudo-moral prohibitions. Vivid examples for such "surplus moral constraints" (Buchanan & Powell, 2017) are the stigmatization of menstruating women in India or discriminatory attitudes towards the disabled. Disgust, or lack thereof, plays a crucial role in eliminating such harmful prohibitions. Proper moralization, on the other hand, happens when people start to disapprove of practices whose moral odiousness had previously gone unnoticed.[3]

The prospects for a convincing response to the challenge from evolutionary conservatism would greatly improve if disgust could be "appropriated" (Kahan, 1999) for such progressive causes. Historically, disgust has frequently played a role in sustaining unjustified moral norms, such as rules against "miscegenation", religious satire, or various medical innovations (Kass, 1997). But we don't have to be disgusted by interracial marriage, provocative music videos, or stem cell research. People can get rid of their intuitive revulsion towards such actions. After a while, people cease to develop it in the first place. Instead, they can become disgusted by such instances of harmful moralization themselves. Disgust towards sacrilegious art can turn into disgust towards those who want to regulate art for the sake of religious ideology (Kahan, 1999, p. 65ff.). Disgust towards homosexuality can turn into disgust towards "Don't ask, don't tell" advocates. The list item (i) thus seems to be satisfied: there are some normative attitudes that can secure the moral gains typically associated with extended empathy, disgust being one example.

Indeed, some authors have recently tried to save disgust from its dubious reputation. Alexandra Plakias (2013), for instance, argues that due to its properties as a detector of vectors of disease, disgust is suitable for tracking

social contagion as well. Immoral behavior can spread through populations in a way that is more or less analogous to disease and contamination. Disgust detects social contamination. In a similar vein, Kumar (2017) writes:

> that disgust is implicated in important moral norms and values that are shared by liberals and conservatives. Disgust is repurposed in ways that support these norms and values, by motivating an important form of punishment, tracking the spread of moral violations, and expressively coordinating collective action. Disgust accurately detects the nature of certain wrongs that commonly elicit moral revulsion.
>
> (13)

What about condition (ii)? Moral disgust (and other emotions) can play an important role in pushing towards further expansions of the moral circle. But the main advantage of disgust, from an evolutionary perspective, is that it does not suffer from comparable evolutionary constraints. What we are and are not disgusted by is highly malleable (Tybur et al., 2013). Some authors have suggested that, if anything, disgust is *too* malleable to enjoy any normative authority (Knapp, 2003).

In many cases, redirecting disgust towards progressive causes is not even necessary. All that is required is for people to shed their revulsion. And clearly, there are virtually no evolutionary constraints on how *few* things people can become disgusted by. Narrowing down what's considered gross to feces and food would be a major moral accomplishment. Disgust can also incorporate and update on morally salient information. Famously, people can become disgusted by something, such as the sight and smell of meat, and they have found to be morally objectionable on disgust-independent grounds (Rozin et al., 1997). The conservative argument from hardwired evolutionary constraints thus cuts no ice against disgust.

Perhaps most importantly, repurposed disgust is not an evolutionary luxury good. Buchanan and Powell argue that inclusivist attitudes only become psychologically feasible under favorable socioeconomic conditions of economic prosperity and political stability. This is a problem for advocates of moral progress, because it means that societies can only enter the track towards moral gains if they are already on it. Disgust does not suffer from a similar problem. People can easily become disgusted by all kinds of different things, regardless of whether conditions are good or not. Besides a minimal core, what cues disgust is triggered by seems highly cultural variable. Empathy and altruism, on the other hand, are either parochial or inclusivist, depending on the circumstances. Condition (iii) is thus satisfied, too.

One may doubt that moral disgust should count as genuine disgust (Gert, 2015). Since we are talking about what is and isn't disgusting, introspective evidence seems admissible. Consider the litany of human atrocity

that is Livingstone Smith's *Less Than Human* (2011). The list of disembowelments, cut off breasts, severed limbs, or baby's heads flung against walls is endless. The descriptions of genocide by the Spanish against the Native American population, by the Germans against the Jews, by the Hutu against the Tutsi, or by the Japanese against the Chinese are frequently viscerally disgusting.

It is likely that a fair amount of talk about how "disgusting" a reprehensible action is purely metaphorical. Moreover, immoral actions frequently just happen to be disgusting. On the other hand, there is neuroscientific evidence suggesting that even so-called "pure" moral violations—that is, actions which aren't independently disgusting such as mutilations or unusual sex acts—can elicit reactions of disgust (Kumar, 2017, p. 4f.). This link is buttressed not just by self-reports but also by facial activity, implicit measures, and behavioral evidence (Chapman et al., 2009). On the other hand, some have suggested that the link between disgust and the severity of our moral attitudes is less strong than initially thought (May, 2014; Landy & Goodwin, 2015).

Some may object, on moral grounds, to letting disgust perform inclusivism's work at all. An "exclusionary" attitude such as disgust, which is designed to "keep things out" rather than let them in, seems like an ambivalent ally in the fight for expansions of the moral circle. But such expansions are not ends in themselves. What we want, ultimately, is for human beings not to enslave, torture, kill, or otherwise dehumanize others. It is of course desirable when individuals come to respect each other as equals. But ultimately, the value of respect lies in the moral protection it affords. We want to be recognized because we don't want to be harmed, violated, and infringed upon. If exclusivist attitudes never led to such violations, we would have little reason to care about them. Expanding criteria of membership in the moral community are instrumental to that goal. Buchanan and Powell accept this point, because they object to exclusivist ideologies precisely because they can lead to disenfranchisement, dehumanization, and genocide. If disgust performs better at this task, there is little reason to complain.

The expanding circle of moral concern is not always the most suitable form of social change for the better for progressives to focus on. But it seems that what I am actually suggesting is not to give up such inclusivist shifts as a valuable end but to look for different *means* of accomplishing the same end.

This impression is, to a large extent, justified. As I have tried to emphasize, one of the virtues of disgust, along with other moral attitudes such as anger or indignation, is that they can be used to secure *the same moral gains*. At the fundamental level, these moral gains consist in reducing discrimination, xenophobia, and the gratuitous or instrumental infliction of suffering on members of the outgroup. Such gains can be achieved either by (de)moralizing certain behaviors or by adjusting criteria of membership in the moral

community. My point was that the latter strategy is less promising because of the constraints our evolved psychology imposes in such adjustments. By focusing on inclusivism as the main form of moral progress, we are more likely to end up focusing on inadequate psychological means of achieving it. But even if my alternative strategy succeeds, the expanding circle of inclusivity would remain, at the very least, an extremely welcome byproduct.

Second, one may doubt the empirical side of my story. According to this worry, there is simply no reason to believe that disgust fares any better than empathy or altruism do when it comes to how strongly its development and subsequent content is canalized by evolution. Disgust may be equally vulnerable to a challenge from evolutionary conservatism.

It may of course simply be true that the remedy for evolutionary conservatism I have identified falls short because the empirical details of my account do not pan out. This would not entail, however, that, *given what we know about the limitations of inclusivist attitudes*, the general recommendation of looking for other, less evolutionarily constrained ways of securing progressive moral gains would not be worth making. Consider, for instance, the emotion of shame. Like disgust, shame is highly malleable in terms of the sheer variety of (sometimes bizarre) things people can become intensely ashamed of doing or failing to do. Appiah (2011) famously describes the intense pressure, which sometimes persisted for centuries, to participate in gentlemanly dueling or foot-binding in Europe or China, respectively. Not doing so would result in enormous shame, social stigma, and ostracism. And yet these centuries-old norms evaporated within a generation or less, never to be encountered again.

The historical record suggests that disgust is malleable in just this way. Homophobic attitudes used to be (and in many cases still are) underwritten by strong feelings of disgust. The same holds for so-called "usury" or "miscegenation". But the amplification of inclusivist attitudes hardly played a role in ending discrimination against gay people, Jews, or people in interracial relationships. Disgust, on the other hand, is less constrained by evolution because seemingly unlimited range of things people can become disgusted by and the staggering speed with which even whole societies can cease to be disgusted by various actions. This remains true even if one concedes that disgust is subject to some evolved limitations, and that it sometimes does play a role in maintaining ingroup/outgroup divisions.

I have focused on disgust here, but the same point generalizes to other feelings. Take attitudes of respect/honor. Contractions of the moral circle can also be progressive, such as when excessive belief in and respect for the (alleged) authority of powerful people is curtailed (Huemer, 2013). Here, too, there seem to be no evolutionary limits as to how many and which things we can *un*learn to respect. Disgust supports dynamics of demoralization, developments of respect, and honor support contractions of the moral

circle. Both of these forms of progress support the kinds of developments we are after when we think about expansions of the moral circle: reductions in violence, genocide, discrimination, or dehumanization. But neither is subject to comparable evolutionarily inherited psychological constraints.

2.5   Does Evolution Constrain Moral Progress?

To counter the evolutionary challenge, many authors try to identify features of our moral psychology as well as of morality itself that are not amenable to an evolutionary explanation. One example for this can be found in the literature on so-called evolutionary debunking arguments (Street, 2006; Sauer, 2018).

A prominent type of debunking argument tries to undermine some or all of our moral beliefs by pointing out their epistemically dubious etiology (Nichols, 2014). If our moral beliefs result from processes which we know to be unreliable, or at least have no good reason to consider reliable, then those moral beliefs must be reconsidered or abandoned. At the very least, we ought to adjust our confidence in them in proportion to the impact of their contaminated genealogy.

The evolutionary version of such debunking arguments has it that a commitment to evolutionary explanations of our moral beliefs together with a commitment to robust moral truth yields moral skepticism. If moral knowledge consists in getting the objective moral facts right, and our moral beliefs are heavily influenced by what it would be adaptive to value, and if what it would be adaptive to value has no recognizable connection to what the moral facts are, then our moral beliefs are likely to be saturated with morally irrelevant influences, and are thus thoroughly unjustified.

Perhaps, however, it is possible to identity a subset of our moral beliefs for which there is no plausible evolutionary account? If so, then those moral beliefs would stand vindicated, or would at least be immunized from evolutionary debunking. Buchanan and Powell claim that no selectionist explanation of *subject-centered morality* or the inclusivist anomaly is available. In some cases, they make this point somewhat confusingly. For instance, Buchanan and Powell criticize Kitcher's functionalism because of how it identifies moral progress with the gradual removal of "altruism failures", which happen when people's strategic self-interest motivationally overpowers their concern for their fellow cooperators. However, they argue, it "is worth noting that on this interpretation . . ., many of the putative achievements of subject-centered morality, such as basic rights for persons with disabilities or children, will not count as instances of moral progress" (Buchanan & Powell, 2018, p. 53). It is not entirely clear why they would think that treating one's children well cannot be understood in terms of adaptiveness, since caring for the well-being of one's children is

close to the most adaptive thing one can possibly do. There is simply no puzzle here over what the evolutionary explanation of close kin altruism is.

But leaving that to one side, is it true that the inclusivist anomaly cannot be accounted for in selectionist terms? And even if it were true, does that mean that there is no evolutionary explanation of non-strategic moral concern, period? Consider this example: a cultural-learning account of evolution can explain why in a highly cooperative niche (i.e., increasingly large communities such as cities), there are adaptive advantages to people having a non-discriminating disposition to cooperate even with strangers and outgroup members (because of rewards from trade or information transmission). The next generation of norm learners (Sterelny, 2019) then acquires this set of undiscriminating norms of cooperation from the previous generation, such that at the end of this process, we have people acting on the simple norm of "be nice to people in general". The distinction between strategic and subject-centered morality is a misleading alternative. There may be a strategic explanation for why people's cooperative dispositions become purely subject-centered, namely because that kind of norm is crucial for being able to navigate the highly advantageous (because safe and resource-rich) environment of hypersocial, large communities.

Railton (2017) makes the following, equally plausible, suggestion:

> Why might this be [that people can recognize the moral status of the outgroup, H. S.]? The ethnographic record indicates that many hunter-gatherer bands practice exogamy (i.e., marriage takes place with individuals outside the group), and field observations suggest that individuals not infrequently shift from one band to another, whether as an outcome of warfare, social exclusion, migration, decay of group size, or attempts to secure better prospects. . . . Hunter-gatherer social and trading networks can be extensive, and "functional social proximity" through shared activity or exchange is often more influential than actual "kin proximity" in shaping behavior. . . . Engaging effectively in these more flexible and less directly group-centric ways of life puts a premium on the ability to "size up" and interact with strangers and other groups in light of assessments of general, modal, morally-relevant characteristics such as cooperativeness, trustworthiness, competence, knowledgeability, aggressiveness, or tendencies to help or share. Group selection and sexual selection likewise involve such traits, and have been identified as processes that can favor the emergence of significantly altruistic behavior that cuts across genetic relatedness.
>
> (Railton, 2017, p. 179)

It is also unclear why Buchanan and Powell resist "misfiring" explanations of the expanding circle, according to which the proper function of

empathy is parochial, but maybe simply be activated and misfired when around unrelated/unfamiliar people. They seem to suggest that a misfiring story would help the evolutionary conservative, but it is hard to see why. Consider an analogy with sexual arousal, *vulgo* horniness: horniness has been selected for a particular function. But people still get horny even when using contraceptive measures, thereby sabotaging the original purpose of horniness, that is, to promote reproduction—a classic case of misfiring. In a way, we are "expanding the circle" (to allow me the silly analogy) of what we consider "sexy" towards (temporarily or permanently) non-reproductively relevant individuals. Here, misfiring leads to expansion; why can't moral concern be like that as well? If this story were correct, this would in no way support a form of evolutionary conservatism about sexual attractiveness. Compare the plausibility of an evo-conservative account of horniness: we can recognize that horniness has a specific evolutionary function. But due the fact that it can be made to misfire, people have an "open-ended" capacity to become attracted even by people who wear condoms or use the pill.

Most importantly, we now do have a fully worked out account of the expanding circle that does explain the inclusivist anomaly in selectionist terms: the cultural evolution of impersonal prosociality:

> [W]e propose that the spread of the Church, specifically through its transformation of kinship and marriage, was a key factor behind a cultural shift towards a WEIRDer psychology in Europe. This shift eventually fostered the creation of new formal institutions, including representative governments, individual rights, commercial law and impersonal markets.
>
> (Schulz et al., 2018)

I will not spell out this explanation in detail, but let me sketch how it may work. In order to function properly, modern economies require social norms regulating and governing how to interact with strangers in cooperative ways. Once these generalized norms for how to treat anybody, not just kin, are in place, positive-sum interactions become possible that further entrench those very social norms. And once one has social norms that facilitate indiscriminate gregariousness of this sort, subject-centered morality is only a small step away. And indeed, the main shifts towards such subject-centered moralities seem to coincide precisely with the cultural-evolutionary replacement of kinship norms in favor of more "impartial" norms. The key thing to explain is the emergence of "impersonal prosociality" (Henrich, 2020), that is, the expanding circle of moral status, and this can be explained in terms of cultural evolution. This does not entail, again, that evolutionary conservatism is vindicated.

It is possible to concede the central evolutionary conservative point without sacrificing the core progressive aims worth wanting. This is because even if we admit that there are some evolved constraints on social change, it is much less clear just how tightly constrained the resulting space of social change ends up being. Suppose, for instance, that we conceded to an evolutionary conservative about gender the basic claim that there are some significant innate differences in skills and interests between women and men that are impossible, as good as impossible, or prohibitively costly, to change. Perhaps we are thereby conceding that a society with full-blown gender equality—that is, a society in which it doesn't matter at all which gender one belongs to, and in which gender has literally no discernible systematic effects whatsoever—is out of reach. This admission, however, has no real bearing on how the question of social gender equality approached *today*. All existing societies seem to have some degree of gender inequality and indeed inequity. Consider, however, the fact that there are human societies, like the African Mbuti, who practice hardly any gender inequality, or the Brazilian Mundurucu, on the other hand, who have starkly different norms for the genders, including separate housing and highly polarized social rituals (O'Connor, 2019, p. 198ff.; see also Oakley, 2015). So even if there are some evolved, biologically hard-wired, and impenetrable differences between the sexes, the potential for the social changes we want remains huge.

This account of the emergence of impersonal prosociality also makes most sense *regionally*. It is clear that the trends Buchanan and Powell mention—animal rights, human rights, moral universality—first emerged in a particular region of the world, namely Western societies, spreading only later. If "open-ended normativity" were the specific explanation for these progressive trends, then it would remain inexplicable why these trends first emerged where they emerged, rather than several times in different places.

Notes

1. It is important to stress that Smith himself did not intend this observation to support any normative defense of partiality. The quote continues:

   To prevent, therefore, this paltry misfortune to himself, would a man of humanity be willing to sacrifice the lives of a hundred millions of his brethren, provided he had never seen them? Human nature startles with horror at the thought, and the world, in its greatest depravity and corruption, never produced such a villain as could be capable of entertaining it. But what makes this difference? When our passive feelings are almost always so sordid and so selfish, how comes it that our active principles should often be so generous and so noble? When we are always so much more deeply affected by whatever concerns ourselves, than by whatever concerns other men; what is it which prompts the generous, upon all occasions, and the mean upon many, to sacrifice their own interests to the greater interests of others? It is

not the soft power of humanity, it is not that feeble spark of benevolence which Nature has lighted up in the human heart, that is thus capable of counteracting the strongest impulses of self-love. It is a stronger power, a more forcible motive, which exerts itself upon such occasions. It is reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct.

(158)

2. See, for instance, www.theguardian.com/global-development/2015/dec/22/india-menstruation-periods-gaokor-women-isolated
3. Note that, not infrequently, this moral odiousness has gone unnoticed by the *victims* of those practices. See Fricker (2007).

3   The End of Utopia
Does Moral Progress Have a Goal?

Introduction

If nothing else can make you doubt that there is progress in history, perhaps the fact that Hegel thought so should succeed in making you skeptical:

> It has already been shown and will again emerge in the course of this enquiry that the history of the world is a rational process, the rational and necessary evolution of the world spirit. This spirit [is] the substance of history; its nature is always one and the same; and it discloses this nature in the existence of the world. (The world spirit is the absolute spirit.) This, as I have said, must be the result of our study of history.
>
> (Hegel, 1975 [1830], p. 29)

In his *Philosophy of Right*, he would sum it up as concisely and infamously as possible: "What is real is rational, and what is rational is real".

But surely history cannot be headed for what's right and what's good? It seems that this idea has been discredited a long time ago:

> At the start of the twentieth century, reflective Europeans were . . . able to believe in moral progress, and to see human viciousness and barbarism as in retreat. At the end of the century, it is hard to be confident either about the moral law or about moral progress.
>
> (Glover, 1999, p. 1)[1]

This is a truism. But is it true? The statement that we are no longer "able" to believe in history as inevitable progress dictated by reason suggests a psychological unwillingness, a certain kind of cognitive resistance that many contemporary writers—myself included—seem to encounter when contemplating this possibility. How *could* history be rational? We, the transcendentally bankrupt, simply cannot afford such luxuries anymore. Some time ago, we defaulted on our eschatological debts, and haven't been able to get back on our feet since.

That we remain condemned to irreparable metaphysical homelessness was perhaps most forcefully articulated by Adorno and Horkheimer. The enlightenment, they observe, was humanity's best shot at progress, that secular sibling of salvation. And yet, far from delivering on its promise, "the fully enlightened world is ablaze with triumphal doom" (Horkheimer & Adorno, 1998 [1944], p. 9; translation H. S.). The 20th century, at the latest, lays waste to the idea that moral progress is possible, much less that there is any guarantee of it happening.

This chapter has four sections. In section 3.1, I address the issue of whether moral teleology can be naturalized. Section 3.2 is about the normative ambivalence of the teleological conception of moral change. Section 3.3 deals with the concept of teleology and shows that the usual objections to it, while in principle valid, primarily apply to the strongest version of the teleological thesis. A more moderate form of teleology may still turn out to be defensible. Finally, in section 3.4, I introduce some preliminary reasons for why moral teleology should be taken more seriously than it currently is.

3.1   Naturalizing Teleology?

Don't get me wrong: I understand why skepticism towards teleology is tempting. The idea that history is somehow biased towards moral improvement just seems too fanciful and self-congratulatory, a sign of delusional privilege more than healthy realism, a lullaby for passengers on a sinking ship. And I agree that the teleological structure of social change, even if real, isn't *obvious*; Hegel knew that we need to look *at* history with reason for it to look *back at us* with reason as well. Or, as Paul Auster once put it: stories only happen to those who are able to tell them.

Contemporary philosophers try not to succumb to teleological temptations, regardless of how alluringly the narrative is crafted. Perhaps this is for fear of appearing superficial. There is an exploitably thin line between depth and gloom that is all too easily trespassed; still, the strength of this anti-teleological resolution is at least somewhat surprising, because for much of history, the idea that human development had some sort of goal was virtually taken for granted (Herman, 1997; Nisbet, 1980). Theorists of decadence like Rousseau or Nietzsche thought so, however dubious a goal they thought it was moving towards. Even cyclical theorists like Oswald Spengler assume goals, albeit many, bound up in cycles of cultural blossoming and decay. And, of course, teleological thinking is central to many religious traditions and their visions of judgment day, a return to a golden age, a rediscovery of paradise lost—in this world or the next.[2]

The very attempt is seen as hopelessly naïve and whiggish, the product of wishful thinking more than a sober scrutiny of the facts. In fact, Buchanan and Powell formulate not as a result, but as one of the

desiderata for any viable theory of moral progress that it cannot be teleological: "As a scientifically informed secular theory, our account eschews teleological thinking about nature, human nature, and the nature of society" (Buchanan & Powell, 2018, p. 29). This, they claim, is due to the fact that such directional accounts of history lack naturalistic credentials and empirical support. One main

> defect of some secular conceptions of moral progress was that they claimed, without evidence, that moral progress was inevitable, not merely feasible. Given a near total lack of solid empirical grounding, the claim that moral progress was inevitable was even shakier than the claim that it was feasible
>
> (Buchanan & Powell, 2018, p. 27)

Moreover, they contend that a satisfactory account of moral change "should not assume that moral progress is necessary or inevitable or that there are "iron laws of progress" (Buchanan & Powell, 2018, p. 31)". Teleology, they hold, cannot be naturalized.[3]

Why did teleological ambitions have to be shelved? Much of it, I speculate, has to do with the notion that modern philosophy—that is, roughly speaking, philosophy after the Holocaust and the Linguistic Turn—had to be conducted under so-called "post-metaphysical" conditions (Makkreel, 1992). It is hard to articulate what precisely these conditions are supposed to amount to, but the general idea seemed to consist in the admission that reality does not have an objective metaphysical structure which can be accurately deciphered by the powers of reason and thoroughly laid out in a unified, coherent philosophical system. Instead, humans are contingent beings with finite powers of understanding. Modern philosophy's teleophobia stems from a recognition of this fact.

Teleological thinking seemed to depend on several metaphysically dubious propositions without which its very possibility appeared to collapse. It may seem, for instance, that teleological accounts of history inevitably remain committed to some version of theism. If history is directed towards some goal, then there must be someone who embraces this goal and actively pursues it via history; arguably, this someone would have to be a being with incredible intelligence, awesome power, and a passionate interest in morality—that is, a divine entity.

Non-theistic accounts of teleology, on the other hand, may seem committed to implausible assumptions about the possibility of backwards causation. If society develops towards a goal, but this goal isn't represented and promoted by some supreme being, then how can it exert its guidance upon the course of history? This seems possible only if the *telos* of history can shape its own genesis. History's effects must precede their causes.

Neither option seems very attractive from a naturalistic perspective. But perhaps most damagingly, the very idea that moral progress could be teleological appears preposterous when one realizes appreciates the sheer contingency that is baked into social developments, however morally desirable they may be. Consider, for instance, the somewhat mysterious decline in violent crime rates in the USA. In 1993, the US murder rate peaked; since then, violent crime of all types has gone down by 300%. But why? One reason for this appears to be the fact that in 1976 and 1978, respectively, the USA banned the use of lead in cars and paint (Easterbrook, 2018, p. 111ff.). Lead poisoning has a well-known negative effect on impulse control and delinquency (Reyes, 2015). And the decline of violent crime and homicide are clearly morally desirable. However, if the latter can, at least in part, be attributed to the former, moral progress seems to be accidental all the way down.

Moral progress, even if occasionally real, may be unlikely to last (Norlock, 2018, p. 5). In his landmark study on the idea of Western decline in history, Arthur Herman concludes: "The notion of history as progress stands largely discredited today among intellectuals, and especially among historians" (Herman, 1997, p. 13; for a notable exception, see Godlovitch, 1998).

In light of this in part naturalistically, in part morally motivated skepticism about the alleged Panglossianism inherent in teleology, some grounds of which are admittedly reasonable, I want to take a step back to see what a teleological account of history would entail. Only then will it become possible to assess the prospects of defending it on the basis of naturalistic premises. But before I do so, let me briefly address whether the truth of teleology would be good or bad news.

3.2   Normative Ambivalence

The truth of teleology seems inherently desirable, from an evaluative point of view. Things have gotten better, and will continue to do so—what's not to love? However, there are a few reasons why teleology may be more normatively ambivalent than it initially appears. I will mention five.

For one thing, a normative problem for teleological accounts of progress is that they seem to nurture an attitude of quietism and inaction. Why do anything to make the world a better place, if teleological progress takes care of itself? The later Marx, for instance, is sometimes (falsely, but commonly) read as abandoning the idea that anything would be asked of us to be done about the injustices he diagnosed in capitalist exploitation (Leiter, 2015). To him, the overthrow of capitalism was supposedly an inevitable result of the internal contradictions between the relations of production and the productive powers. There was little or nothing that could or had to be done for or against the end of capitalism, just like there was nothing to be done

for or against capitalism when it first emerged to inevitably replace feudal economies. Conceptually, however, the idea that teleological forces make moral action for the sake of progress superfluous seems based on a confusion: teleology does not entail that moral improvements will occur in a way that *bypasses* human action. The claim that history gravitates towards moral improvement is true (if indeed it is) because individual people and social movements make it so. Teleology does not warrant complacency. Whether it leads to it as a matter of psychological fact is a separate, empirical matter.

The converse problem for teleology is that when history is thought to move towards a utopian end state of moral perfection, anything seems justifiable to bring this state about (Jones & Paris, 2018). And the maths check out: when the potential gains are infinite, arbitrarily high sacrifices can be justified, even when the chances of success are arbitrarily low. A concern for justice, Camus notes, can forge human solidarity and peace; utopian thinking, on the other hand, leads to the "logical crime" of bureaucratically organized mass murder: topsy-turvy obscenities such as "slave camps under the flag of freedom, massacres justified by philanthropy or by a taste for the superhuman" are the result (Camus, 1991 [1951], p. 4). When one has found the ultimate recipe for utopia, who gives a damn if a few, or a few million, must perish to attain perpetual bliss? Once justice has prevailed, those sacrifices will have long been forgotten, and we can spend the rest of eternity counting our blessings.

Third and relatedly, the idea of teleology can be used to justify inequality and colonial interventionism, allegedly for the sake of those cultures or nations who haven't heard the good news yet (Levy, 2020, p. 182). Then again, it is a far cry from claiming that things can and do get better to the idea that this entitles some nations to violently coerce the people of other nations to make those same changes now whether they want to or not. It seems that in many cases, critics of teleology conflate its truth with whether it has the potential to be abused for bad ends. If we don't wish to admit this strategy as epistemically relevant when it comes to other theories that have led to abuse—think: Social Darwinism—why should we admit such arguments here?

Fourth, there seems to be a possible route from moral progress to moral skepticism. Frequently, the fact that we have come to recognize racism or sexism as wrong is offered as a ground for epistemic optimism: *look*, we *can* acquire moral knowledge—well, done, humanity! However, the opposite conclusion may actually be more natural. Via a kind of "pessimistic meta-induction", we may come to ask ourselves that if we've been so radically wrong about morality before, who can guarantee that we aren't radically wrong *now* about what morality permits and requires (Stokes, 2017)? Moral progress becomes a genuine skeptical hypothesis, akin to the familiar "brain in a vat" scenario.

Finally, a teleological account of the social world can foster nihilism in the individual. In his *Science as Vocation*, Max Weber makes the interesting observation that in ancient times, it was possible for a person's life to reach a stage of completion and roundness. When there is progress, this becomes close to impossible (Weber, 2004 [1919], p. 13). Paradoxically, moral and social progress, while being the source of tremendous benefits for all, can be a source of existential anxiety for each. This is, perhaps, why declinist phantasies continue to hold such sway over the excitable imagination of the chattering classes, where pale-skinned pessimist intellectuals sometimes reveal their ghoulish fascination with fantasies of a future wasteland of environmental destruction and social mayhem.

Critics of teleology often instinctively respond to the claim that there has been moral progress by coming up with counterexamples: but what about the rise of authoritarian populism? Are you unaware of the return of white supremacy? Haven't you read the news today? In one sense, these points are well taken. But the claim that there has been moral progress is not incompatible with the admission that there are also pockets of regress, reaction, and revanchism. Progress will always remain under threat.

The assumption that individual counterexamples should disconfirm progressive trends is, in a sense, undialectical: progressive developments are constituted by a movement that generates its own obstacles. Societies always develop in waves and swings, actions that provoke reactions, revolution and restoration, protest and pushback. That's what progress *is*—it's not a smooth, uninterrupted ride into the sunset, but a fight. Fights require enemies. But the fact that there are some battles along the way doesn't entail that the war isn't being won.

3.3   What Is Teleology, Anyway?

People do not agree on what teleology is, but what they do agree on is that it is false. In this section, I wish to bring some clarity into this debate, and zero in on what I take to be the most plausible candidate for what "teleology" could possibly mean. The remainder of the book is dedicated to whether teleology, so understood, may be true.

Teleological claims vary along several dimensions that can be adjusted for strength. Most of its critics believe that teleology is false because its strongest and least plausible version likely is. This rather clearly does not follow.

The most obvious dimensions of teleology are:

(i) Directionality

A necessary but insufficient feature of teleological accounts of history is that according to them, social change has a *direction*. The reason why this

claim is not enough to amount to full-blown teleology is that it is (close to) trivially true: unless one thinks that humanity will end tomorrow, society will move towards *some* state of affairs, however random and insignificant it may be. For teleology to be true, the direction of history must be in some sense a meaningful one.

When it comes to the direction of history, a more important issue seems to be whether it is merely goal-directed or positively goal-*driven*. History is goal-driven only if the process of history is determined by an actual *telos*, a set goal it unfolds towards. This seems implausible to defend without divine intervention or an objective spirit. That social change is goal-*directed*, on the other hand, amounts to no more than the claim that it tends to move in a particular, non-arbitrary direction. This weaker understanding seems much more plausible on its face.

Another issue has to do with whether the directionality of historical developments should be understood in terms of an *end state*. But the claim that there is an end of history—a state of completion or perfection it culminates in—is merely one specific form of teleology. The traditional term for this view is *eschatology:* the study of the final things. Indeed, some argue that the belief in progress as such is a "heritage of Christian eschatology" (van Bavel, 2018, p. 45). This is not true, however: teleological notions are found in many cultures with little or no cross-fertilization with Judaeo-Christian ideas. But even if it were true, it would not by itself undermine teleology's soundness: "if it's a mistake to suppose that nobility will have exalted progeny, it's also unwise to assume that children must inherit the sins of their parents" (Doris, 2009, p. 706).

It may seem as if talking about moral progress specifically presupposes eschatology. When there is progress, there must be some sort of end goal that it is progress towards. This impression, however, is misleading. We are perfectly capable of making comparative assessments of relative improvement even in the absence of an anticipated end state of perfection, just like we do not need to know what the perfect painting would look like to know that, sadly, our beloved daughters' paintings fall short of it.[4] Such "transcendental" recipes for utopia are neither necessary nor sufficient for judging progress. They may indeed become positively harmful when the quest for articulating a vision of the perfect car prevents me from changing the flat tire of my actual car (Huemer, 2016b; see also Levy, 2016).

What would it mean to deny the directionality of history? Moral nihilism involves no commitment to its denial, but merely to the notion that whatever the facts of history are, these facts lack any moral flavor; moral nihilists can agree that history has a direction even if, for them, that direction is morally empty. The claim that things are actually *in decline* doesn't reject directionality either: it is just teleology for the morose. Alternatively, one could think that social change isn't about improvement *or* decline, but

about stagnation. This, too, however, assumes a direction, albeit a morally neutral one (things develop in a direction, but they get neither better nor worse overall). A proper denial of directionality thus amounts to the claim that social change is a random back and forth with no discernible pattern at all, moral or otherwise (Pleasants, 2018).

(ii) Agency

Teleological claims can vary depending on what they identify as the guiding force behind history's trajectory. Does the long march of humanity have to be secured by blind forces, or is something—*someone*—taking care of the journey? Faith-based trust in moral improvements gives us a straightforwardly affirmative answer to this question. Naturalistic accounts of moral progress try hard to steer clear of any divine puppetry. Instead, they identify mechanisms of social change that can unfold without intelligent oversight. This is not to suggest that they can unfold without individual and social action, much of which is driven by human agents with moral values. If teleology is true, it will be because human beings (partly) make it true. Whichever way history is headed, moral agents are involved in it. But they will be involved in the trenches, not as a "Weltgeist zu Pferde".

(iii) Probability

Third, different versions of teleology differ with regard to how certain they take the realization of their envisioned *telos* to be. Is moral progress *bound* to happen? This question, too, seems to allow for a variety of answers, ranging from the absolute inevitability of moral progress to the concession that nothing in life has better odds than 50/50. Some even hold that the feasibility of moral ideals has no bearing at all on whether those ideals are worth pursuing (Estlund, 2019). Buchanan and Powell, as we have seen, reject moral teleology because they do not believe that moral improvement is "inevitable", or that there are "iron laws" of progress. Fair enough. But the interesting question here is that if the main teleological claim is that there are forces that reliably push societies in the direction of moral improvement, how reliable must this trend be? More than 0%? More than 50%? 99%? This issue is almost always left unaddressed.

(iv) Morality

Teleology is about whether history develops in some direction. It does not strictly speaking entail that this direction be one of moral improvement (or decline, or stagnation). *De facto*, however, teleological accounts of history have almost always taken the goals of social change to be morally imbued.

Optimists see the potential for improvement; for pessimists, disaster is always waiting around the corner.

Eschatology holds that history has an end state, not that that end state is necessarily a good one. Some eschatological narratives, in addition to anticipating an end state of social development, are also *soteriological*, in that they claim that humanity is advancing not just towards some final goal, but towards a morally good one. Soteriology is teleological, but not necessarily the other way around. The most promising form of teleology, I will argue, is non-eschatological and thus a fortiori, non-soteriological: there is no such thing as an "end of history", since further improvements are always possible.

(v) Transparency

It may be that there is moral progress. But how do we know which changes should be counted as improvements, and which as setbacks? One could think that it is impossible to know what the goal of moral progress is because under current conditions of moral imperfection, any positive articulation of the ideal of social development would inevitably become skewed and wonky—tainted, as it were, by the perverseness of the current shape of life. Adorno's "methodological negativism", for instance, embraces this result (Freyenhagen, 2013). Moral and political philosophy, on this account, consists primarily in the "denunciation of the inhumane" rather than any concrete representation of the good.

Critics reply that this negativistic attitude renders the very practice of critique in the name of progress hollow and impossible, for a purely negative ethics cannot justify its demands for improvement. However, both sides underestimate the prospects of a third way between pure negativism and overly concrete and thus epistemically dubious utopianism, which is the possibility of pairwise comparisons between social states in terms of *better* and *worse*. Such non-ideal comparisons are based neither on a mere identification of what's not good nor on an epistemically dubious vision of what's ultimately and pristinely good: "We must retool our values from within" (Prinz, 2007, p. 289). It may seem that we can never know whether moral progress actually occurred, for in order to do so, we would have to be able to avail ourselves of some sort of external, Platonic moral standard in light of which we could assess whether a particular instance of social change should count as progress or not. This, however, overshoots the target.

(vi) Scale

Who is subject to those teleological forces, if there are any? Is it the whole universe, careening towards reconciliation? All of humankind? Or merely

this or that society, culture, or group? The second law of thermodynamics sees the universe develop towards some irreversible state of thermodynamic equilibrium, *vulgo* heat death. This narrative is arguably directional, but admitting the inevitability of this fate would strike few as distinctively teleological. It just seems that the whole universe isn't the kind of thing whose trajectory should be evaluated in terms of a potential goal-directed structure to begin with. I aim to focus on the possibility of at least somewhat more regional forms of progress that allow for reliable trends of enhancement in some places, without postulating that *all* humanity must therefore be swept up in the same maelstrom. However, the quantitative question remains interesting: does teleology require that all of humanity, that is, each person without exception, should undergo the required moral improvements? This seems implausibly strong. On the other hand, suppose there were a highly specific pocket of progress somewhere: a relatively small social group that, by all available metrics, reliably and steadily improves morally. This would be puzzling and astonishing, but would it be sufficient to vindicate teleology? Does teleology require moral progress to at least keep an eye on the whole world?

(vii) Uniqueness

A related worry about the possibility of moral progress has to do with its potential *path-dependency*. What did the Irishman tell the tourist who asked for directions to Dublin? "I wouldn't start from here". *We* may be that tourist. If we want to reach a state of moral perfection, we inevitably have to start from where we are now—there is no other starting point. But it may be that the point of departure we have inherited, and the possible trajectories of social development that are accessible from it, blocks us from ever reaching the juiciest fruit. When it comes to the prospects of progress, we may have painted ourselves into a corner.

(viii) Timing

If the forces pushing society in the direction of moral improvement are in place, then why did moral progress of the kind we praise today (abolition of slavery, animal rights, and so on) not happen earlier? This is an excellent question. Note, however, that this is an explanatory burden any account of social change has to shoulder. Why, if whatever account of historical change is correct, do certain shifts happen at one point in time rather than another? Thus the question of timing, though a problem for a teleological account of moral progress, is not a problem *specifically* for such an account. According to Buchanan and Powell's non-teleological theory, for instance, inclusivist moral gains are enabled by a human capacity for

"open-ended normativity" (Buchanan & Powell, 2018). Presumably, however, this capacity has been around for quite some time. Why, then, did the dynamic of expansive morality not happen earlier, or indeed later? Why did it happen, and when it happened? This is, again, not a problem that afflicts their theory specifically; rather, it is a formidable puzzle any theory of moral progress (or regress) must contend with.

Where does this leave us? We face a *spectrum* of potential teleological views, the strongest of which max out on all of the aforementioned dimensions: history is guaranteed (iii) by a supremely powerful intelligent agency (ii) to push all humankind (vi) towards a knowable (v), unique (vii) end state (i) of moral perfection (iv) that had to happen precisely when it happened (viii). On the other end of the spectrum, we find metaphysically much more humble varieties of teleology, according to which there are some mechanisms in place that more or less reliably push some human societies towards some degree of moral improvement. Such accounts still deserve to be called teleological because they preserve the crucial feature that history is biased in favor of change for the better.

Cavalier dismissals of teleology almost always end up—deliberately or not—strawmanning it. The aforementioned quotes all reject teleological accounts of moral change according to which there are supposed to be "iron laws" of "steady progress" towards "moral perfection". Moderate forms of teleology are almost never given a fair hearing. Such a moderate form of teleology, according to which there are some mechanisms in place that push groups towards gradual, open-ended moral improvement, is the one I aim to articulate and evaluate.

3.4   Taking Teleology Seriously

Teleological thinking has, ironically, been in decline. But before teleology became a four-letter word, it had many distinguished adherents. I have already mentioned Hegel, but the further we go back, the easier it is to find prominent voices who claimed to see fragments of reason in social transformations (Nisbet, 1980). In some cases, this took the form of (quasi) transcendental arguments, according to which practical reason compels us to read certain signs of hope into past events, "with a frank confirmation bias" (Kleingeld, 2011, p. 175).

In the following century, and perhaps rather more importantly, Martin Luther King espoused the idea that there is something about the trajectory of humanity that makes it—in Theodore Parker's famous words that King was fond of—"bend" towards justice:

> I do not pretend to understand the moral universe; the arc is a long one, my eye reaches but little ways; I cannot calculate the curve and complete

the figure by the experience of sight; I can divine it by conscience. And from what I see I am sure it bends towards justice.

<div style="text-align: right">(Parker, 1853, Sermon III)</div>

In contemporary philosophy, Peter Singer has toyed with a similar idea. The type of moral progress he focuses on—the expanding circle of moral concern—is at least partly shaped by moral reasoning. Reasoning, however, is in principle boundless and "expansionist" (Singer, 2011b, p. 100). Applying it can take us anywhere. Once custom is subjected to scrutiny, there is no going back.[5]

Note that few people seem to have a problem with *individual* at the level of ontogenetic development teleology (think: Kohlberg's stages). It would be implausible to deny that people develop in certain predictable ways from birth to adulthood, and that the maturation of their moral powers forms part and parcel of this process. Note, also, that many people are still attracted to the claim that "ontogeny recapitulates phylogeny". Together, these two propositions entail phylogenetic teleology. This is not supposed to be an argument for social teleology. What it is supposed to show is merely that historical teleology does not occupy an "alien" place in our overall web of beliefs, but that it coheres quite neatly with various claims that are not considered ridiculous.

Even data monger such as Steven Pinker occasionally can't resist teleological temptations. In his *The Better Angels of Our Nature*, he identifies six trends that have contributed to the impressive decline in violence he documents over millennia, centuries, and decades: the pacification process, the civilizing process, the humanitarian revolution, the long peace, the new peace, and the rights revolutions. But what, one may ask, binds these together, and consolidates their power against our countervailing "inner demons"?

> We should not expect these forces to fall out of a grand unified theory. The declines we seek to explain unfolded over vastly different scales of time and damage: the taming of chronic raiding and feuding, the reduction of vicious interpersonal violence such as cutting off noses, the elimination of cruel practices like human sacrifice, torture-executions, and flogging, the abolition of institutions such as slavery and debt bondage, the falling out of fashion of blood sports and dueling, the eroding of political murder and despotism, the recent decline of wars, pogroms, and genocides, the reduction of violence against women, the decriminalization of homosexuality, the protection of children and animals. . . . At the same time, all these developments undeniably point in the same direction. It's a good time in history to be a potential victim. One can imagine a historical narrative in which different practices went in different directions: slavery stayed abolished, for example, but parents

decided to bring back savage beatings of their children; or states became increasingly humane to their citizens but more likely to wage war on one another. That hasn't happened. Most practices have moved in the less violent direction, too many to be a coincidence.

<div align="right">(Pinker, 2011, p. 672)</div>

He concludes that the "implication of directionality superimposed on the random walk of ideological fashion, may seem Whiggish and presentist and historically naïve. Yet it is a kind of Whig history that is supported by the facts" (692). What are the facts that contradict it?

A powerful argument for the existence of society-wide moral improvements—that is, moral progress—and against a "mere change" (Kitcher, 2017, p. 173ff.) account of social development is that many practices that used to be widespread but have meanwhile been abolished have become "beyond the pale". It seems inconceivable for, say, Denmark to decide to try slavery again, or roll back women's suffrage. Indeed, it is comical to even *entertain* these possibilities. Why is this? One explanation is that as societies reach certain levels of moral development and maturity, it becomes impossible for them to fall behind those levels (without disintegrating entirely). Rowland Hunt, Member of Parliament in the English House of Commons between 1903 and 1918, had this to say against women's suffrage:

> There are obvious disadvantages about having women in Parliament. I do not know what is going to be done about their hats. How is a poor little man to get on with a couple of women wearing enormous hats in front of him?

Such an argument could and would not be made today. I do not wish to suggest that stupid arguments for bigoted views have completely fallen out of fashion, but some debates really are over and done for, and moral progress seems to be responsible for it.

> Some moral realists take the apparent irreversibility of certain moral advances to show that moral facts play a causally guiding role in social change. For why, if slavery or the disenfranchisement of women were merely inexpedient, were there no attempts for their reimposition? Cohen (1997) argues that the "injustice of a social arrangement limits its viability. . . . Social arrangements better able to elicit voluntary cooperation have both moral and practical advantages over their more coercive counterparts".
>
> <div align="right">(93)</div>

There is something deeply right about this idea. *Res nolunt diu male administrari*: things refuse to be mismanaged for long.

Notes

1. This is a recurrent theme in the literature on moral progress. Here is Martha Nussbaum:

> We should probably abandon the nineteenth-century expectation for a steady progress of humanity toward greater and greater overall moral achievement. The wars of the twentieth century extinguished that teleological expectation, and the twenty-first, so far, gives us no reason to revive it.
>
> (Nussbaum, 2007, p. 939)

And John Gray:

> The core of the belief in progress is that human values and goals converge in parallel with our increasing knowledge. The twentieth century shows the contrary. Human beings use the power of scientific knowledge to assert and defend the values and goals they already have. New technologies can be used to alleviate suffering and enhance freedom. They can, and will, also be used to wage war and strengthen tyranny. Science made possible the technologies that powered the industrial revolution. In the twentieth century, these technologies were used to implement state terror and genocide on an unprecedented scale.
>
> (Gray, 2004, p. 106)

2. Modern teleologists are mostly content with bastardized versions of the real thing. A recent volume dedicated exclusively to the problem (and to a clumsy nominal style of writing) starts out by saying that it wants to

> casts doubt on the idea that a single, if powerful, conception of a directional movement of peoples and societies over time could function as the unifying principle of all modern historicity; instead, the essays here seek to track the plurality of modern historicities and their underlying structures.
>
> (Trüper, H. Chakrabarty, D. & Subrahmanyam, S. 2015, p. xi)

And, indeed, none of the contributions to that volume take it upon them to seriously defend a teleological account of history at all (see also Carr, 2017).
3. Even comparative optimists like Ronald Inglehart agree with this point:

> Progress is not inevitable. Socioeconomic development brings massive and roughly predictable cultural changes, but if economic collapse occurs, cultural changes start to move in the opposite direction. Development has been the dominant trend of recent centuries: most countries are far more prosperous today than they were two hundred years ago. But this rising long-term trend shows numerous fluctuations.
>
> (Inglehart, 2018, p. 44)

4. For this useful analogy, see Sen (2009, p. 15f.):

> If a theory of justice is to guide reasoned choice of policies, strategies or institutions, then the identification of fully just social arrangements is neither necessary nor sufficient. To illustrate, if we are trying to choose between a Picasso and a Dali, it is of no help to invoke a diagnosis (even if such a transcendental diagnosis could be made) that the ideal picture in the world is the *Mona Lisa*. That may be interesting to hear, but it is neither here nor there in

the choice between a Dali and a Picasso. Indeed, it is not at all necessary to talk about what may be the greatest or most perfect picture in the world, to choose between the two alternatives that we are facing. Nor is it sufficient, or indeed of any particular help, to know that the *Mona Lisa* is the most perfect picture in the world when the choice is actually between a Dali and a Picasso.

5. Among those who are attracted to a teleological account of nature, teleology about the products of culture should raise even fewer eyebrows. In his much-maligned critique of materialist Neo-Darwinism, Thomas Nagel asserts:

> [T]he development of value and moral understanding, like the development of knowledge and reason and the development of consciousness that underlies both of those higher-order functions, forms part of what a general conception of the cosmos must explain. As I have said, the process seems to be one of the universe gradually waking up.
>
> <div align="right">(Nagel, 2012, p. 117)</div>

Non-teleological theories, he argues, are simply not up to this task:

> If we recall the three potential types of historical explanation—causal, teleological, and intentional—it is hard to see how a causal explanation would be possible. Even if there were a partly reductive answer to the constitutive question about the existence of value—if the value of an experience of pleasure were constituted, for example, by the combined value of it protomental parts—that doesn't lead anywhere with regard to the historical question. It is difficult to imagine what form of psychophysical monism could make possible a reductive historical explanation of the origin of life, the development of conscious life, and the appearance of practical reason that would make it anything other than a complete accident that what we care about has objective value. By contrast, once we recognize that an explanation of the appearance and development of life must at the same time be an explanation of the appearance and development of value, a teleological explanation comes to seem more eligible. This would mean that what explains the appearance of life is in part the fact that life is a necessary condition of the instantiation of value, and ultimately of its recognition. I will again set aside the hypothesis of an intentional explanation, even though it, too, could meet this condition. That leaves teleology.
>
> <div align="right">(117)</div>

4 Looking Forward
Towards Teleology 2.0

Introduction

The account of teleology developed in this book is supposed to be moderate and naturalistically credible. My main thesis is that social change is biased for moral improvement, and that there are defeasible mechanisms in place that reliably push societies in an ameliorative direction. A naturalized form of teleology operates without goals. Moral progress has a *direction* but no endpoint. That direction is morally non-arbitrary, but it's not guided by an objective spirit, either.

This is not the best of all possible worlds. Nor is it the only good one: for instance, there is evidence that hunter-gatherer life was, if not exactly arcadia, then at least pretty tolerable, and for the vast majority of individuals almost certainly preferable to life in early civilizations with its drudgery, domination and disease (Widerquist & McCall, 2015). But I will maintain that some people in some places right now inhabit one of the better compromises available, and that most other people in most other places will, in the long run, likely converge on something that is reasonably similar to this compromise.

If someone told you that they wanted to argue for a version of moral teleology, you would have some questions. What do we need to vindicate the teleological thesis? It seems to me that, at the very least, we need four things:

1. The case for *regress* should not be comparatively stronger than the case for progress.
2. Our assessments of progress should not be *epistemically unreliable*.
3. Teleology should be empirically supported: there should be empirical evidence that progress is actually *happening*, and that it is happening robustly and not just haphazardly.
4. We should be able to sketch some sort of broad *explanation* for why social change is biased for moral improvement. This explanation should also support the expectation that progress is likely to continue.

If one succeeded in showing that progress is actually happening, that the amount of progress that is happening is not counteracted by an even greater amount of regress happening elsewhere on the socio-moral balance sheet, that we have no particular reason for thinking that we would be epistemically inept at identifying progress, and that we have an idea for why all of this progress should be occurring at all, we would be fairly close to showing that social change is indeed headed for further moral gains.

4.1   The Case for Decline

Modernity was, very roughly speaking, supposed to bring us three things: freedom, equality, and happiness. Freedom was supposed to flow from increased political participation, expanded opportunities, and withering oppression. Equality was supposed to come from more inclusive institutions and reduced discrimination. Happiness was supposed to emerge from a combination of the two: if more people are put in a position to do more of the things they want because they have been emancipated from previously existing obstacles, their lives will improve.

This has indeed happened to some extent. In recent years, however, a number of authors have tried to highlight the Faustian bargain we may have made, stressing how the socio-political revolutions and institutional upheavals of modernity—especially *late* modernity—have begun to slowly undermine the promise that made us want to embark on the project of modernity in the first place.

The first problem is that modern societies have created a kind of "really existing nihilism", an exhaustion of their evaluative energies that has led to precipitous cultural and economic decadence. Modern capitalism has created jobs that are objectively and subjectively pointless: they accomplish nothing or do more harm than good. The people performing them are perversely aware of this, and of how spending their lives as sycophantic flunkies, useless bureaucrats, and time-killing administrators slowly suffocates their souls (Graeber, 2018). The economy is slowing down (Cowen, 2011); people have fewer babies; society is torn apart by distrust and polarization, even outright hatred; mainstream culture creates sequels and remakes, pastiches, parodies, and reboots (Douthat, 2020); and our political institutions are in a constant state of irreparable gridlock, held together only by an increasing number of barely functional kludges. Taken together, this state of affairs creates the distinct impression of lateness and decay, the dawn of a shriveling civilization.

Whatever material gains have been created by this shiny hollowness are unequally distributed, by which I mean. . . . As a result, this threatens to tear the social fabric apart. We have—perhaps inadvertently—fashioned a society in which people are sequestered into highly isolated socioeconomic

bubbles. They don't speak to, nor would they increasingly be able to understand, each other. The beneficiaries of modern meritocracy are a narrow "cognitive elite" that has been able to channel most of the perks of current society towards its own ilk, and to secure the same privileges for their heirs via the iron cage of assortative mating, lubricated by exclusive Ivy league schooling. This has undermined social solidarity and created a morally deficient underclass that revels in self-destructive behavior and has lost all sense of virtue. Honesty, industriousness, marriage, and religiosity—the main pillars of American success thus far—have become quaint words whose real meaning will soon be forgotten (Murray, 2012).

Modernity is eating away at its own foundations. The enlightenment was supposed to liberate humanity from the triple yoke of superstition, oppression, and poverty. But instead of unleashing peace, creativity, and self-realization, modern society has created crippling isolation, maddening monotonousness, destructive egoism, ferocious conformity, and technological alienation. Only a rediscovery of communal identity, civic participation, and genuine cultural diversity can save us from this carnage.

Or so the argument goes. It's worth emphasizing that the case for decline is often based not on evidence for actual decline, but merely for slower growth. This means that much of the debate here is about a choice between progress and somewhat decelerated progress. Second, for all of the complaints about contemporary materialism and the hollow cultural nihilism it has nurtured, contemporary declinists tend to ignore many of the most significant *non-materialist* advances that have been made. They complain about materialist attitudes, but their yardstick of progress—or regress— often still comes down to "where are those flying cars we were promised?" and "why can't I vacation on Mars yet?", which are actually *ultramaterialist* complaints. What is rarely brought up is that, while it may be true that wealthy nations are on a less impressive economic slope now, recent decades have seen people make *vast* changes in many of their central moral attitudes. Tolerance, freedom, aversion to inequality, and the protection of minorities have all become central concerns, and the idea that modern societies haven't made tremendous social progress in terms of the evaluative atmosphere that we breathe is so difficult to take seriously that not even the most ardent declinists want to take up that argument.

Meanwhile, the alternative on offer, at least if we look at it without the conservative's romanticism-colored glasses, is deeply unappealing. What are the anti-modernists' economic and political proposals? Patrick Deneen, for instance, recommends a return to what he refers to as "household economics", that is

> economic habits that are developed to support the flourishing of households but which in turn seek to transform the household into a small

economy. Utility and ease must be rejected in preference to practices of local knowledge and virtuosity. The ability to do and make things for oneself—to provision one's household through the work of one's own and one's children's hands—should be prized above consumption and waste.

(Deneen, 2018, p. 193ff.)

I am inclined to reply: *your* children's hands first. Further: new

practices will be developed that will benefit from the openness of liberal society. They will be regarded as "options" within the liberal frame, and while suspect in the broader culture, largely permitted to exist so long as they are nonthreatening to the liberal order's main business. Yet it is likely from the lessons learned within these communities that a viable postliberal political theory will arise, one that begins with fundamentally different anthropological assumptions not arising from a supposed state of nature or concluding with a world-straddling state and market, but instead building on the fact of human relationality, sociability, and the learned ability to sacrifice one's narrow personal interest not to abstract humanity, but for the sake of other humans. With the demise of the liberal order, such countercultures will come to be seen not as options but as necessities.

(Deneen, 2018, p. 193ff.)

There is a sinister, almost postapocalyptic atmosphere in such visions of bucolic simplicity, where children are conscripted into the workforce and misfit communities of zealous weirdos become the dominant form of social organization.

The claim that earlier times didn't suffer from comparably rampant inequality sometimes borders on the obtuse, and indeed comical. In the 1960s, we are told that inequality was of an altogether different type:

Take, for example, the woman who was the embodiment of the different world of the rich, Marjorie Merriweather Post. Heiress to the founder of the company that became General Foods, one of the wealthiest women in America, she owned palatial homes in Washington, Palm Beach, and on Long Island, furnished with antiques and objects from the castles of Europe. . . . Hers was not a life familiar to many other Americans. But, with trivial exceptions, it was different only in the things that money could buy. When her guests assembled for dinner, the men wore black tie, a footman stood behind every chair, the silver was sterling, and the china had gold leaf. But the soup was likely to be beef consommé, the main course was almost always roast beef, steak, lamb chops, or broiled chicken, the starch was almost certainly potato, and the vegetable was

likely to be broccoli au gratin. The books on the shelves of her librar-
ies were a run-of-the-mill mix of popular fiction and nonfiction. She
screened the latest films in the privacy of her homes, but the films her
guests watched were standard Hollywood products. The wealthy had
only a very few pastimes—polo and foxhunting are the only two I can
think of, and they engaged only a fraction of the rich—that were differ-
ent from pastimes in the rest of America.

<div align="right">(Murray, 2012, p. 20f.)</div>

At the end of the day, such diagnoses seem to be more reflective of the fact
that we are vastly less tolerant of social inequalities than we used to be—
which is a good thing—rather than that we witness historically extreme
forms of inequality—which may be a bad thing. I will leave it to the reader
to imagine how impressed those servants and footmen would have been
by the fundamental equality established between them and the foxhunting,
yacht-sailing upper class merely by their common appreciation of pota-
toes and broccoli. If anything, the post-WWII era was a highly unusual
and exceptional time, rather than an egalitarian default state we are only
now departing from. At least for the past 10,000 years, all (large) societies
everywhere have been fiercely unequal. This is not, of course, a defense
of inequality, but merely a rejection of the idea that modern inequality is
somehow an altogether novel threat.

Another common refrain is that we have traded material riches for our
mental health. Many people, after they have been forced, by the uncoercive
coercion of the available evidence, to admit that life has improved along
many lines (life expectancy, wealth, etc.), resort to the claim that we have
nevertheless paid a steep price for this progress, and that the trade-off may
not have been worth it. The idea is that even though we may be materially
better off, we are psychologically worse off, as evidence for increasing rates
of mental illness such as depression or anxiety shows.

The main problem with this line of argument is that this mental health
crisis does not exist. In fact, the evidence strongly suggests that mental ill-
ness is *not* on the rise. Rather, what seems to have happened is that accurate
diagnoses of mental illness such as depression have increased, that effective
medication with psychopharmacological drugs is now more widespread,
and that people are more open about their psychological struggles. People
were always depressed. The difference is that we increasingly diagnose men-
tal illness correctly, treat it properly, and no longer stigmatize it as harshly
(Dornes, 2016; Richter et al., 2008; Richter & Berger, 2013; Baxter et al.,
2014). A full appreciation of the evidence regarding depression and anxiety
thus shows that here, too, there is more progress than regress.

Finally, a powerful intuition against progress is that while it may be real,
it is irrevocably *tainted*. Here, the idea seems to be that the progress made

by some people in some places would not have been possible without the violent expropriation of other people in other places. Since the progress thus achieved has been facilitated by unspeakable horrors and unjustifiable wrongs, we cannot speak of genuine progress. The main examples here are colonialism and slavery. If the modern world of affluence and opportunity was built on colonial looting and the pillage of black and brown bodies, progressive triumphalism is a lost cause (for one example, see Beckert, 2014). However, it just doesn't seem to be true that colonialism or slavery were necessary, or even *helpful*, in bringing about the modern world (Acemoglu & Robinson, 2013; Deaton, 2013; Brennan, 2020). It is true, of course, that the modern world has a horrendous history of genocide, exploitation, and plunder (as does the premodern world). But the evidence suggests that these things were neither required nor even contingently causally responsible for jumpstarting modern progress. In many cases, colonialism and imperialism were downright inimical to progress, merely enriching social elites in the colonizing countries as well as the colonies, at the expense of poorer people in *both* places. The worst colonizers were not always the richest countries, and many countries became wealthy without any imperial plunder at all. Likewise, it is true that slavery did make some people rich. But in addition to its immorality, slavery was also harmful as an economic arrangement, and the regions in North or South America that most extensively relied on slave labor neither were, nor are, the globally or domestically richest places.

I am not trying to dismiss the case against progress wholesale. Just like every cloud has a silver lining, every improvement has its dark side. It would be naïve to suggest that the recent trajectory of social change had no costs and only benefits. However, I wish to insist that a sober look at the evidence does not vindicate declinist conclusions in any straightforward way.

4.2   Debunking Teleology? Anti-Narrativism

One may still wonder: are we even in a position to know whether or not society is improving? Are our assessments of the relative state of moral development likely to be accurate?

An important reason *not* to take teleology seriously would be that we would find it compelling even if it weren't true. Probably the most fundamental and significant recent challenge to historical teleology comes from Alex Rosenberg's naturalistic anti-narrativism (Rosenberg, 2019).

One way of framing what teleology amounts to is to identify it with the claim that (human) history is headed in a certain direction, that is, that we can discern a certain narrative structure in patterns of historical change. Alex Rosenberg advances an error theory of teleology. Thinking about

history in terms of *goals* and *direction* is both extremely common and systematically mistaken:

> Teleological "thinking" about—divining purposes in—the past was narrative history's bane for as long as it was its raison d'être. Darwin banished purpose from biology just as rigorously for humans as he did for other animals. But no one noticed. Academic historians sought with great success to drive teleology out of their discipline, though not for the right reasons. They rejected the notion that history was going anywhere because they rejected the Christian, Muslim, Marxian, capitalist, racial, patriarchal, and nationalist eschatologies that identified history's end, goal, or purpose. . . . The search for meaning in particular episodes, eras, or epochs in national narratives is driven by this teleological mistake.
>
> <div align="right">(Rosenberg, 2019, p. 248)</div>

Rosenberg argues that teleological thinking was abandoned, but not radically enough. To the extent that it has been given up, it was for reasons of substantive disagreement not over whether history had any goals, but over *which* goal it had. All particular suggestions were weighed and found wanting. However, no principled reason was given for why the quest for teleology as such is nonsense—or so Rosenberg complains.

The main explanation for our teleological delusion is that our minds simply don't work according to the "boxological" framework of folk psychology (Rosenberg, 2019, p. 165). Folk psychology, according to Rosenberg, has it that the mind consists of something akin to containers—beliefs and desires being the most important ones—whose content is determined by perceptual, inferential, and somatic mechanisms and which, if shaken correctly, produce behavior. He argues that this basic picture is undermined by recent neuroscientific discoveries. Since this basic picture is incorrect, the account of history that is based on it must be incorrect as well.

Why did Talleyrand betray Napoleon? Your answer, if you can think of one, is probably thoroughly off track:

> It turns out that none of the biographers got it right. None identified the pairings of beliefs and desires that moved Talleyrand into treasonable conversation with the emperor's enemies. The reason is not that there was another set of beliefs and desires Talleyrand had, which no one discerned, not even Metternich, the close student of Talleyrand's machinations. The reason is that there were no beliefs and desires anywhere inside Talleyrand's mind as he went through the process of deciding. . . . The real inside story is that there was no story.
>
> <div align="right">(159)</div>

Thus, one of the main recent challenges to teleology is based on a debunking argument: we are wired to crave stories that make sense, but the world just doesn't work that way. The anti-narrativist debunking argument against teleology goes something like this:

(1) We are psychologically wired to think about the world in terms of stories.
(2) Neuroscientific evidence suggests that the mechanisms that make us construct stories are not truth-tracking. Therefore,
(3) We are unjustified in believing that the world has any kind of narrative structure.

But there are several problems with this argument. For one thing, a radical rejection of belief/desire psychology leaves us unable to account for differences in plausibility between competing mentalistic explanations of the events: "Talleyrand acted the way he did because of his strategic and political goals" is at the very least *more plausible* than an absurd statement such as "Talleyrand acted the way he did because he hated fish and thought the name Napoléon was a fish".

However, even if we grant the claim that teleological explanations of historical events, that is, explanations of historical developments couched in terms of a belief/desire folk psychology, are systematically mistaken, this does not bear on the question whether there are other, *non-psychological* mechanisms in place that make societies gravitate towards moral improvement. In his discussion of Jared Diamond's work (Rosenberg, 2019, pp. 219–241), Rosenberg does admit to the possibility that there are some systematic trends in history. It's just that these trends, whatever they are, are not brought about by people's beliefs and desires (because these things don't exist). And there is nothing else *in the anti-narrativist argument* that precludes the possibility of directional, collective moral improvement over time. The falsity of folk psychology at the lower level does not undermine the possibility of moral teleology at the higher level of large-scale social change.

If anything, the psychological deck is more strongly stacked *against* believing in teleological progress than it is biased in its favor. There are various powerful cognitive mechanisms that make us wary of the future and incline us towards a considerable degree of pessimism about the prospects of social improvement. It is the anti-teleological attitude that is the comparatively more debunkable one.

Some of these mechanisms are psychological in nature. Famously, there is an asymmetry in how adaptive it is to tolerate false positives and false negatives, respectively, in our assessments of risks and danger. To see a threat when there isn't one—a false negative—is typically only a minor cost; to overlook an objectively real threat can be disastrous. Therefore, our evolutionary, and

indeed individual, learning history made us overly aversive to bad developments and events. This effect is further compounded by the availability heuristic: we tend to estimate the probability of an event, which is of course difficult to gauge directly, by relying on examples that we find easy to recall. A gruesome child murder in the news, for instance, may stick in our memory forever. This can severely bias our estimates of how common child murders are. Sensationalistic media reporting caters to this disposition.

There are at least two further psychological mechanisms that exacerbate this tendency. The hedonic treadmill makes sure that we get used to the good things, which makes it hard to consciously appreciate just how well off one is (Brickman & Campbell, 1971). Status quo bias creates an irrational partiality towards whatever happens to be the current state of affairs, thus further nurturing a skeptical attitude towards the potential of the future (Bostrom & Ord, 2006).

Second, there are various epistemic mechanisms that act as distorting influences on our assessments of progress. One could say that good things and bad things belong to two fundamentally different ontological categories: good things are (typically) *developments*, while bad things are (typically) *events*. Since developments are often gradual, they can be imperceptible; events, which are often sudden, can be hard to miss. One cannot overlook the 3,000 victims that died on 9/11. But we need boring data and counterintuitive statistics to appreciate the millions (and indeed billions) of people that have been lifted out of extreme poverty over the past decades and centuries. Progress is thus intrinsically hard to perceive.

The reminiscence bump (Jansari & Parkin, 1996) makes sure that we don't remember all times equally; as it happens, people's most vivid memories tend to be from the time that also happens to be their best time, that is, the phase between 15 and 30. This is, incidentally, the time when people go to their first parties, have their first sex, go to university, and start a career. This basically guarantees that the past seems more appealing to most people than the present and future. Finally, the optimism gap (Whitman, 1998) is the phenomenon than people's assessments of their own life prospects are much more optimistic than their perspective on society at large. Even if they think that *they* will be fine, they think that the rest of society is doomed. This is, of course, incoherent, just as not everyone can be an above-average driver.

Third, assessments of moral progress are *themselves* subject to moral evaluation. The claim that society is improving may, to some, appear to constitute a morally odious form of acquiescence towards the objectionable status quo. Admitting that there has been progress, and that it is more likely than not to continue, can seem to license inaction and passivity. If things will improve anyway, why not sit back and enjoy the ride? Finally, to concede that there has been progress of various kinds can make one seem insensitive to the plight of those left behind. These effects mutually

reinforce and stabilize each other, making the perception of societal decline psychologically even more irresistible.

My debunking explanation of anti-teleological pessimism draws on cognitive processes which, due to their evolved or culturally widespread character, should be expected to be (more or less) universal tendencies. This, however, seems unable to explain the fact that there seem to be trends of pessimistic and optimistic thinking: some societies or eras seem to be more prone to pessimism, some less. But if the pessimistic dispositions I identified are psychological universals, why do we observe such fluctuations at all? First of all, not all mechanisms I mentioned are psychologically hard-wired (think of the role of the media, which may be subject to variation). Second, conspiratorial thinking can serve as an apt analogy here: conspiracy theories are based on an unholy alliance between certain personality traits, motivated reasoning and confirmation bias, all of which are dispositions universally found in all humans or, at least, tendencies whose average distribution in society should not be subject to much variation. How come, then, then conspiracies are sometimes more, sometimes less *en vogue*? These trends, it seems to me, are best explained in terms of the changing social conditions that allow certain universal traits to be more or less strongly expressed. For instance, when there is a high degree of social volatility and economic insecurity, people may become more likely to cling to conspiratorial thinking. Under certain circumstances, then, humans, though universally attracted towards declinism at all times, will be subject to waves of pessimism and (relative) optimism.

However, the evidence that people are, in general, excessively pessimistic is very strong. The vast majority of people is unaware of decreases in child mortality, extreme poverty, or illiteracy, and falsely believes bad things such as crime to be on the rise. As a result, almost no one believes that the world is getting better.[1] Tellingly, it is fair to say that the better off people are, the less likely they are to think that the world is getting better. For instance, only 4% of people in Germany and 3% of Australians believe that the world is improving, compared to 23% in Indonesia or 41% in China, and old people everywhere inevitably believe that the generation succeeding them consists of a bunch of spoiled, ungrateful brats (Protzko & Schooler, 2019). Who's right?

4.3   The New Optimism: Empirical Evidence for Progress

There has been a recent wave of publications that all make roughly the same point: the state of the world is much better than you think it is, and it is likely to get better still. Perhaps the most visible contribution to this genre is Steven Pinker's *Enlightenment Now*. But there are numerous other authors milking the same cow: a sober look at the facts and the evidence

shows that the world is improving in almost all important respects. Some focus on poverty and inequality (Deaton, 2013); some on risks and existential threats (Cohen & Zenko, 2019); some on the prospects of progressive politics (Teixeira, 2017); some on the benefits of freedom (Norberg, 2016; Norberg, 2020); some on health and human development (Rosling, 2018); some on declines in violence and war (Pinker, 2011); some on the benefits of innovation and enterprise (Ridley, 2011); some on the essential goodness of human nature (Bregman, 2020); some on our unhealthy preoccupation with what goes wrong rather than right (Duffy, 2018); some on the state of moral development (Shermer, 2015); and some, finally, try to develop an overall assessment of the condition of the modern world based on most or all of those aspects (Easterbrook, 2018; Schröder, 2018; Boudry, 2019).

I won't rehearse all the evidence and empirical details of this trend here. But I want to highlight some of the main points and explore some of their most important moral and philosophical implications. The philosophical reception of this research so far desperately stands in need of improvement: most reactions to what I like to refer to as the "new optimism" have been either overly enthusiastic or downright dismissive (Goldin, 2018).[2] I aim to take the available data seriously without ignoring relevant blind spots or criticisms.

One final caveat: In what follows, I will try to provide a very quick summary of the case for optimism. I will necessarily have to gloss over many complications and ignore many details of this debate. Moreover, claims such as "X has improved" or "Y has become better" are almost necessarily lacking in nuance, and present an overly crude and strictly speaking false picture. It is usually correct to emphasize that such general statements of progress are rarely true for literally everyone literally everywhere.

There are several key themes that keep coming up in the new optimism. One is safety. Humans, like all animals, are fragile creatures, vulnerable to disease, injury, abuse, and death. It is hard to dispute the fact that our ability to accurately assess risks is seriously flawed. Almost as hard to dispute is the fact that the modern world has successfully eliminated, mitigated, or neutralized various threats that used to be taken for granted. Various horrendous diseases have been eradicated, can be prevented, or can at least be effectively treated. Various random and seemingly trivial risks, such as being struck by lightning, or dying an accidental death at home, have been greatly reduced. These days, people rarely get mauled by wolves or bears. Much of this is difficult to appreciate or even to become aware of. Other risks, however, are greatly exaggerated, and occupy an excessive amount of space both in the perception of the public and in individual worrying. Terrorism is perhaps the clearest example: the dangers of politically motivated acts of violence are in fact negligible, yet terrorism, particularly

of non-domestic origin, remains a news staple. Our lives are, on average, much safer than they ever were.

Another is health. For most of history, human average life expectancy was horrendously low, hovering around 30–35 years. Much of this is attributable to humans' extremely high rate of infant mortality (the number of individuals reaching the age of five or older). The list of ghastly diseases that used to bedevil humanity is long: diphtheria, smallpox, tuberculosis, cholera, polio, influenza, malaria, and the plague. These afflictions used to be deadly scourges, and they still are, to some extent, in some regions. But improved general health, effective prevention, and therapy have, in many places at least, solved the problem for many people. One major task for the future of humanity is to accomplish this for all of humanity. Improved nutrition, access to clean water, sanitation, and antibiotics have doubled life expectancy. Overall, people are much healthier than they ever were.

Wealth is another important aspect of progress. I cannot even begin to summarize the history of economic growth here (see, for instance, Clark, 2007). One helpful proxy, however, that can illustrate (quite literally) how strikingly different modern life is when compared to life 200, 2,000, or 20,000 years ago is the history of light (Nordhaus, 1996). Access to and consumption of (artificial) light have increased by several orders of magnitude from open fires and prehistoric fat lights over tallow candles and whale oil lamps to fluorescent light bulbs. But there is not merely the quantitative question of how much cheaper light has become, and how inadequately traditional price indices measure rising living standards. Rather, the history of light serves as a case study for how drastically conventional economic measures underestimate the explosion of wealth more generally. That we have become wealthier doesn't mean that we can afford more of the same, as if to be wealthy today meant, in contrast to 17th-century Europe, being able to afford not one, but one billion whale oil lamps. It means that we have access to entirely new categories of things, things that solve previously unheard-of problems and cater to previously unimagined wishes. It's not that, centuries ago, only the richest of the rich could afford a car, a television, or a flu shot, while hardly anyone else could, whereas today, the number of people who can afford those things has grown. Rather, these *types* of things did not yet exist, and in many cases, *had not even been imagined by anyone yet*.

The issue if inequality is intimately tied to the issue of wealth. Modern inequality, for all its ills, is as much a cause of social problems as it is a symptom of their (partial) solution. It is easy to maintain equality when everyone is poor. In reality, of course, even in the desperately poor societies that existed everywhere before the Industrial Revolution, there has always been a very small group of people that was comparatively vastly richer than the rest. But those aristocratic or clerical elites rarely became wealthy

due to their productivity; rather, they had devised ways of extracting whatever meagre wealth was generated by the powerless rest of society to line their own pockets with it. Moreover, even though said elites were vastly richer than the rest, they were, by today's standards, still extremely poor, simply because—despite the monetary income and labor power they were in a position to command—the vast majority of the things that define our wealth today didn't even exist at the time.

How much money would I have to offer you to travel back to and live in the 16th century? Chances are there is no amount that could compensate you for the absence of refrigerators, dental care, Netflix, or safe travel. Thus, while the "great divergence" ushered in by the Industrial Revolution has made some countries as much as 50 times wealthier than others, it has also created new forms of staggering inequality both within and between societies. Inequality remains a comparative measure, however, so the fact that we live in a more unequal world does not mean that the poorest in society are not vastly better off as well.

This wealth has made people happier. That money doesn't make you happy is one of those truisms that has been empirically investigated and found not to be true at all (Schröder, 2018, p. 178). It is of course correct to observe that there are some miserable rich people. But it is equally true that individual counterexamples don't disconfirm averages, and the averages are clear: while there are some places (mostly in South America) that enjoy comparatively high average levels of individual happiness without comparatively high levels of material riches, people's average level of life satisfaction is highest in affluent countries, and lowest in poor places. Indeed, there are no rich countries, but plenty of poor ones, where people are, on average, unhappy. Money is sufficient for happiness (even if it's not necessary).

Some may think that, as far as the issue of moral progress is concerned, this is all neither here nor there. Sure, this objection goes, that many people are living healthier, more prosperous lives is worthy of our approbation. It is desirable, from a moral point of view, that significant parts of the world are becoming a better place. But moral progress is an entirely separate issue: here, we are not talking about morally welcome developments but about developments in and of our morality. Moral progress, in short, is about moral improvement, not just morally laudable improvement. I have argued against such a narrow notion of moral progress. But even for those who insist that genuine moral progress is different from merely morally desirable social progress, there is plenty of evidence suggesting that such narrowly *moral* progress is happening as well.

I will focus on two trends: one concerns improvements in the prevalence of political participation both within countries and globally speaking. These improvements allow people to reduce domination and oppression, while providing the opportunity to develop their autonomy in deciding

about the rules they want to live by. The other concerns the corresponding changes in people's values, and the way they reflect a move away from a focus on survival and hardship towards values embracing liberty, tolerance, and authenticity.

First is political empowerment. The simple fact is that, not too long ago, there were no democracies, whereas now, there are many. Proto-democracies such as Ancient Greece were on the right track; ultimately, however, they don't really qualify because of their highly exclusive criteria of access to the franchise: in Athens, only male adult citizens were granted democratic participation. Women, children, slaves, and *métoikoi* were left out.

This has changed significantly, especially over the last decades. The Polity IV index assigns a democracy score to regimes that ranges from –10 (completely undemocratic) to 10 (consolidated democracy).[3] By subtracting the autocracy score from the democracy score, one arrives at a final value. Proper democracies score 5 or higher, autocracies –5 or lower. Everything in between is classified as an "anocratic" government. The features of political regimes that these values latch onto mostly have to do with how competitive the selection process for governmental responsibilities is, how free elections are, how transparent and serious the constraints on executive action are, and so on. There is strong evidence that the number of democracies internationally is on the rise, and that the quality of democracies internally is increasing. There are authoritarian setbacks, of course, but the long-term trends unequivocally point in the direction of further democratization.

There is strong evidence that this rise of democracy is causally attributable to changing moral values. The causal direction seems to "flow" from an increasing adoption of modern evaluative attitudes towards increased demand for political participation, enfranchisement, and governmental accountability. Political freedom goes hand in hand with civil rights. Racial segregation was legally entrenched in the USA until civil and voting rights were expanded to racial minorities in 1965. Of course, racist attitudes and practices are often not fully abolished, but merely channeled into new institutions (Alexander, 2020).

Meanwhile, lip service to democracy is already near perfect. As of today, almost all countries grant all of their adult citizens the *pro forma* right to vote. Only Saudi Arabia, Somalia, and the United Arab Emirates are exceptions. Late adoption of a fully inclusive franchise is widely considered scandalous: Switzerland, to its great embarrassment, only adopted female suffrage in 1971 (at least nation-wide). The public commitment to democracy, however brittle it may be in practice, often works as an ideal standard that oppressive regimes are held against.

But what about people's *values*? It would cast doubt on the idea of moral progress if, as people's lives and living conditions are improving, their values remained stagnant. Fortunately, this is not what we find. Instead, we

see a consistent trend towards more progressive, liberal, and emancipative values, both locally within cultures and globally between them.

The authoritative resource on the fluctuation of moral and cultural attitudes is the *World Values Survey*.[4] It systematically keeps track of changes in people's evaluative orientations since 1981, both between countries (or regions) and generations. It distinguishes two basic dimensions of value change: one axis concerns the development from so-called *traditional* to *secular-rational* values. Paradigmatic items on the survey are not only about the value of hard work and a safe job, the necessity of social order and stability, the importance of faith and respect for parents or elders, and toughness on crime, but also skepticism towards female emancipation or gay rights. A secular-rational orientation displays the opposite direction. The second axis concerns *survival* values versus values of *self-expression*. Here, the survey asks people about their trust in foreigners, the importance of freedom and political participation, the significance of tolerance towards and inclusion of minorities, or the relative importance of physical and economic security over individual freedom or authenticity. The evidence clearly shows that people's attitude changes in the direction of secular-rational/self-expression values.

The obtained values are highly correlated with levels of economic development and individual happiness. Importantly, these patterns of moral change do not merely reflect differences between younger and older people. If that were so, we there would be no lasting change because the young people who are postmaterialist now would become just as traditionalist and survival-oriented as older individuals over time. That this is not the case can be seen because the survey tracks people's attitudes over time by clustering them into generational cohorts (Welzel, 2013). This shows that each subsequent cohort is more progressive, in absolute terms, than the previous one, so that the normative effect of youthful progressivism doesn't evaporate as those young individuals grow older.

Norms surrounding gender equality, fertility, and attitudes towards atypical sexual orientations provide a useful specific test case. People have become much more supportive of women's rights and much more tolerant of alternative sexual lifestyles. At the same time, people are now much less likely to endorse anti-black racism, antisemitism, or other forms of xenophobic prejudice.

The world is becoming an increasingly more hospitable place. But the new optimism is not without its own shortcomings and inaccuracies. One massive blind spot in the new optimism is animal rights. As far as I am aware, none of the aforementioned contributions to the genre devote appropriate attention to the fact that the scale of animal exploitation and cruelty in contemporary society is historically unprecedented. Many, such as Pinker, do not even mention this fact at all. Human treatment

of non-human animals is an obvious case of tremendous moral regress. Depending on how one gauges the amount of suffering inflicted on animals in factory farming, the moral losses may even eat up all of the gains made elsewhere, in the realm of human life (Huemer, 2019).[5] Depending on what we think about the moral status of various animals, and depending on the scale of their mistreatment, the case for overall progress fails.

What about declines in poverty? Critics of the new optimism like to point out that whether, and if so to what extent, global poverty has been in decline depends on the numbers one uses, the criteria one applies, and the region one looks at. All of these are reasonable objections that deserve to be taken seriously. In the end, however, they don't disconfirm the fact that poverty has been in decline, and indeed impressively so. Hickel (2016), for instance, argues that what he refers to as the "good news" narrative rests on a tendentious presentation of the data. One often hears that the number of people in extreme poverty has been in steady decline for decades; extreme poverty, here, is understood as living on (roughly) $1/day, adjusted for purchasing power. But for one thing, Hickel argues, almost all of the people that have escaped such dire straits in recent decades have been concentrated in very few places, most prominently China. Many other regions, such as big chunks of Sub-Saharan Africa, have been left out of those gains. Second, he stresses that, while the *share* of people living in extreme poverty may have declined, their *absolute* numbers have been increasing. Today, only 10% of people live in extreme poverty. But those 10% amount to 1 billion people, a number that stands for an intolerable amount of horrendous suffering. Third, depending on where the cut-off is set, poverty hasn't been declining at all. If we set the threshold at $4/day, we would see that 60% rather than 10% of people in the world today are poor.

It is plausible, and important to emphasize, that the numbers used to evaluate the success of poverty reduction programs are always somewhat arbitrary. They are also frequently embellished and gerrymandered to let governments and NGOs look good. But this doesn't mean that the general trend is not real. Note that the objection that almost all of the successes in reducing poverty have been in China is at least somewhat odd. Maybe most of those gains have been concentrated in one place, but . . . so what? It's strange to suggest that the fact that millions of people who don't have to suffer anymore from the harshest circumstances matters more or less, depending on whether those people are in China or scattered across the globe. Moreover, the claim is strongly myopic, for it completely ignores the fact that everyone *in the West* used to be extremely poor, too, whereas now, almost no one is. Even if the real number of poor people today is 60%, then that is, at the very least, an improvement of 40% relative to the 100% of poverty that used to exist before the "Great Escape". For the issue of moral progress, this long-term perspective is the one that counts.

Second, the fact that there are more poor people, in absolute numbers, than before, is largely driven by the fact that there are *more people* than before. It is far from clear whether the absolute numbers or the proportions are more morally relevant. There is something to be said in favor of either approach: it's regress when there are literally more poor people. But it's progress when there are relatively fewer. Finally, it is fair to ponder how the threshold for poverty should be set. Then again, it does not seem normatively implausible to hold that we should prioritize the poorest of the poor. If *their* numbers and/or share go down, this should be considered a comparatively more significant gain. Finally, the problem of poverty poses unique problems since many of the other metrics of progress discussed earlier are not subject to similarly unclear, vague, or potentially arbitrary cutoffs. Infant mortality goes down or it doesn't. Average life expectancy goes up or it doesn't, as do literacy rates or rates or other indices of progress. These facts are not controversial.

Perhaps the biggest concern for the new optimists remains climate change. I cannot satisfactorily address the climate change objection to moral progress here (van der Vossen & Brennan, 2018). Of course, social progress and moral progress don't amount to much if the price we ultimately pay for them is the end of human civilization and the extinction of our species (and many others). However, let me at least sketch how a not entirely pessimistic response to the very serious threat of climate change might look. (i) The very worst effects of climate change, such as a runaway greenhouse effect, are improbable. One should not reject the progress narrative *simply because* of the tail risks of climate change, because these are by definition unlikely to manifest. Now obviously, extremely large but improbable risks should be taken into account. But there is no need for highly unlikely worst case scenarios to dominate the response to global warming and other forms of environmental degradation. (After all, there are some highly unlike best case scenarios to counterbalance those.) According to mainstream predictions, climate change will create huge costs. How high will these costs be? Within the chosen timeframe (typically until 2100), these costs will make people all over the globe worse off than they would have been had no global warming occurred. But this does not mean that they will be worse off than today. In fact, we can expect people to be much *better* off in the future, even with global warming taken into account (Stern, 2006). (ii) Climate change is the flipside of (some of) the positive developments mentioned earlier. The reason why dangerous global warming occurs in the first place is (largely) due to the fact that as people have escaped poverty and destitution, their energy consumption has gone up. Global warming is an unintended side-effect of otherwise desirable social developments. It is not clear how, or indeed whether, these other positive developments could have been accomplished without doing

the things (such as burning fossil fuels) that led to climate change. (iii) Climate change also has positive effects, such as increased habitability or agricultural potential in regions that are currently too cold or too hostile for most significant human and/or non-human activity, or a decrease in deaths from cold weather, a significant problem globally speaking. These positive effects—however unclear their magnitude may be—have to be traded off against the negative ones. (iv) Many of the bad effects of climate change are themselves endogenous to the issue of moral progress: the displacement of people living in coastal regions in the global south is a humanitarian catastrophe. But *how much* of a catastrophe it ends up being depends on how much international cooperation on this problem there will be, and how willing less affected countries are to allow the expected migration to occur safely and peacefully. (v) The richer people become, they more they start caring about the environment, simply because they can afford it. This plays a role in promoting political action and creates demand for a more sustainable economy. This means that the processes, such as economic growth, that generate climate change also play a big role in nurturing the evaluative attitudes needed to curb it. (vi) As countries become richer, their relative levels of pollution and environmental damage go down (the so-called "environmental Kuznets curve", see Cole et al., 1997), mostly because their technological prowess and pollution management become more advanced. (vii) The damage inflicted by climate change is getting worse *in virtue of progress*. This creates the illusion that things are getting worse when, in some cases, they actually do not. That natural disasters such as extreme weather events (storms, fires, or droughts) are worsening is largely due to the fact that we are becoming richer, so that the destruction caused by natural disasters is costlier (Shellenberger, 2020). Rich people have more to lose. (viii) Finally, and perhaps most importantly, the very mechanisms that lead to economic growth, thereby accelerating climate change, are also the mechanisms that tend to make societies more resilient against external threats. Wealthy nations with better institutions, a more advanced infrastructure, and a healthier and better educated populace are better able to cope with and adapt to whatever problems the world throws at them. Climate change is a daunting challenge for humanity. But it is not clear that its bad effects outweigh the gains that modern societies have already made and that developing countries are bound to make in the future.

The recent global pandemic is often instinctively cited as another example to show that humanity hasn't progressed that much. Things are improving? Were you hibernating during the past two years? In fact, the situation is much more complex, and in some ways the opposite is true. When the issue is how much progress humans have made—morally, socially, intellectually, and politically—the real question is not how this pandemic has

been handled, or whether this pandemic has been handled well, but how this pandemic has been handled *in comparison to the great pandemics of the past*. And in light of this question, it is hard to deny that Covid-19 has been handled vastly better than its predecessors. Conservative estimates suggest that the Black Death killed *one third* of the global population; it went on for *hundreds of years*; meanwhile, the world had no knowledge of where it came from, or what caused it, and I mean *absolutely none whatsoever*. Indeed, the discipline of medicine hadn't developed to a stage where anyone could come even remotely close to understanding, in the roughest of terms, the origins of this disease; as far as effective therapies or preventive measure were concerned, there were none—again, *none whatsoever*, if we don't count the creepy beak masks.

Meanwhile, the nature of Covid-19 was thoroughly understood within weeks, and as far as its origins are concerned, there is at least a narrow range of known options. The degree to which an outgroup was blamed for poisoning the wells, which inevitably used to happen whenever people were harassed by another inexplicable (to them) catastrophe, was relatively minor. Effective preventative and therapeutic measures were readily and widely available, and rapidly implemented, however imperfectly. The fatality rate of the disease was compressed, to some degree, by behavioral adjustments, likely saving millions of lives. An effective vaccine became widely available within a year from the first emergence of the virus, while becoming de facto available within *days*. This vaccine was made possible by successful international global cooperation between nations that used to be at war, discovered by dedicated individuals whose access to education and resources was facilitated by migration, and manufactured with technologies created by previous, but preserved and honed by current, generations of scientists and engineers. A similar, though of course less extreme comparison, holds for the so-called "Spanish flu" pandemic in the early 20th century.

Modern forms of cooperation, knowledge generation and distribution, preparation and planning, international governance, regulation, and technological innovation are far from perfect, and often operate in frustratingly ineffective ways. But the responses to major disasters that modern institutions of global cooperation have made possible are unambiguously better than anything that would have been possible at any earlier stage in human history. This constitutes, has been made possible by, and is reflective of, moral progress.

The case for socio-moral decline is weak. We have no particular reason to be skeptical of our epistemic ability to identify instances of progress, because we are predisposed towards social pessimism; this means that we do have reason to be skeptical of our ability to accurately perceive *decline*, because we are prone to exaggerate its probability and severity.

The evidence suggests that people are much better off today in terms of health, wealth, and safety, and that they enjoy much better levels of political participation. The trajectory of their values points in a highly progressive direction. All of these constitute progress.

4.4   The Cunning of Reason: Teleology Without Goals

The account of the mechanisms of moral progress sketched here has a surprisingly radical implication: collective moral progress often depends on individual moral vice (Levy & Alfano, 2020).

The prospects of teleological moral progress hinge on the possibility of cumulative cultural evolution. Moral gains are driven by various technological, epistemic, cultural, informational, or material mechanisms that become stored in the extended mind of a variety of institutional norms and practices. These externally stored forms of moral knowledge—successful ways of securing the benefits of cooperation for a growing number of people—are absorbed by each subsequent generation and built upon via piecemeal cultural learning. This process is crucial for intergenerational progress to occur at all; at the same time, it critically depends on a host of individual-level traits which it is difficult to classify as anything but vices.

Cultural practices are highly complex and causally opaque. This means that they cannot be invented from scratch. In order to secure their reproduction and incremental improvement, then, individual cultural apprentices must embody various behavioral and epistemic dispositions which, when seen at the individual level, seem anything but morally commendable. Cultural evolution

> portrays many of the unique aspects of human cognition as serving, essentially, interdependence and interface functionality, that is, having evolved in the face of pressures selecting for the ability to sync up with others in ways appropriate for different collective activities. These have a somewhat different flavor than the usual suspects, the types of capacities typically advanced as representing the pinnacle of human cognition, for example, the hallmarks of individual genius like problem-solving prowess and creative firepower, or whatever virtues are distinctive of moral exemplars and those possessed of the most incisive understanding and soundest judgment. Indeed, on the radical reading, many of our unique psychological capacities may even work against common ideals associated with self-reliance, such as establishing a distinct identity or achieving basic personal autonomy, because the main purpose of those capacities is to allow us to fluidly mesh with others, making us effective nodes in larger networks.
>
> (Kelly & Hoburg, 2017, p. 841)

Without obedience, conformity, gullibility, cowardice, compliance, imitativeness, docility, deference, servility, susceptibility to peer pressure—in short: the *opposite of individual moral autonomy*—cultural evolution, and, hence, the main mechanism for keeping progress alive would be a complete non-starter.

Why should we comply with the cryptic demands of social norms, practices, and institutions at all? Because of the way in which cultural practices contain accumulated knowledge that outstrips any individual's capacities. Society, in short, is smarter than you.

At least in this sense, morals are like manioc. Consider the marvelously intricate detoxification procedure manioc needs to undergo to prevent chronic cyanide poisoning which can result from eating unprocessed manioc (Henrich, 2015):

> In the Colombian Amazon, for example, indigenous Tukanoans use a multistep, multiday processing technique that involves scraping, grating, and finally washing the roots in order to separate the fiber, starch, and liquid. Once separated, the liquid is boiled into a beverage, but the fiber and starch must then sit for two more days, when they can be baked and eaten.
>
> (97)

Over the course of this procedure, the cyanogenic content of the root is reduced from 100% to below 10% on the third day. Note that skipping any one of the steps will undermine the whole process. Note, also, that the underlying rationale of the procedure remains enigmatic to those involved, since the symptoms of cyanide poisoning build up slowly and only become visible after years of consumption. This is why humans need to be over-imitators, slavishly dedicated to following each and every step of the procedure handed down to them by previous generations, however unintelligible those steps may seem (Levy, 2011). Institutions that successfully facilitate social cooperation are similar in this regard.

To counter the challenge posed by evolutionary conservatives, we need to find, implement, and maintain cleverly designed institutions that solve the problems posed by our evolved psychological constraints. This is a possibility that overly reductive, psychological responses to the conservative challenge do not take seriously enough. We don't need more empathy or altruism to make human hypersociality work. We need smart institutional support. The good news is that the possibility of such clever scaffolding can, in turn, be given its own naturalistic defense. The prospects for moral progress aren't so bad, given how much human behavior can be shaped by cumulative cultural institution building and self-domestication (Henrich, 2015; Richerson & Boyd, 2005).

The idea that evolved limitations to individuals' psychological attitudes could impose constraints on moral progress underestimates the extent to which cooperative dispositions can be "offloaded" into an institutional environment. We don't need to reinvent norms of cooperation with each new generation by finding ourselves sentimentally inclined to be nice to others. Indeed, current models of human ultrasociality and beyond-kin cooperativeness place virtually no emphasis on prosocial attitudes, which is readily admitted to being unable to sustain anything beyond cooperation in small groups (Bowles & Gintis, 2011, p. 4ff.). Instead, late Pleistocene environment set us on a path towards ever greater cooperativeness, over the course of which we constructed a highly specific niche which increased the benefits of cooperation that accrue to our species. Human beings accumulate normative capital which becomes embodied in norms of punishment, communication, and internalization (Sterelny, 2012). These normative institutions are what creates and sustains expansions of the moral circle as well as other types of progressive gains such as improvements in cooperation, the demoralization of harmful norms or increases in welfare.

On a closer look, it is not hard to appreciate why the various mechanisms of moral change should be biased in favor of moral improvement. The typology of moral progress developed earlier gives us the criteria by which to judge a given moral transformation as progressive or not: moral progress occurs when more people are made better off, when unnecessary and/or harmful norms are eliminated, when the circle of moral status is appropriately expanded, when people are empowered to more freely participate in their own self-governance, and when societies manage to secure an appropriate level of compliance with its set of norms and values without engaging in excessively harsh and ubiquitous social sanctioning.

Increase in group size, stored in institutions that facilitate mutually beneficial cooperation between strangers (McCullough, 2020), creates a teleological bias in favor of an expanding circle of moral concern. Populations grow, both globally and regionally.[6] (Demographic shifts, such as dwindling birth rates in country A, don't disconfirm this, because country A can still grow via immigration.) In order to successfully cope with an increasing number of people, a society must, in the long run, find ways of securing stable and equitable terms of cooperation between its current or aspiring members. An expanding circle is the natural result. And once certain differences between groups—ethnicity, gender, religion, or skin color—are acknowledged as morally arbitrary, more and more such markers, including species membership, will be recognized as normatively insignificant.

Larger groups automatically bring a higher degree of cultural diversity. The resulting clash of idiosyncratic social mores—differences in how people dress, what they eat, the art they enjoy—creates a moral bias in favor of increasing liberalization. When people see that alternative ways of living

are possible and allow people to thrive, it reveals the contingency of many social norms. The unchallenged authority of traditional lifestyles crumbles, and a greater variety of morally permissible courses of action opens up. Social norms become demoralized and optional.

This bias in favor of demoralization of moral norms is exacerbated by the basic inertia of cultural transmission. Rituals, skills, or norms don't pass themselves on to the next generation. Something is always lost in every transition from one cohort to the next. Social norms that have nothing going for them—no emotional resonance, no social power, no promise of protection—suffer a competitive disadvantage (Nichols, 2004). Over time, the number of arbitrary social norms, that is, norms that constrain behavior but produce no tangible benefit, will tend to shrink.

When larger groups of people are unshackled from the disabling coercion of unnecessary pseudomoral norms, they start to question whether the way their social network happens to be organized is justified to begin with. Who is in charge? And why? Whose interests are best served by the current arrangement? What justifies the exclusion of this or that segment of society? What do we have to lose? Social movements demand freedom from oppression and participation. Not all of them are successful, of course. But overall influence of such movements will be in the direction of moral improvement, because the unsuccessful ones leave their society relatively unchanged, and the ones that prevail leave their mark.

Technological developments solve problems and allow people to do things more efficiently, more easily, or, when certain tasks are outsourced to machines, to stop doing certain chores altogether. This leaves people more time to do what they want and frees up resources for them to improve their lives.

All of these developments are naturally accompanied by punitive improvements. For societies to achieve the aforementioned moral gains, they must be able to become more effective at securing cooperation from its members. But once such solutions are found, they will almost automatically lead to a simultaneous decline in the severity of punishment and a corresponding increase in the relative effectiveness of punitive practices.

Teleological explanations have long been consigned to the dustbin of history. Some authors flirt with moral teleology, but stop short of endorsing it wholeheartedly. Here is Pinker:

> Though I am skittish about any notion of historical inevitability, cosmic forces, or mystical arcs of justice, some kinds of social change really do seem to be carried along by an inexorable tectonic force. As they proceed, certain factions oppose them hammer and tongs, but resistance turns out to be futile.

> (Pinker, 2018, p. 109)

Robert Inglehart writes: "cultural change is a process through which societies adapt their survival strategies. The process operates as if evolutionary forces were consciously seeking to maximize human happiness" (Inglehart, 2018, p. 140).[7]

Another option, which I will not pursue further here, would be to vindicate moral teleology transcendentally. The question of whether there is reason in history is, to some extent, about how we rationally reconstruct it. If there is no alternative for us, epistemically speaking, to viewing history under teleological auspices, then teleology is transcendentally justified. For instance, there is evidence that we judge persistence in teleological terms (Rose et al., 2020) Whether or not we think a society or culture continues to exist at all depends on whether it does well or fulfills its (real or imagined) purpose. The Roman Empire didn't *literally* collapse: the Italian peninsula is still there; no natural disaster wiped out its population. And yet we speak of the fall of Rome because various social upheavals led to its downfall as a political unit. The empire ceased to exist because it didn't maintain its strength and glory.

What would we expect to see, respectively, if a teleological or a non-teleological account of socio-moral change were true? If anti-teleologists were right, we would not expect to see *no* moral improvement anywhere ever, but we would expect historical patterns to consist in a more or less random, patternless drift between periods of moral improvement, stagnation, and decline. Likewise, if the teleological picture were accurate, we would not expect to see *no* hiccups, blunders, catastrophes, and phases of reaction. But we would predict more or less steady tendencies of improvement along the main axes of the good, and more or less steady pushback against the, as it were, axes of evil, resulting in long term progress for ever larger groups of people.

The teleological version of history is more strongly confirmed by the empirical evidence. What we see is a fairly robust development of societies in the direction of moral improvement. These trends do not start at the same time in all places, and do not unfold at the same pace, but eventually, more and more societies, however culturally different they may otherwise be, are swept up in the same dynamic. One important thing to realize is that we do not just see societies bumble through towards moral advancement. Rather, we see the forces of social change pushing society in the direction of a coherent moral outlook (Huemer, 2016a). We do not find that societies develop norms against discrimination and xenophobia while at the same time opposing democracy and embracing aggressive wars. Societies don't abandon gratuitously harmful norms such as female genital mutilation or a highly rigid and repressive Victorian sexual morality, while otherwise becoming less tolerant and liberal. Rather, we see that socio-moral change is biased in a coherent direction of recognizing the individuality

and fundamental equality of everyone (who is someone). It does so by implementing, over the course of socio-cultural evolution, institutional solutions to cooperation problems. Every once in a while, a new solution is discovered. If it makes people better off and allows more people to treat each other better, that solution thereby enjoys a competitive advantage in the process of cultural evolutionary selection of moral norms and values.

Moral progress occurs because of the kind of being we are. We are intelligent problem solvers capable of cultural learning, and prone to storing our accumulated knowledge in social practices and external institutions. Societies gravitate towards a state of affairs where more people are better off, cooperation is arranged more smoothly, people enjoy increasing and more far-reaching liberties, and where they are given more—and more meaningful—opportunities. The resulting changes in social norms, cultural practices, technology, economic reproduction, and group size mean that more members are admitted and more blatantly arbitrary norms are eliminated. These developments stabilize and reinforce each other.

According to the moderate form of teleology I am advocating, social change is biased towards moral improvement. This version of teleology is metaphysically humble and naturalistically respectable; it is a teleology without goals: there are no iron laws of progress; there are no guarantees that it will happen; there is no end state of moral perfection we are heading towards; and there is no objective spirit in charge of the whole thing. Of course, it is important for societies to get on the path towards progress in the first place, and not every society does. But once certain mechanisms are in place, the journey towards greater freedom, opportunity, well-being, inclusion, and equality has started.

That there is teleological moral progress does not mean that all societies will ultimately converge on one unique way of living. However, it does mean that many (successful) human groups will over time show at least a solid degree of convergence. The existence of the occasional oppressive hell hole or backward theocracy does not decisively refute the teleological thesis. There will always be pockets of moral suffocation. Moral change is path-dependent, and societies can become locked into suboptimal trajectories that are hard to redirect.

This story also allows for the possibility of regress. Whenever a shift to new forms of cooperation needs to be made due to increases in group size, changes in the natural environment, new forms of knowledge or social practices, disruptive psychological tendencies, and unwieldy social practices, bequeathed to us by biological evolution, cultural evolution, or individual idiosyncrasies, stand in the way of reintegration and only some groups make the cut. Those who figure out how to institutionalize suitable new forms of cooperation that are adequate for coping with the forces of social combustion may just make it.

The resulting modestly teleological account of moral change is that the driving force behind social progress is grounded in a combination of intelligence, cultural malleability, and institutional niche construction. Once there are beings that are (a) intelligent problem-solvers and (b) malleable learners who can (c) construct their own corner of the world (Sterelny, 2007b), moral progress will occur. Intelligence and malleability together entail that individuals as well as groups will likely stumble upon new solutions to perennial problems such as safety and survival, which leads to increases in scale. These increases require new forms of securing cooperation, which some groups figure out and offload into their norms and institutions. This leads to further increases in scale, which drives a new round of innovations for how to make cooperation happen. All of this will make human beings more norm-sensitive and cooperative, and increase the circle of potential cooperators. This *is* moral progress. Due to the resulting success of those groups, other groups will be inclined to copy the practices and institutions of the former group. This is how progress spreads. History is biased towards changes that it makes sense to describe as moral improvements. This means, simply put, that moral teleology is true.

Again: no naturalistically respectable theory of moral progress will operate with any sort of divine intervention/objective spirit/intelligent design account of teleology, according to which the march of history is secured by the guidance of a (quasi)personal entity such as god or the "Weltgeist". The ontologically frugal version of teleology I am advocating will end up hinging on processes of socio-cultural evolution. Teleology 2.0 is a story about how cultural evolution is sensitive to moral concerns. A peculiarity of cultural evolution is that the satisfaction of human interests—that is, moral values—constitutes one of the most important selection pressures for cultural variants. Because this is so, processes of cultural evolution are steered, at least in part, by what's right. This means that there are forces in place that push human groups in the direction of moral change for the better. This, in turn, is another way of saying that history is biased towards moral improvement.

Improvements in security or prosperity or stability will lead to corresponding shifts in the direction of more inclusion, more cooperation, more freedom, more autonomy (Inglehart, 2018; Buchanan & Powell, 2018). But if this is so, then we may be caught in a positive, self-propelling feedback loop where greater institutionally secured prosperity leads to greater willingness to tolerate difference, stop discriminating, and cooperate more. This then leads to further prosperity and security, which accelerates the process even more.

A key feature of humanity that makes societies *globally* gravitate towards moral improvement are the ability to deliberately copy and migrate. In animals, migration and copying do occur, but blindly. Animals don't migrate

to different groups because they are being treated more fairly there, enjoy more freedom, or because they can create a better life for themselves and their children. They don't deliberately adopt foreign lifestyles because these have turned out to be more successful, or more conducive to their welfare. They don't adopt new technologies, or remix technologies in innovative ways, to create smarter solutions for the problems they encounter. But humans do all of that, all the time. And if they keep doing so, their way of living has an inherent tendency to improve.

The key idea to make good on for a naturalistically credible teleological account of moral progress is whether cumulative cultural evolution selects for social practices that promote and/or constitute moral improvements. If cultural evolution accumulates moral capital like that, social change will be teleological in shape. We already know that cumulative cultural evolution can, and does, produce improvements in knowledge, skills, or technology. But morality is, crudely speaking, just one type of social technology: its function is to address various obstacles to social cooperation and integration. Cultural evolution will gravitate towards arrangements in which more people cooperative more freely, peacefully, and equally, because that is what improvements of social technology consist in.

Ultimately, then, the teleological story is a story about discovery, transmission, and storage of moral innovations (Henrich, 2015). Within a cultural evolution framework, the teleological claim—that social change is biased towards moral improvement—becomes this: we can imagine that "society" (group C at t_1) starts with some kind of normative infrastructure, some set of rules to generate, secure, and police cooperative arrangements. Such rules are transmitted from one generation to the next. Some of that transmission probably has a strong genetic component (for instance, we care about our offspring, and bequeath this attitude towards them). Other rules belong more firmly to our cultural heritage. Now this culturally accumulated body of knowledge, norms, and practices get transmitted from one generation (G_1) to another (G_2). But there are various factors that make sure that the copy of the moral code M_1 from G_1 created by G_2—that is, M_2—will differ slightly from its previous instantiation. That difference will be due to a mix between a random component (we can call these random "mutations" of the moral code) and a more systematic, non-random component. *A naturalized teleology claims that the intergenerational copying process of moral codes is such that the non-random component of the mix of factors, some of which remain genetic, that modify M2 compared to M1 contains morally non-neutral filters.* Individuals of G_2 will copy M_1 (thereby creating M_2) in ways that introduce ever so slight moral improvements. This marginally improved moral code will not only do better, ceteris paribus, in cultural transmission than its marginally worse predecessor, but it will also be more readily absorbed horizontally, by members of the

same generation. Now, a number of moral innovations are introduced at roughly the same time, and those that successfully "fit" together or are more easily coupled with each other—say: equality and individual freedom go together better than equality and support for violence—will combine to create an even more viable moral code in G_n.

This process is deeply path-dependent. Not all possible moral innovations are discovered, which is why social change is biased in favor of comparative moral improvement, not headed for absolute moral perfection. Some authors agree that cultural evolution is biased, to some extent, towards moral improvement:

> Ancient and tribal instincts provide different tricks to identify others in the group as likely cooperators and, more importantly, to detect and sanction defectors. It has been posited that together these can collectively act as a "moral hidden hand," or a source of pro-social behavior and psychological stickiness that influences the spread and evolution of norms. Recall that on its own punishment can stabilize any norm: the useful and the pointless alike, as well as the just and unfair, the cruel and the kind. The moral hidden hand, however, can act as one of the pressures that drive social change toward the better—or at least the more prosocial—by giving a cultural fitness boost to norms that lead individuals to act for the good of the group, paying personal costs for the sake of others. Thus, norms that activate our feelings of empathy, our sense of fairness, or our aversion to gratuitous harm, receive a transmission advantage over those that do not. This, in turn, acts as a gentle but persistent selection pressure favoring more equitable and compassionate social arrangements over the long run, because the norms that prescribe such arrangements are more likely to "mesh" well with the range of human cooperative instincts.
>
> (Kelly & Davis, 2018, p. 70f.)

Some moral innovations spread more easily or rapidly (or both) than others. Which moral discoveries are more likely to be adopted, copied, and proved sustainable? Probably those that will serve the interests of a comparatively larger group of people, those that attract new followers, those that successfully secure cooperation, intragroup helping, and so on. Among the reasons that make moral innovations likely to be successful are that these innovations constitute moral improvements.

Consider the case of "Magdeburg Law":

> [I]n 965 CE, Church records note that "a group of Jews and other traders" had set up shop in Magdeburg (Germany), along the Elbe River at the edge of the old Carolingian Empire. A decade later, the Holy Roman

emperor Otto I formally granted "privileges" to this community. Gradually, Magdeburg's approach to civil administration, the regulation of guilds, and criminal laws were forged into what became known as Magdeburg Law. By 1038 CE, Magdeburg's success had begun to inspire other communities to copy its laws. In the next several centuries, over 80 cities would directly and explicitly copy Magdeburg's charter, laws, and civil institutions.

<div align="right">(Henrich, 2020, p. 310ff.)</div>

What made Magdeburg so successful that other places wanted to copy and live by its distinctive set of laws and institutions? For instance, according to Magdeburg Law, it was no longer possible for a man to be held responsible for a murder committed by his son. In earlier customs and legal codes, criminal liability remained suffused with the intensive kinship ties that form the basis of social organization in all (historically known) human societies. A number of other rules gradually implemented a greater focus on individuals and their actions, granting more expansive freedoms whole doling out more targeted sanctions. It is not difficult to see why such an institutional arrangement would be both popular and highly adaptive for communities to copy. Whatever deterrent effect of external sanctions there is will be stronger when the punishment is provided exclusively to the actual transgressor; otherwise, a thief or murderer may be able to shift the costs of their deviant behavior on other people in their immediate family. More effective forms of punishment increase the robustness and scale of cooperative structures (Fehr & Gächter, 2000), which makes cities that adopted those new regulations better places to live. At the same time, it would be hard to reject the idea that allocating punishment only to those who actually committed a punishable offense constitutes a tremendous moral improvement in and of itself.

What is the fundamental "metaphysical" explanation for the teleological structure of social change, that is, the moral bias in historical progress? One possible suggestion (Christakis, 2019, p. 111ff.) could be that the evolution of social practices and experiments in living will make sure that over time, more and more of the "morphospace" of possible social arrangements is being explored. Not all *logically* possible forms of life and packages of norms, practices, and institutions actually end up being tried out because of various natural, physical, psychological, emotional, motivational, or cognitive constraints. Of the forms that are being tried out, only some turn out to be actually sustainable, livable, desirable, or stable. Due to the peculiar characteristics of human beings, who can learn from other social groups or individuals or choose to be absorbed by and join them, some of these social arrangements will prove to be more viable. But—and this is the crux—since the very viability of the social arrangements that

enter this experimental morphospace depends on whether or not they are responsive to genuinely morally significant values, the overall trajectory of social change will be one of moral improvement. People are more likely to remain in, emulate, or join societies that more fully realize human values—that make people better off, treat them more equally, and give them more freedoms. Moral values are, in this sense, part of the selective pressures acting on cultural evolution.

Naturalized moral teleology is, at its core, a form of historical materialism. It contends that ultimately, broadly "material" forces are responsible for driving people's values or, at the very least, for providing the fertile ground that allow values whose time has come to thrive. Another way of putting this point is this: it was only a matter of time until a culture evolved with comparatively more progressive features: a society with a focus on individualism and weaker kinship ties, impersonal sociality, positive-sum thinking, peaceful cooperation and commerce, or impartiality. And once that culture evolved, it was bound to be comparatively more than many others successful.

This is not to endorse a form of ethnocentric Western triumphalism. The values that shape moral progress—greater freedom, greater well-being, and greater equality—are not Western values any more than mathematics is, in any interesting sense, Arabic simply because that's where the numbers people do it in originated. Rather, there happened to be one region of the world that, after 10,000 years of oppression and abuse, made these values central to its political organization and moral culture. But that doesn't make these norms and values Western in any interesting sense any more than there is a Western physics or an Eastern biology. Certain scientific discoveries were, as it happens, first made in certain places rather than others. But these scientific discoveries are part of the universal heritage of humanity. They had to emerge somewhere, but they belong to everyone. The same holds for the values of progress. Indeed, we see that very similar cultural differences evolved *within* other cultures such as China, for instance (Henrich, 2020, Chapter 7.)

It is true, however, that this account involves very strong commitments to the possibility of moral comparisons between cultures: on the teleological account, societies in which children are beaten, most women disenfranchised, most men oppressed, and which are joyless, impoverished hellholes (think of various theocracies or authoritarian regimes today or throughout history) are *less morally developed* than some existing alternatives.

4.5   The Arc(s) of History

The teleological thesis, as I understand it here, is that social change is biased in favor of moral improvement. But pointing out that things have gotten better in some places sometimes rather obviously doesn't support

this claim; for the teleological thesis to be true, progress has to be systematic rather than lucky.

But doesn't history disconfirm the teleological thesis? For progress to be systematic in the required sense, there must be mechanisms in place that reliably (even if unsuccessfully on occasion) push societies in an ameliorative direction. For the longest time, however, virtually nothing happened that would seem to display this pattern. For the longest time, virtually nothing happened *at all* (at least in terms of big *structural* upheavals in human society).

Whether this confirms or disconfirms the moderate teleological story I advocate here depends on what we expect the shape of the moral arc to look like. If we include their pre- and protohuman ancestors, humanity's journey thus far has been millions of years long, and *Homo sapiens*' hundreds of thousands. And yes, for the longest time, little to nothing happened—as in, no progress at all. The kinds of developments covered in this book are mostly about extremely recent phenomena—the abolition of slavery was witnessed by people who knew someone *you* knew, female emancipation was witnessed by someone you knew, the gay rights revolution was witnessed by *yourself*.

What are we entitled to assume about the shape of history? Any attempt to answer this sort of question will be largely speculative. Though perhaps not entirely: the idea that the actual trajectory of history, if there is such a thing, disconfirms the moderate teleological thesis I have articulated here seems due to the assumption that the arc of social change is *linear*, such that the magnitude of the moral improvement undergone by the respective group at issue remains constant per unit of time.

But this assumption is implausible, for this is not the shape we observe in any other field important to human history. For the longest time, there was little to no economic growth; population was mostly stagnant, technological innovation slow, and scientific advancement was distinctly unimpressive. It was only after various social tipping points were reached that growth in these areas accelerated, and then kept accelerating.

Will things continue to develop like this? Can we expect moral progress to be exponential(ish)? Even I find this hard to believe, but what does seem conceivable to me is for the moral arc to consist of a series of logistic(ish) growth phases that start slow, accelerate enormously for a limited amount of time, and then start tapering off again. At this point, another period of slow development may ensue, and at one point the question becomes whether the society at issue will be able to jump to the next sigmoid curve or face decline on a downward slope of moral deterioration. This (sketch of a) model would suggest that history is biased in favor of moral improvement, and that there are forces that push social change in an ameliorative direction. It would also, however, suggest that humanity isn't as tightly

hinged to this escalator as we may wish it to be, and that humanity and its various subgroups will forever remain on a series of boons and crises, not all of which may be resolved fortuitously.

4.6 Circularity and Smugness

Much of the awkwardness and epistemic angst in judging the past in terms of progress and regress stems from the impression that judgments of progress are self-congratulatory: it seems that such assessments are necessarily *circular*, in that the present will inevitably appear morally superior to the past, since the set of substantive moral standards we avail ourselves of in judging earlier generations and social practices is the set of *present* moral standards. If we imaginatively turn the tables on ourselves, and judge present society in light of what we know about past moral norms and values, the present seems worse. And the teleological thesis is the pinnacle of this smugness: not only are "we" superior, but it also *had to* happen this way. What gives?

To judge the past in terms of present norms and values—such as improved well-being, greater equality, an expansion of the moral community, or the abandonment of superfluous moral constraints—seems like an arbitrary choice, where the second option (the present) vindicates itself.

But, while tempting, this worry is unjustified. First, let me emphasize the same kind of worry applies to any other field as well. When we judge present technology, science, or mathematics to be in some sense "superior" to earlier incarnations, what we avail ourselves of are the substantive criteria we have *right now*. What other criteria would we use?

Moreover, it is simply not true that the choice between moral standards is an arbitrary one. Consider the following example: judged by the rules of poker, you ought to do this; judged by the rules of chess, you ought to do that. If we were chess players who, one day, woke up as poker players, it would seem arbitrary to denounce chess as inferior simply because we happened to have switched from one to the other miraculously, overnight, for no intelligible reason whatsoever.

But this is not what happens with moral progress: actually, the later stage emerged from and within the first. This is a process of learning and experience, not an arbitrary, symmetrical choice between two distinct axiological systems that we could imagine going just as easily in the reverse direction. It's as if once upon a time, there was a society of chess-players; it worked just fine for some, and not so fine for many. Then, some of the chess players started to protest, and to demand more poker-like rules; a few of them were implemented, then more, then even more; the poker society worked out fine for a larger number of people; moreover, none of the adjacent societies seemed to have a problem with it. In fact, some—like the

backgammonists—started copying some of the new poker-rules to become more poker-like as well; so they did, and it agreed with them. Now, there are always some who want to go back to the good old days of chess, but they are in the minority.

The circularity worry is fueled by the idea that we have two different systems of moral norms and values, and that from the perspective of either of them, the other must seem bad and/or unjustified just like, from the perspective of the Bible, modern biology is unjustified, whereas from the perspective of modern biology, the Bible is. But this, again, is not what happened: rather, modern biology emerged from and within a theistic framework, and we deliberately developed in that direction because the theistic outlook proved unsustainable when judged *in light of its own* standards at the time.

Notes

1. https://ourworldindata.org/optimism-pessimism#individual-optimism-and-social-pessimism
2. www.jacobinmag.com/2018/10/steven-pinker-enlightenment-now-review; www.newstatesman.com/culture/books/2018/02/unenlightened-thinking-steven-pinker-s-embarrassing-new-book-feeble-sermon; www.nytimes.com/2018/02/28/books/review-enlightenment-now-steven-pinker.html
3. www.systemicpeace.org//inscr/p4manualv2016.pdf
4. www.worldvaluessurvey.org
5. For useful discussions of the moral status of animals from a variety of moral perspectives, see Singer (2011), Korsgaard (2018).
6. https://ourworldindata.org/world-population-growth
7. David Haig tries to revive teleological explanations in the realm of biology. He argues that teleological language is part and parcel of the vocabulary that causally explains evolutionary processes, because it is just as true that genes cause behavior as it is true that differences in goal-directed behavior cause shifts in the frequency of genes between generations—which is to say that goals influence the trajectory of evolution (Haig, 2020).

5   Beyond Expansion

Which Types of Moral Progress Are There?

Introduction

According to the currently dominant account, the story of moral progress goes something like this: once, moral recognition used to be the privilege of a select few. Full moral status was only accorded to people of a certain class, age, gender, ethnicity, religion, or all of those combined. Over time, the moral franchise was gradually extended, however slowly and imperfectly, to include human beings of all races, creeds, or genders. Eventually, even species membership started to become recognized as morally irrelevant, and the moral circle is extended towards non-human animals as well. One of the most influential contemporary ethicists, Peter Singer (2011b), and others call this the "expanding circle" of moral concern. Moral progress, on this account, mainly consists in further expansions of the moral realm beyond ethically arbitrary features.

The expanding circle account of moral progress has recently been given a systematic update. In their 2018 book, Allen Buchanan and Rachell Powell argue that moral evolution, particularly over the last centuries and decades, is best described in terms of a series of inclusivist shifts over the course of which moral status is granted ever more generously. In particular, they single out four key developments as evidence for this trend (which they refer to as "the inclusivist anomaly" (153ff.): the increasing recognition of the moral standing of non-human animals, the reconceptualization of morality as a domain with universal rather than regionally indexed applicability, the emergence of a culture of human rights, and, finally, the development of so-called "subject-centered" moral outlooks which disentangle moral status from strategically relevant capacities for cooperation—which can be either completely or partially absent in, for instance, severely disabled people or small children.

In the previous chapter, I addressed the question of whether evolution, and the mark it has left on our psychological dispositions, imposes any significant feasibility constraints on the prospects of further inclusivist gains,

and I argued that it does not. This is important, of course, for the prospects of moral teleology as well, because social change can't be biased in favor of moral improvement if the improvements that history is supposed to be headed towards are impossible to attain.

What are the types of progress we want to accomplish? In this chapter, my goal is to develop a typology of moral progress. I aim to correct the somewhat reductive focus on the expanding circle as the main, or perhaps even the only, form of moral progress worth zooming in on. Instead, the typology of moral progress I will offer displaces the inclusivist anomaly from center stage, and argues that, while important, the expanding circle is merely one of several types of moral progress and may in many cases merely be a welcome by-product of other, more fundamental progressive shifts. Second, I will correct for some—to my mind—rather striking omissions in current inclusivist theories of moral progress, which often have surprisingly little to say about highly morally significant improvements in well-being or increases in liberty. This chapter is supposed to fill these gaps.

The ultimate aim of this chapter lies in its contribution to the teleological argument developed in this book as a whole: we won't be able to properly address the teleological structure of social change until we have a theoretical apparatus in hand that allows us to connect the dots between types of progress, the mechanisms that are driving them, and the institutions that sustain them.

5.1   Well-Being

One of the biggest and most striking blind spots in contemporary theories of moral progress is the relatively peripheral role granted to increases in people's *well-being*. It seems that one of the main concerns when it comes to understanding moral progress should be to understand how to make people better off—what this means, what it amounts to, and how to make it happen. Consider, in contrast, that increases in well-being are not even mentioned as a distinct form of moral progress by Buchanan and Powell (2018). Likewise, Victor Kumar and Richmond Campbell (2022) write: "the fight for moral progress is principally a fight for greater inclusivity and equality". This may be so for the already well-off, but the severely poor ones—a group which literally everyone used to belong to not so long ago—understandably have different priorities. For the vast majority of people then and now, improvements to their living conditions matter the most. In this section, I want to very briefly explain what I mean by "well-being", and I will make a few remarks about why well-being is highly morally significant, and thus why changes in well-being should count as an important index of moral progress.

One thing that needs to be shown is that well-being is a morally relevant category. Here, it cannot be emphasized enough that conceding this point is perfectly non-partisan: all moral theories, not just utilitarianism, agree that people's well-being is morally important. Kantianism recognizes imperfect duties to promote one's own and others' well-being and identifies the highest good as the state of affairs when people's happiness will be proportional to their moral worth; virtue theories like to describe moral action as the main avenue to proper human flourishing; and so on.

Ultimately, the issue is whether well-being has been subject to progress. Here, the evidence is complex, but it seems to point in the direction that it has been and continues to be. For now, the problem is how we should understand the concept of human well-being so that we can bring it to bear on the issue of moral progress.

As is well known, extant philosophical theories of well-being come in three flavors (Tiberius, 2014). Hedonistic theories hold that well-being is a subjective state of mind: people are happy when they (on balance) feel good, that is, when their subjectively agreeable states of mind outweigh the disagreeable ones in quantity and quality. Desire-satisfaction theories hold that well-being is about people getting what they want, or having their preferences satisfied. Finally, some theories of flourishing insist that happiness is about whether a person's life objectively contains various goods.

Much of the philosophical debate has been about showing that these three features—pleasure, desire-satisfaction, and objective goods—can come apart. This is unfortunate, because it has clouded the fact that in real life, they almost never do. In reality, these three theories highlight different aspects of flourishing that are reliably connected, except in unusual and/ or pathological circumstances of, say, wealthy people with great friends and a meaningful career who are nevertheless depressed (think: Ebenezer Scrooge, but affable), or people in what appear to be objectively miserable conditions who are nevertheless happy (think: Tiny Tim), such that there is a disconnect between subjective pleasure, preference satisfaction, and objective goods. It should be clear, however, that in the vast majority of cases, obtaining various important goods, sustaining crucial capabilities, getting what one wants, and feeling good as a result are robustly correlated. This is not to say that it isn't philosophically important to keep these aspects of flourishing conceptually distinct. It does suggest, however, that for a theory of moral progress which aims to offer an empirically rich account of social change for the better, such theoretical niceties can very frequently be ignored.

Jeremy Evans (2017) suggests that, in lieu of an agreed-upon account of moral progress, we use people's well-being—more precisely, the slightly more technical notion of population welfare—as a proxy property. He argues that if we can't track social change for the better directly, we may

still be able to track it *indirectly*, by identifying a property that tends to be reliably correlated with moral progress (see also Luco, 2019). Population welfare, he argues, does the job, since increases in population welfare (regardless of whether these are understood and/or measured in terms of subjective self-reports or more quantifiable indicators) tend to be highly correlated with increases in cooperativeness, social capital, autonomy, and even (opportunities for) political participation—that is, the moral improvements we care about.

There are two main theoretical problems with using well-being as an index of progress, one epistemic, one psychological. The first one has to do with whether we know how people's levels of well-being change; the second one is about whether well-being is a useful measure of progress (regress) at all, given how easily exhaustible welfare gains are.

Regarding the first point: much of the skepticism towards adding human welfare to the typology of moral progress, it seems to me, has to do with skepticism about whether we are able to trace development in people's well-being over time. How can we know how people used to feel centuries or millennia ago? How do we know if, and to what extent, people are happier or unhappier today? How could we even begin to measure this?

Assessing the well-being of past generations has obvious limitations. We cannot include them in our surveys, for instance. It is worth mentioning that, though we cannot rummage through their browser search history either, we can take into account other inadvertent expressions of subjective well-being in poetry, song, and religious ritual. Here, the picture we see is bleak, as we receive a litany of misery and despair, and the constant declaration that life is but a valley of tears. This is further exacerbated by the observation that only the very wealthiest individuals, those with access to some means of transgenerational preservation, were able to bequeath their respective *De Profundis* to us at all. There is a strong selection bias in the historical record. All the others were, presumably, too busy suffering and starving to leave anything on record. Moreover, if it were impossible to know how happy people were in the past, pessimistic narratives would also be undermined. Agonosticism about past well-being supports neither pessimism nor optimism.

Skeptics such as Branko Milanovic (2010) hold that except in sufficiently close cases (say a generation or two), intertemporal wealth comparisons lead to absurd conclusions, which means that we must adopt time-relative standards of wealth. Following his line of reasoning, we would otherwise be forced to conclude that, because we have light bulbs and Croesus didn't, Croesus was poor. Moreover, because our descendants may vacation on Mars and people alive today can't, so is Jeff Bezos.[1] (Milanovic suggests that we should measure wealth in terms of the amount of labor a person can command with her income, which would make John D. Rockefeller the wealthiest person of all time.) It seems even more absurd, however, to

go all-in with the intertemporal incommensurability of wealth. Granted, it rings false to say that Croesus or Jeff Bezos was/is poor *sans phrase*, but it doesn't seem implausible to describe them as poor *relative* to a comparison group of healthy, 200-year-old interplanetary vacationers from 2150.

The second, psychological problem is due to the fact that seemingly objective improvements in people's circumstances often do not translate straightforwardly into corresponding gains in how well people's lives are going *for them*. A version of this has become known as the "Easterlin" paradox (1974). When subjective reports of individual well-being are compared, we see that within rich countries, wealthier people are happier than less wealthy people. But when we compare average well-being *between rich and poor countries*, we do not see that rich people are happier than poor people. What is going on here?

Various explanations for the paradox have been offered. It now seems that the best response to the paradox is to deny that it is real. More recent and thorough analyses of the available data show a clear link between wealth and happiness (Stevenson & Wolfers, 2008; Veenhoven & Vergunst, 2014). The gains do taper off, of course, and beyond a certain threshold (the exact level of which remains controversial), further material improvements have considerably weaker clout. At any rate, it is worth noting that the Easterlin paradox has always been essentially about diminishing marginal utility: once you make $20,000 (or so), making a little more won't make you that much happier. Note, however, this if you make that much money, you are already very rich in global comparison, so the claim that money isn't important sounds compelling only to those who have it. More importantly, once we look at the data properly, even this effect largely disappears. What matters is not so much how absolute increases in wealth produce changes in happiness, but how changes of the same "magnitude" do. Roughly speaking, a fourfold increase in GDP per capita in a country yields a gain of one point on a 1–10 life evaluation scale. This relationship becomes highly salient once the correlation is plotted logarithmically (Deaton, 2013, p. 21). Richer people *are* happier.

The relevance of the Easterlin paradox for using well-being as a metric of progress is further diminished by the fact that it has no little or bearing on non-subjective accounts of welfare. On either preference-satisfaction or objective list theories, well-being increases when people get what they want or obtain more of the important goods and functionings more reliably. People want to live longer, healthier lives, with more teeth in their mouths, fuller bellies, and fewer deceased children. They want security, leisure, community, participation, and opportunity. All of these things can be assessed in objective terms, thereby evading many of the epistemic problems there are with accessing levels of subjective happiness, from adaptive preferences to hedonic adaptation or positional effects (Haybron, 2007).

One positive reason for including well-being as an axis of progress is that people who are better off are, on average, better people. Why? Because affluence breeds moral improvement. This is a proposition that many will find scandalous, but on closer inspection, it becomes hard to deny. Deirdre McCloskey (2006) agrees that if, in gaining the world, we should lose our soul, it wouldn't be worth it. However, she argues that we haven't lost our soul; rather, increases in material well-being have almost always resulted in corresponding moral improvements. This is largely due to the fact that gains in economic productivity and access to resources liberate people from the yoke of subsistence, enabling them to pursue more high-minded goals:

> Richer and more urban people, contrary to what the magazines of opinion sometimes suggest, are *less* materialistic, *less* violent, *less* superficial than poor and rural people. Because people in capitalist countries already possess the material, they are less attached to their possessions than people in poor countries. And because they have more to lose from a society of violence, they resist it. You can choose to disbelieve if you wish some of the things said to go along with the capitalist revolution of the past two centuries, such as the emerging global village, the rise in literacy, the progress of science, the new rule of law, the fall of tyrannies, the growth of majority government, the opening of closed lives, the liberation of women and children, the spread of free institutions, the enrichment of world culture.
>
> (McCloskey, 2006, p. 26)

I would also like to reiterate a point I made in previous chapters, which is that improvements in well-being deserve to be classified as a form of moral progress because the most important of them—improvements in health, wealth, and safety—have been brought about by value-guided cooperative efforts to improve the human condition. They are the direct result of moral action.

A third reason why gains in welfare are an important source of moral progress is that such gains represent trade-offs we no longer have to make. This point has been forcefully made by Tyler Cowen (2018). Many seem to think that human welfare is an ambiguous source of progress because in the pursuit of wealth, we may have to face unpleasant trade-offs and sacrifice other, equally important values, such as meaning, satisfaction, spirituality, and other wholesome-sounding things. But Cowen shows that what economic growth does is to relieve us of many of these trade-offs altogether: the wealthier we are, the fewer ugly trade-offs we face. Other aggregation problems disappear as well. The wealthier we are, the more individual rights we can "afford" to respect. Deontological rights become more robust as utilitarian gains accumulate.

It is of course true that increases in GDP are not directly sensitive to distributional concerns. Wealth per capita can be high when few people have a lot while most have little. But while this is strictly speaking true, living in high-GDP countries is nevertheless best for the poorest and living in a low-GDP country worst for them. So anyone but the most literal-minded egalitarians should be able to appreciate the benefits of affluence. We don't have to heap all our praise on markets, either: when it comes to improvements in human well-being, technological innovations or novel forms of cooperation are just as crucial. We don't just care about the size of the loot; we also want to know how it is distributed.

5.2   Equality

When it comes to giving examples for moral progress, gains in social equality often come most readily to mind. The abolition of slavery is a prime example; so are the achievements of the women's movement, the gay rights revolution, and many other ongoing efforts to undermine prejudice, oppression, and discrimination.

In all of these cases, equalizations of status and condition constitute moral progress. Victor Kumar and Richmond Campbell (2022), for instance, suggest that equality is one of the two main metrics of moral progress, the expanding circle of moral status being the other. Thomas Scanlon goes even further: "[t]he increased acceptance of the idea of basic moral equality, and the expansion of the range of people it is acknowledged to cover, has been perhaps the most important form of moral progress over the centuries" (2018, p. 4). Many are thus attracted to the idea that equality should count as an important dimension of moral progress. And I won't deny that it is—at least, not quite.

The problem is that there are fairly convincing objections to the idea of recognizing equality per se as morally significant. The leveling down-objection is perhaps the most famous one: if equality were morally relevant, we should consider a more equal state of affairs in which no one is better off but some are worse off as (at least pro tanto) morally preferable (Parfit, 1997). This seems odious, and indeed spiteful. Following Harry Frankfurt (1987), many now recognize that equality *as such* isn't really important: what matters is not how people do relative to others, but how well they do, period. Contemporary egalitarians have taken such arguments to heart, trying to object to existing inequalities not because they are offensive to literal equality, but because inequality tends to have all sorts of other objectionable consequences. If anywhere, these egalitarians hold, the badness of social inequality can be found in how differences in social status can cause feelings of shame and humiliation, in how economic inequality leads to unfair political influence and unjust relationships of

domination (Scanlon, 2018), or in how social hierarchies encourage wasteful positional arm races (Pickett & Wilkinson, 2010).

To make matters worse, it is becoming increasingly clear that egalitarianism lacks metaphysical credibility (Husi, 2017). It is not only very difficult to find a plausible candidate for a property that is both (a) morally significant enough and (b) shared equally by all and only the entities we would like to populate the coveted kingdom of ends with. The quest for grounding moral equality starts to appear thoroughly hopeless once we appreciate that we don't just need an egalitarian feature of this sort; we also need to make sure that there aren't any other, additional features that disable moral equality again. Some authors now suggest that to get around this problem and maintain our egalitarian commitments, we should simply pretend to be ignorant of some of people's morally relevant differences (Carter, 2011). This seems hardly promising.

Finally, there is the pragmatic problem that, even if equality were inherently morally relevant, it is simply not that easy to find good and/or effective ways of achieving robust socio-economic equality that are worth the cost. Ever since we began living in larger groups (starting around 10,000 years ago), human societies have been intensely stratified, perhaps lamentably. Since then, every new generation has tried to return to the more egalitarian social structures that are most agreeable to our evolved psychology (Boehm, 1999). The main ways of successfully compressing inequality, however, were almost always tremendously violent: only warfare, disease, or systemic failure is known to disrupt inequality (Scheidel, 2018). Peaceful measures hardly make a dent.

On the other hand, one should not exaggerate this point: it is clear that many social measures, from taxation to outright revolution, have succeeded at achieving at least somewhat greater equality. The abolition of feudalism and segregation were surely desirable and effective. The point here is that luck and liberty tend to do their best to upset egalitarian patterns, and that even if one finds this morally and politically undesirable, it is not always clear what to do about it.

Does this mean that we should cross equality off our list of important types of moral progress? I don't think so, and here's why. One major reason why we should recognize the reduction of inequality as progressive is that inequality is intersectional: disadvantages tend to cluster (Wolff & De-Shalit, 2007): first you become unemployed, then you lose your health insurance, then you become homeless, and then substance-dependent. This piles disadvantage upon disadvantage, and creates a vicious circle of disenfranchisement and impoverishment. Moreover, highly unequal societies tend to suffer from eroding social trust, and harmful status competition (Pickett & Wilkinson, 2010).

Almost all social movements that fought for social justice under egalitarian auspices were resisting prevalent social *in*equalities, and even nonegalitarians can agree that removing unjustified inequalities is progress.

Non-egalitarians can comfortably be staunch anti-inegalitarians. Since non-egalitarians suggest that we disabuse ourselves of *any* comparative considerations, they also reject that we justify inequalities on the basis of whatever features—class, caste, race—advocates of inequality may favor. Almost all historically existing forms of inequality were clearly unjustified. It is true that non-egalitarians will often refuse to condemn certain states of affairs *just because* they are unequal. But actually existing egalitarian social movements did not demand equality for the sake of equality. Rather, they called for the elimination of discrimination and oppression, and for access to meaningful opportunities. "Equality" was the catchy label they chose for those demands.

All of these suggest that when it comes to the relevance of equality as a type of moral progress, the philosophical disquisitions regarding the genuine moral relevance (or lack thereof) of equality are neither here nor there. Equality is important because reducing inequality (usually) is. This is something that egalitarians and non-egalitarians can agree on.

5.3   Moral Status: The Expanding Circle

One of the most striking facts about human history is the amount of sheer hostility and violence that would ensue, with astonishing regularity, upon two groups of people's first encounter. One group arrives at a new shore, and almost immediately, bloodshed sets in (Livingstone Smith, 2011). Moreover, most of those involved find this completely unremarkable and indeed commendable, describing their victims as primitive savages or threatening vermin to rationalize their abuse, which almost always bears the marks of lustful, excessive cruelty rather than any kind of strategically necessary, efficient, and proportionate neutralization of a threat.

Such forms of violence have by no means disappeared. And more or less recent events, such as the Rwandan genocide or the Srebrenica massacre, remind us that those tendencies are all too easily reactivated. However, it is a very good sign that we look at such tales with horror rather than an acknowledgment that gratuitous torture and sadism are still the mundane facts of life they have been for the longest time. This surely has to do with the decline in our tolerance towards violence more generally (Pinker, 2011). Partly, however, it suggests that an increasing number of people are no longer as ready to vilify members of different ethnic groups or social identities as entities with little or no moral standing.

Developments such as these seem to form such a central aspect of how modernity has reshaped the moral infrastructure of society that some authors have suggested the "expanding circle" of moral concern to be the core of moral progress (Singer, 2011b). The by now familiar story is that moral standing—being the bearer of rights or entitlements on the one

and the subject of obligations and responsibilities on the other hand—gradually expanded over the past millennia, centuries, and decades from a privilege only enjoyed by a select few to a status granted to an increasing number of people. Over time, it was recognized that wealth, creed, sex, ethnicity, or indeed species membership does not carve the moral domain at its joints. Rather, these are morally arbitrary categories we choose to impose on the world, the only important unit ought to be the (sentient or sapient) individual.

The idea (and indeed the phrase "the expanding circle" itself) that inclusivist shifts constitute perhaps the most significant moral trend of modernity is due to William Lecky's 1869 *A History of European Morals*. Lecky argued that expansions of the moral circle are rarely, if ever, perfectly executed, and that the amount of moral concern afforded to different people is, as a psychological matter of fact, usually watered down when moving from family to compatriot to complete stranger.

In saying that I wish to deemphasize the importance of the expanding circle for an account of moral progress, I am not denying that the concept is to some extent useful for understanding social change for the better and the psychological constraints it frequently encounters. For instance, recent studies confirm that people's level of inclusivism has an important political dimension (Waytz et al., 2019). Conservatives, on average, draw the moral circle more tightly, assigning higher moral weight to a more narrowly confined ingroup rather than to all of humanity or the class of all sentient beings.

Recently, researchers have started to develop a "moral expansiveness scale" (MES; see Crimston et al., 2018) to measure this variability and account for individual differences in how many entities are included in one's moral circle, ranging from the "inner" (family and friends) over the "outer" circle (fellow citizens) to the "fringes" (so-called "power" animals such as fish or insects) and, finally, to entities falling outside of moral concern altogether (the "outer circle", such as "villains"). Phenomena such as compassion fade, and pseudoinefficacy (see previous chapter) show how surprisingly parochial our moral attitudes can sometimes be, bolstering the evoconservative suspicion that if moral progress is about inclusivism, then the prospects of moral progress may be more limited than we are inclined to admit. Often, compassion remains restricted to the vivid plight of just one individual. More generally, there is evidence that both developmentally and proximally, there are "centripetal" and "centrifugal" forces at work. Attachments to family and loyalty to a particular group constrict the moral circle, whereas a sense of impartial justice or aversion towards prejudice loosen it (Graham et al., 2017).

On the other hand, understanding what, precisely, the expanding circle amounts to is much less straightforward than it seems, and accurately describing how, precisely, the dynamic of the expanding circle unfolds is

even more complex. Many writers, in trying to figure out how moral concern is allocated, resort to implausibly biologistic notions which essentially imagine the expanding circle to be structured by Hamilton's rule, according to which one's degree of altruism is determined by a relationship coefficient. The idea being that this is how altruism must work, because that's the only way for it to enhance an individual's inclusive fitness (Birch, 2017). Michael Shermer (2015, p. 20), for instance, places the individual self (and, tellingly, identical twins) at the center of a person's moral concern, followed by that person's parents and siblings, followed by half siblings and grandparents, until it eventually reaches the whole ingroup, then the outgroup, the species, the animal kingdom, and so on.

However, there is simply no guarantee that the distribution of moral concern will be dictated by kinship in such a neat way. Many people arguably care more about their children than themselves, or more about themselves than about their identical twin, or more about their pets than their grandparents, or more about their friends than their siblings. Moreover, tying the expanding circle to morally arbitrary features such as degrees of genetic relatedness undermines much of its moral appeal. If the expanding circle is supposed to play a role in theorizing moral progress, its description should not make it seem regressive to begin with.

There are strong reasons to suspect that, when it comes to understanding moral progress, the expanding circle cannot be the whole story. Huemer (2016a, p. 1998) writes:

> Singer seems to assume that morality is exclusively or almost exclusively concerned with promoting others' interests, and that moral progress consists in the move from considering the interests of a small group to considering the interests of ever larger groups. That is indeed one important aspect of the moral progress that we have seen. No longer, for example, is it viewed as acceptable to make war on a neighboring society for no reason other than to capture their resources. But the notion of an expanding circle of moral concern is far from capturing all of the moral evolution that we observe over human history.

Huemer plausibly holds that the elimination of taboos on, for instance, premarital sex, the decline in social punitiveness, or the abolition of dueling and foot-binding cannot plausibly be described as expansions of moral concern.

However, even the core examples usually offered for increasing moral inclusiveness are open to interpretation. Perhaps cosmopolitan attitudes and a concern for animal welfare are best captured by expansions of moral status. Perhaps, however, they should be described as the demoralization of tribal boundaries and species membership. The increasing number of

entities enjoying moral status may well be a welcome by-product of more fundamental changes in what we regard as valid moral norms.

Finally, there is the normative problem of improper expansions. Not all expansions of the moral circle are unambiguously desirable and the expanding circle account by itself cannot distinguish between morally better and worse expansions of the circle. Many people, for instance, find it doubtful that corporations should count as people either morally or legally. We need independent moral reasons to figure out which of these expansions should count as good or bad.

5.4   Moral Status: The Contracting Circle

Sometimes, *contractions* of the moral circle are called for. I once spent a weekend in a small village in the German Mosel wine region. Our residence used to be a church. There we were 20 people, half of them were young children, cooking, drinking Riesling, and having fun, all facilitated by *Airbnb*. It is hard to imagine that over all of its centuries-long history, when sinister priests were threatening gullible peasants with eternal damnation to whip them into obedience, that church generated as much fun as we had on that one weekend. The point of this story is that there are sometimes tremendous benefits to a place no longer being considered *morally elevated*. The dignified status that church was once imbued with has faded away. It has now become a place of real value, whereas previously it was the site of excessive reverence.

In March 2017, the Whanganui River in New Zealand was recognized as a legal person, including the corresponding rights and responsibilities (Hutchison, 2014). I have no doubt that as a symbolic act of recognizing its importance to the local Māori tribe, this act was entirely appropriate, and it may also be an excellent means towards securing its ecological preservation. Similar protective measures are currently being debated regarding the legal status of Lake Erie. Still, it is hard to avoid the feeling that less animistic solutions could have been pursued. No offense lies in acknowledging that rivers and lakes aren't persons, however worthy of protection and sensitive stewardship they may be.

For most of history, membership in the ruling classes of society was assigned "ascriptively", granting elevated moral status, and the corresponding political and economic privileges that were supposed to flow from it, to a narrow circle of hereditary nobility.[2] The constitution of the Weimar Republic eliminated aristocratic titles and privileges in 1919; Italy gave up legal recognition of its nobility when it became a republic in 1946.

As in the case of expansions of moral status, the normative question is pertinent. Not all contractions of the moral circle are justified, and improper contractions of the moral circle surely happen. I would be surprised,

however, if proper contractions of the moral sphere weren't the rule, and if most of them didn't make sense. It is not an urban legend that animals used to be put on trial for their crimes (Kadri, 2007). Today, the execution of pigs and exile for donkeys have fewer and fewer defenders.

Membership in the moral community does not merely expand and contract; sometimes, entirely new circles pop up, or single circles split into several. Despite their often vast theoretical differences, many sociological theories of social change agree that one key development in modern societies is that they have become *functionally differentiated*. As a result, society has become radically decentralized: the economy, law, politics, religion, science, or art—all of which deal in their own meanings, codes, mediums of communication, modes of interaction, and criteria of success—have become thoroughly disentangled, whereas in premodern times, what's true and what's just, what's beautiful and what's holy would be regarded as fundamentally convergent, or indeed indistinguishable.

Nineteenth-century social theorists such as Tönnies or Durkheim described this as the move from community (Gemeinschaft) to society (Gesellschaft) or as the replacement of mechanical with organic solidarity. In the early 20th century, Weber and Parsons would echo this observation in terms of a separation of value spheres or an inchoate structural systems theory. During the second half of the 20th century, authors such as Habermas or Bourdieu began to speak of a differentiation of distinct social systems from the life world or the budding off of separate social "fields" of action. Niklas Luhmann, finally, saw the functional differentiation of modern society as largely complete, arguing that different social systems constitute autopoietic strings of communications which are based on their own respective binary codes (true/false, holy/profane, etc.) that are mere environmental noise for the neighboring systems.

It would be surprising if the major social upheaval characterizing modernity happened to lack any kind of moral flavor. Why would we think of the functional decentralization of modern society as progress? For one thing, the separation of previously interwoven social spheres may constitute progress because it rids us of a variety of conceptual confusions. *We* know today that what is funny need to be virtuous, or that what is true need not be beautiful. It is progress when morality no longer intrudes upon spheres it has no bearing on. The disenchanted world we inhabit may be a world in which we feel less at home. But it is a world that we understand better and conceptualize more accurately than the previous one.

Morality is often overly vigilant. This claim does not contradict the earlier one according to which expansions of the moral circle would often be sensible as well. The point is not that we are naturally good at drawing the circle of moral concern; rather, it is that we rarely withdraw moral status without good reason. The dehumanization of minorities does not

disconfirm this, because dehumanization tends to paint the outgroup as a positively vicious or disgusting enemy, rather than expelling the other into morally *neutral* territory.

As we have seen, Buchanan and Powell argue that expansions or contractions of the moral circle are sensitive to (perceptions of) outgroup threat and/or parasite stress. When the going gets tough, people become more stingy with their recognition of equal moral status. Under more favorable socio-economic conditions, they become more generous. At the same time, this account is somewhat incomplete in that it remains tailored to deciphering the dialectic between inclusivism and xenophobia. Many moral issues where the proper delineation of the moral circle is at stake do not seem modulated by outgroup threat at all. The US abortion debate, for instance, is about whether fetuses are persons; in bioethics, there are lively discussions regarding the moral status of PVS patients; business ethicists (and indeed lawmakers) disagree about whether corporations are people. In none of these cases does it seem plausible that which side one comes down on has anything whatsoever to do with the activation of our exclusivist tendencies through socioeconomic instability.

5.5 Liberty and Autonomy

Increases in liberty, freedom, and autonomy are another important dimension of moral progress. One doesn't have to go back to Rousseau to appreciate that people don't want to live in chains. Many, if not most, historical social movements were animated, at least in part, by a desire to be freed from domination, oppression, bondage, or the iron cage of repressive social norms and conventions.

Whether a particular instance of social change for the better should be described as a gain in welfare, equality, autonomy, or expanding moral status will often be difficult to decide. Fortunately, we do not have to. The struggle for female suffrage was a struggle for all of those things, and so was the resistance towards racial segregation. Real-world cases of moral progress are placed in an *n*-dimensional space of progressive dimensions, all of which can apply to the same process (to various degrees) at the same time.

The importance of increases in liberty and autonomy as a dimension of moral progress, and their central role in the vocabulary with which various social movements have articulated their demands, makes it all the more striking that neither Buchanan and Powell (2018) nor Pinker (2018) single out improved social freedoms as significant progressive concerns. It would of course be implausible to attribute to them the view that liberty isn't important. And yet they seem to subsume gains in human autonomy under inclusivist shifts or improvements in political participation and economic opportunities.

Liberty comes in (at least) two basic flavors (cf. Pettit, 1997). Negative liberty concerns the absence of external constraints, whether they are created by social practices or by the impositions of a stepmotherly nature. Negative liberty is liberty *from x*. Positive liberty concerns the presence of genuine opportunities to live one's life a certain way and to pursue various available options. This is the concept of freedom at play in the caustic observation that everyone is free to dine at the most expensive restaurant in New York City. Everyone is indeed free to do so, but most people, of course, actually are not. Correspondingly, progress in terms of liberty and autonomy can consist in the removal of external (natural or political) obstacles as well as in the provision of a richer set of meaningful courses of action.

According to Christian Welzel, moral progress along the axis of liberty and autonomy has three components: existential conditions, psychological orientations, and institutional regulations. Together, these components constitute the elements of human emancipation. Socioeconomic conditions determine the resources that are available to people in a society, which of them are scarce and why, how productively resources are used and how efficiently an economy is organized. Cultural conditions contain the corresponding values, and are oriented either towards conformity or towards self-expression and autonomy. Finally, the institutional framework specifies which freedoms and opportunities people are entitled to, which can range from very little to the rather expansive freedoms enjoyed by many citizens in the affluent world.

5.6   Fewer Bad Norms

All moral norms constrain—the set of morally permissible actions is a subset of the set of all feasible actions. Still, it is possible for societies to impose a needlessly fine-grained normative infrastructure on its members, resulting in too many action types being ruled out as morally transgressive. When there is such a reservoir of "surplus" moral constraints (Buchanan & Powell, 2017), societies can benefit from a type of moral progress that is often referred to as demoralization. Conceptually, it is not always trivial to distinguish between cases in which actions have been demoralized and cases in which people have made a switch from judging something to be wrong to judging something to be permissible. The former happens when something becomes genuinely morally *neutral*: people do not even recognize a practice as something that belongs within the purview of moral evaluation at all.

The US writer Julia Galef collects "unpopular ideas about social norms".[3] To a large extent, this list contains suggestions by other people

regarding practices that, in their view, ought to be demoralized. Here are some examples:

- Non-offending pedophiles should be more widely accepted by society. It's unfair to ostracize someone for a desire they were born with, and integrating them into society makes them less likely to cause harm.
- Incest that doesn't involve children, coercion, or procreation should be socially accepted.
- Overall, it would be a good thing to have a totally transparent society with no privacy.
- People in BDSM master-slave relationships should be able to be public about their relationship (wear a slave collar at work, introduce their partner socially as their "master", etc.). It's not fair to ask them to hide their sexual identity any more than it's fair to ask gay people to hide theirs.
- We should be *more* willing to lie to others, to protect our ability to be honest with ourselves.

A disproportionate amount of these suggestions concerns social norms surrounding sexual orientation and sexual relationships. This is no accident, since sexual morality is a treasure trove of demoralization-worthy pseudomoral norms that frequently police pleasurable acts between consenting adults. Other possible issues concern vote trading, swift corporal punishment as a more humane alternative to incarceration, or recreational drug use.

Many cases of demoralization illustrate the power of moral progress rather strikingly. Once an action type has become demoralized, the fact that it was ever considered morally loaded can become thoroughly invisible. Learning about what other people in different societies used to consider morally odious can have a jarring and indeed comical effect. For decades, the so-called *Motion Picture Production Code* (colloquially known as the Hays Code after Will H. Hays) would regulate what could and could not be shown on screen and how in Hollywood films. The list of transgressions is one of the most amusing archives of demoralization available. Among other things, it specifies that some things, such as "miscegenation", drug trafficking, or profanity, simply must not be dealt with in movies; other subjects, such as murder, rape, or kissing, had to be handled with utmost delicacy. Perhaps the most shocking thing about Hitchcock's *Psycho* was not Janet Leigh's early death in the shower, but the fact that she was seen flushing a toilet shortly before her demise. The ending of his *Rebecca* had to be altered from du Maurier's source material, because it was considered unacceptable that Maxim de Winter should get away with murder. A whole scene from Kubrick's *Spartacus* had to be cut because it very obliquely and tactfully alluded to homosexuality.

There are plenty of examples from the non-fictional world as well. Nowadays, being an atheist is in many places largely demoralized. In his *Letter Concerning Toleration*, John Locke made an extensive case for liberal religious freedoms. Not for atheists, however, which stood outside the "bonds of society" and, according to him, simply could not be trusted. Other great philosophers are known for similar blunders. Kant, who saw no objection to murdering children born outside of marriage, nevertheless condemned masturbation as a vice so severe that referring to it under its own name was itself immoral (MdS 425). Disability and other unchosen traits used to be considered divine punishment that justified ostracism and vilification. Today, we immediately recognize these bizarre moral views as obvious rationalizations of obsolete prejudices that have rightly been demoralized. As with expansions and contractions of the moral circle, there are proper and improper instances of demoralization. War demoralizes atrocity (Doris & Murphy, 2007) (which is one of the main reasons to avoid it), and ideologies demoralize murder and abuse.

Whether or not to buy life insurance is generally a question of personal risk management and of protecting one's family against financial ruin in the event of one's premature death. Life insurance, too, used to be seen as a scandalous act of putting a price tag on intrinsically valuable human lives, allegedly making wives desire that their husbands die (Brennan & Jaworski, 2015, p. 123). Martha Nussbaum reminds us of Adam Smith's view in *The Wealth of Nations* according to which singing or acting for money was considered a kind of "publick prostitution" (Nussbaum, 1998, p. 694).

Demoralization occurs when societies get rid of some of their moral norms and values. This is different from other forms of moral change, which often involve an inversion of moral polarity: things that used to be considered bad, such as casual sex or rebelling against one's parents, are come to be seen as good. In cases of demoralization, the moral domain itself is renegotiated, and loses one of its members. What used to be considered a morally significant issue becomes neutralized, losing its distinctively moral flavor.

Why do "surplus moral constraints" stick around? What are the forces keeping such arguably useless and even counterproductive prohibitions in place? Buchanan and Powell (2017) have developed a useful account of the various reasons why unjustified moral constraints can persist. Sterelny calls such harmful or useless norms SNAFUS—situation normal, all fucked up:

> [M]any cultures are prisoners of vastly disabling beliefs about the polluting power of female menstruation. So while in some aspects, the people of particular cultures respond to their world in an extraordinarily nuanced, subtle, and informed way, in others these very same people will seem barking mad. The Inuit superbly exemplify the power of the

processes of cultural adaptation to build effective responses to a harsh world. But . . . the Inuit had many (apparently) irrational and costly beliefs about their environment as well. In particular, they populated it with nonexistent dangers: giant fish and birds, ghosts and spirits. These misconceptions of their environment were not free: the Inuit altered their foraging patterns to avoid these supposed dangers, and invested in expensive ritual protections.

<div align="right">(Sterelny, 2007a, p. 318)</div>

In some cases, the respective demoralisanda will simply be due to evolutionary "hangovers", behavioral norms that used to make sense but no longer do. In other cases, dysfunctional or harmful moral norms continue to garner acceptance because they benefit small but powerful groups in society. Sometimes, such norms have failed to unravel earlier because the people subject to them are either unaware that others would rather get rid of those norms as well, or because there is a "first mover" problem that disincentivizes being the first to defect.

Demoralization will usually be a good thing, but it can also have costs that are difficult to appreciate, such as when, by removing its moral flavor from an action, useful social scaffolding of individual behavior is removed as well. Sometimes, the moralization of an issue can provide a helpful external social context supporting one's autonomous decision-making about drugs, crime, or socially and individually irresponsible behavior more generally. Demoralizing an issue, then, can put more of that burden on individual decision-making. Under conditions of demoralization, people can no longer benefit from the formal and informal sanctions imposed by society to discipline their behavior when contemplating the use of drugs (which can often be prudent). This is also an issue of social inequality, because if some people have, for whatever reasons, better capacities to make autonomous decisions without the social scaffolding supplied by the soft coercion of moral norms, removing that scaffolding will differentially affect those who are already comparatively disadvantaged in terms of their agential and deliberative capacities (Heath, forthcoming b).

5.7   More Good Norms

Some societies are overly permissive towards various actions which they improperly fail to flag as morally problematic. In those cases, *moralization*, rather than demoralization, is called for.

Moralization occurs when actions or traits that were previously considered morally neutral acquire moral flavor. As with demoralization, there is proper and improper moralization. One obvious type of improper moralization happens when believing in normatively inert empirical facts

becomes a matter of moral evaluation. How morally offensive a descriptive statement is can affect how much credence people are willing to put in it (Colombo et al., 2016). This seems to happen particularly easily when such factual beliefs become bound up with socio-political group identities (Kovacheff et al., 2018). Moralization can influence what people think about climate change, the effectiveness of vaccination, the biology of sex, the merits of various diets, or the safety of genetic engineering.

The psychology of moralization attracts increasing attention. Our minds can harness the force of powerful affective reactions such as disgust or indignation to convert mere likes and dislikes into morally charged values. This has already happened with smoking (Rozin et al., 1997). Other developments, such as the moralization of meat, are currently under way (Feinberg et al., 2019).

One of the main things social movements tend to aim for is to moralize various issues, and to get the rest of society to recognize that some things are in urgent need of moral opprobrium. Many activists are currently trying to highlight the moral offensiveness of so-called "cultural appropriation", such as the use of various symbolically meaningful signals or behaviors by people with no discernible connection to the cultural group those signals of behaviors originated from, such as white people styling their hair in corn rows or wearing Native American war bonnets. Such processes of (attempted) moralization are often hotly contested, where some people see the trivialization or theft of culturally significant items, others see a mutually fruitful and productive meshing of cultural practices with the potential for reconciliation and innovation—or at least a fun costume.

Many things are, or at least have been, desperately undermoralized. In a fascinating study, Samuel Fleischacker (2009) has shown that (roughly) until the middle of the 18th century, the idea that there could be such a thing as distributive justice—that is, a set of rights entitling everyone to a decent minimum of society's share of resources—was essentially unheard of. The plight of the poor was improperly moralized and demoralized at the same time: improperly moralized, because on the one hand the poor were seen as wretched and sinful, and at any rate necessary (for society to function) or inevitable (for being pleasing to God), while on the other hand, the thought that society could owe anything to its most vulnerable members was simply not on the table.

Sometimes, a special type of metamoralization is called for, as when subjects lack the very conceptual resources to identify something as morally odious in the first place (Fricker, 2007). For the longest time, patting one's secretary on the bottom may have been considered a perfectly normal fact of life (indeed, by everyone involved). It took decades of hermeneutic moral disclosure to appreciate what may be wrong with the sexually charged abuse of asymmetrical power. White collar crime remains severely

undermoralized today, despite causing orders of magnitude more damage than weed-smoking teenagers. Petty theft can make one a social pariah. Large-scale tax evasion does not cause any comparable stain.

There is ample potential for more good norms—proper moralization—and the space of moral discoveries seems far from exhausted. For instance, right now it would probably seem preposterous to everyone except a tiny group of people with a taste for contrarian proposals to suggest that there may a moral imperative to stop wild animals from killing and eating one another. Beyond the human sphere, the carnival of nature is simply beyond moral reproach. But is it? Who knows? And who knows which other "moral catastrophes" (Williams, 2015) are unfolding at this very moment, right in front of our eyes but blocked from view by the limitations of our parochial moral sensibility?

5.8   Improved Compliance

What good is a moral code if people are unable to follow it? In morals, as in life more generally, a bird in the hand is worth two in the bush, and even the noblest ideals and subtlest principles are of little value if the actual human beings who are supposed to be subject to them find them too difficult to understand, or too demanding to accept.

This suggests that there is a type of moral progress that doesn't straightforwardly concern any improvements in norms and values, or improvements in how well people are off, but progress in how smoothly people's compliance with whatever norms and values there are will be secured. Even if the quality of our moral rules remains stagnant, there can still be moral progress when more people comply with them more fully; that is, when there is increased compliance with good norms, lower compliance with bad norms, and generally better—less harmful, less alienating, and more easily absorbed—forms of enlisting loyalty to a society's ethical catechism.

Moral progress in mechanisms of securing compliance has two dimensions: one is about how easily a set of moral norms is acquired, and the other has to do with how austerely those moral norms, or rather their infraction, are policed.

Moral philosophy, where and when it draws on historiography, is replete with gruesome tales about how earlier generations used to punish deviant behavior, and the impressive inventiveness that was dedicated to coming up with novel, ever more exquisitely cruel forms of chastisement and torture that were enjoyed by our ancestors, either as recipients or as witnesses. The most famous of these accounts is perhaps Foucault's graphic description of the botched public drawing and quartering of Robert-François Damiens (Roth, 2014).

We can sincerely congratulate ourselves for having abolished, for the most part, such "cruel and unusual" punishments. On the other hand, it seems hard to deny that most modern societies incarcerate too many of their citizens, for too many things, and for far too long (Surprenant & Brennan, 2019). Moreover, the amount of luck shaping people's lives, including their misdeeds and crimes, seems so fundamental that the idea of anyone "deserving" to be punished becomes difficult to defend in the first place (Levy, 2015). The proper response may well be to do away with most forms of punishment, or to replace them with positive incentives instead.

A major obstacle to such progressive proposals for penal reform gaining any societal uptake is that people's punitive moral intuitions are robustly non-consequentialist (for a good summary, see Greene, 2008, p. 50ff.). People *say* they care about the effectiveness punishments, but ultimately, they do not, and remain insensitive to information about the deterrence yielded by various forms of sanctions. One of the primary goals on the progressive agenda should thus be to think about how to increase the tolerability of policy proposal that, despite their objective benefits, continue to upset people's deeply ingrained retributive inclinations. This is exacerbated, on the epistemic side, by the fact that most people believe that crime is getting worse, when the opposite is true, creating artificially inflated demand for harsher penalties (Duffy, 2018).

Defeatism remains unwarranted, however. Modern societies have proven that the path towards more humane and efficacious sanctions is open. Kitcher (2011, p. 140) recounts the memorable fact that, in some ancient societies, the punishment for raping a man's daughter was to rape *the rapist's daughter*. Since the publication of Cesare Beccaria's 1764 *On Crimes and Punishments*, most nations have abolished capital punishment or torture (Pinker, 2011, p. 144ff.).

Designing a moral code that will be reliably and smoothly absorbed creates a second important compliance problem. There is a reason why sophisticated consequentialists insist on the distinction between a moral theory's criterion of rightness and its decision procedure. "Maximize net happiness!" may be a sound moral rule, but it is far from clear that, in making moral decisions, *this* is the rule agents should consciously entertain. Moral progress can happen in how societies manage the cultural transmission of their normative fabric. A set of rules that is more acceptable is more likely to be accepted; a code that is more likely to be accepted is more likely to be followed. Good moral values that are more likely to be followed are moral progress. This creates a useful bias in the direction of norms and values that can be justified to those who are supposed to accept them: presumably, people are more willing to follow precepts they can see the point of. Purely arbitrary rules will be selected against, as people refuse to comply with them.

5.9   Improved Moral Knowledge

A final type of moral progress I wish to consider is reflective: moral progress occurs when our understanding of the nature of morality improves.

Consider, for instance, the so-called "is/ought"-gap. I am not a historian, of philosophy or otherwise, but it seems that before Hume, it had not even occurred to anyone that the transition from things are a certain way to things ought to be that way is blatantly fallacious. To name just one example for the purpose of illustration, the *Physiologus*, an influential didactic text from late antiquity, attempts to literally read off moral recommendations from the biology and behavior of various animals: the lion is strong and courageous, the hyena duplicitous, the unicorn (sic) chaste, so this is how you, too, must (or must not) be. And despite the popularity of trying to derive values from morally inert facts, it has been an enormous improvement in our understanding of the nature of morality to appreciate this chasm between the normative and the natural.

A related example that I have touched upon before is the differentiation of morality from other domains that it used to be amalgamated with. We now recognize that immoral acts can be beautiful, or that the facts of the matter may be morally unwelcome, or that the illegal can be virtuous and the legal heinous. This internal "rationalization" of independent spheres of discourse has improved our understanding of morality as well.

There is a better understanding of various thick moral concepts: marital rape used to be a conceptual impossibility. It should be considered progress that our understanding of various wrongs is no longer blocked by various conceptual decisions that entrench problematic social power structures and present them as conceptually inevitable.

A better understanding of the nature of morality will also, presumably, be helpful in further accelerating moral progress. Consider an analogy with scientific progress: our knowledge improves the more truths we acquire and the more falsehoods we cast aside. But the better we understand how reality works, the better we will become at reflectively improving our means of accessing it, thereby further boosting scientific knowledge. We can expect a similar effect in the moral domain. Moral progress occurs when social change is biased towards moral improvement. But the more we improve our understanding of morality, the more competently we will be able to target our efforts at boosting such improvements. The better we grasp the nature of morality, the better we will become at taking charge of our moral destiny. Maintaining a culture of moral reasoning and of making morality more and more transparent to ourselves—in civil society as well as academia—becomes a moral imperative in its own right.

5.10   Moral Progress: Towards a Systematic Typology

In some ways, a typology of moral progress will be parallel to a typology of forms of scientific of epistemic progress. In science, we essentially aim to reduce the number of falsehoods we believe, and aim to increase the number of truths we believe, modulated by relevance: we want to avoid cluttering our minds with useless truths, or to invest too much time and energy into disabusing us of harmless falsehoods. This is analogous to processes of demoralization and moralization, which aim to reduce the number of bad norms we follow and boost the number of good norms we adhere to. In science, it is progress when more types of relevant evidence are admitted (think: telescopes rather than scripture), and more forms of irrelevant evidence are ruled out (think: proper experiments rather than superstition). This is analogous to renegotiations of moral status (expansions and contractions of the moral circle). But the structural parallels won't be perfect, for science appears to be more monistic—it ultimately aims to promote knowledge—whereas morality is shot through with a plurality of values.

A typology of moral progress is supposed to provide a map through—more or less—uncharted territory. The success of this map can be measured in terms of the orientation it provides. After consulting it, do we know our way around?

Buchanan and Powell's typology of moral progress contains ten items (53ff.). They list:

 1. better compliance
 2. better moral concepts
 3. better understandings of the virtues
 4. better moral motivation
 5. better moral reasoning
 6. proper demoralization
 7. proper moralization
 8. better understanding of moral status
 9. better understanding of the nature of morality
10. better understanding of justice

The first thing to note here is that it seems possible to group several of these items together. List items 2, 3, 5, 9, and 10 concern improvements in our *knowledge about and understanding* of various morally important categories. List items 1 and 4 concern people's moral *agency*, their capacity to more successfully implement their moral knowledge in corresponding behavior. List items 6 and 7 cover improvements in moral *norms and*

values. List item 8 is about moral status. This suggests that, at the basic level, there are improvements in

1. moral knowledge
2. moral agency
3. moral norms/values
4. moral status

These are the four overarching types of moral progress.

5.11   Evolutionary Conservatism Again

The expanding circle of moral concern isn't the only type of moral progress that is subject to evolved constraints. Most types of moral progress discussed earlier are confronted with their own kind of evolutionary conservative challenge. The expanding circle faces limitations on empathy and altruism. Demoralization faces the problem of cultural evolution, and that we often don't know which social norms work, how they work and why, so that we better not get rid of moral norms without great caution. Something similar holds for moralization and the introduction of novel moral norms: we cannot anticipate how well they will work, and introducing new norms and practices can have major unanticipated and unintended consequences. This is not to say that challenging accepted norms and introducing new ones are never worth a try. Usually, we can experiment with new ways of living in limited and controlled environments, then scale up the ones that turn out to be actually livable.

There is a conservative challenge for well-being as well. We will eventually run up against the limits of how happy we can be, because evolution has not designed us to be content—people that are too easily satisfied suffer a competitive disadvantage. Evolution rewards the hungry. This is the dilemma at the heart of Freud's "civilization and its discontents": modern civilization works better and better, and makes people better and better off, but leaves us "discontent" because it imposes such strict constraints on discipline and self-control.

There is another version of this for liberty/autonomy: our evolved mind may not be ready to handle unlimited amounts of freedom. We become miserable because modernity has removed those comfortable constraints on what we can become, what we can achieve, whom we can marry, where we can live, and who we can be that earlier generations took for granted.

Notes

1. https://marginalrevolution.com/marginalrevolution/2020/02/who-is-wealthy.html
2. Thanks to Enzo Rossi for this example.
3. https://juliagalef.com/2017/08/23/unpopular-ideas-about-social-norms/

6 Mechanisms of Moral Evolution

What Drives Moral Progress?

Introduction

Contemporary theories of progress want to explain what moral progress is, what instances of moral progress there are, or which social developments ought to count as progressive and which ought not to. These are all tremendously important and interesting questions. What they have less to say about, however, is what *causes* moral progress (when it occurs).[1] This is at least somewhat surprising, for one would expect for those who care about how the world becomes a morally better place to show some interest in what it would take to make it so.

Buchanan and Powell (2018), for instance, have an elaborate, illuminating, and (I believe) largely correct account of one of the main causes of one major type of moral *regress*. The circle of moral concern can improperly contract, they argue, when insecure social conditions of economic turmoil and/or political instability allow people's exclusivist tendencies to thrive, or when powerful ideologies, spread by savvy demagogues, make them believe that such conditions obtain. They also develop, as we have seen, an impressively thorough account of what it means for the moral circle to expand towards greater inclusivity, and whether or not evolutionary explanations of our moral faculties can account for these inclusivist shifts. We hear much less, however, about what it is that *makes* the moral circle expand, when it does.

In this chapter, I will offer an account of the mechanisms of moral evolution. One of the main claims of this book is that moral progress is only insufficiently understood in terms of an expanding circle of moral concern, and that inclusivist shifts are at best one form of moral progress among many. An inquiry into the mechanisms causally driving moral progress will further buttress this thesis: the expanding circle cannot be the core of moral progress if and because there is a plurality of drivers of moral evolution, only some of which are related to inclusivist shifts.

This chapter has five sections. I discuss material mechanisms such as technological innovations or increases in group size (section 6.1), and

functionally novel forms of social integration (section 6.2), epistemic mechanisms such as improved knowledge about people and society (section 6.3), social movements (section 6.4), and experiments in living (section 6.5).

6.1   Energy Capture, Group Size, and Technology: Material Mechanisms

One simple yet profound fact that is hardly ever mentioned in theories of moral progress is that societies, in order to reproduce themselves—which seems a rather important thing for progress *or regress* to be possible (in that and only that sense, even moral regress is at least partly a good sign)—need to reproduce the material conditions of their continued existence. That this fact is so often ignored seems understandable, since most moral philosophers, I speculate, tend to be far removed from any actual handiwork, allowing them to indulge the fantasy that life consists mostly of what goes on in people's large heads, a mistake already corrected by Marx and his fellow 19th-century materialists.

Some hold that material processes, such as economic development, may engender progress, but may, at other times, also engender regress. It is a contingent fact, this argument goes, when material mechanisms lead to improvement rather than deterioration. While this is certainly possible, the evidence contradicts it:

> Evidence from around the world indicates that socioeconomic development *does* tend to propel various societies in a roughly predictable direction. Socioeconomic development brings occupational specialization, rising educational levels and rising income levels; it diversifies human interaction, shifting the emphasis from command—obedience relations toward bargaining relations; in the long run this brings cultural change, including changing gender roles, changing attitudes toward authority, changing sexual norms, declining fertility rates, broader political participation and more critical, less easily manipulated publics.
>
> (Inglehart, 2018, p. 42)

The historian Ian Morris argues that societies are largely, though perhaps not exclusively, shaped and constrained by the form of *energy capture* they rely on to sustain themselves. He identifies three main types: foraging, farming, and fossil fuels (Morris, 2015). These three forms of energy capture correspond to three distinct systems of values, each of which is connected to the fundamental moral concerns of equality (vs. hierarchy) and peace (vs. violence):

> I call the first of them [successive systems of human values, H. S.] "foraging values," because it is associated with societies that support

themselves primarily by gathering wild plants and hunting wild ani-
mals. Foragers tend to value equality over most kinds of hierarchy and
are quite tolerant of violence.

<div align="right">(Morris, 2015, p. 4)</div>

Their high degree of egalitarianism is a frequently noted feature of hunter-
gatherer societies (Boehm, 1999). As far as violence is concerned, there
seems to be an asymmetry between a remarkably peaceful and cooperative
intratribal life, together with strong intertribal violence and a tendency
for altercations. The other two systems recombine these two fundamental
evaluative orientations in novel ways:

> The second system I call "farming values", because it is associated with
> societies that support themselves primarily off domesticated plants and
> animals. Farmers tend to value hierarchy over equality and are less tol-
> erant of violence. The third system, which I call "fossil-fuel values," is
> associated with societies that augment the energy of living plants and
> animals by tapping into the energy of fossilized plants. . . . Fossil-fuel
> users tend to value equality of most kinds over hierarchy and to be very
> intolerant of violence.

<div align="right">(Morris, 2015, p. 4)</div>

These different configurations of values do not amount to radical disa-
greements about fundamental values (as Morris, falsely I believe, seems to
suggest: "moral philosophers who try to identify a one-size-fits-all, perfect
system of human values are wasting their time" (5)). Rather, they seem to
be a matter of emphasis: all societies have *some* preference for equality
and *some* tolerance of hierarchy, and all societies have *some* preference for
peace and some *tolerance* of violence. Different forms of energy capture,
then, merely shift around what is emphasized and how strongly.

When it comes to the issue of moral progress, an interesting question is
what the imminent exhaustion of fossil fuels as the main source of energy
entails for our future values. Since this is the only combination of values
not manifested so far, one could think that it will lead to social structures
which are both very violent and highly unequal—a bleak prospect for thus
fossil-fuel addicts like us. On the other hand, the various types of institu-
tional design I aim to sketch in the next chapter open up the possibility of
maintaining a state of basic equality and peaceful cooperativeness as well.

Perhaps the most important driver of moral progress comes from
increases in *group size*. The importance of this aspect seems to be under-
estimated by most political philosophies from socialism to libertarianism,
both of which articulate ideals of social integration which appear to be
designed for "camping trip" (see Cohen, 2009; Brennan, 2014) scenarios

in which the gains of cooperation can be secured through communal solidarity and/or voluntary consent.

Increases in group size tend to be—depending on one's perspective—a virtuous or vicious spiral. Population growth often stems from haphazard ecological or technological improvements that free up resources for an ever increasing division of labor. The resulting specialization achieved through functional differentiation, then, leads to further improvements in living conditions that lead to further growth, which—you get the idea.

Managing such growth, however, is anything but a no-brainer. Growing groups require new forms of social organization to secure continued cooperative success. The problem with this, as Robin Dunbar (1992) famously argued, is that our onboard cognitive, affective, and behavioral resources ate unable to handle the problem of how to engender social cooperation in arbitrarily large groups. Beyond a certain relatively low number of, say 150–200 people, groups need to find new ways of maintaining social cohesion, often involving novel institutional solutions. Stricter forms of hierarchy and punishment are usually the first step after leaving the narrow confines of tribal life (Mathew & Boyd, 2011); beyond that, more and more intricate forms of organization need to be implemented that will allow further "scaling up" to occur (Heath, forthcoming).

As far as different forms of moral progress are concerned, increases in group size have an elective affinity to progressive increases in welfare and improvements in norm-compliance. Henrich (2015, p. 185ff.; see also Hare, 2017) has described this as a process of "self-domestication": under conditions of fierce intergroup competition, culture/gene coevolution will favor individuals and groups with the ability to establish solidarity—which consists, basically, in sharing food and risks—on the basis of increased norm adherence and closer monitoring of norm violations in others (see Cosmides, 1989). This creates individuals that appear, for all intents and purposes, more "domesticated": more cooperative, less aggressive, more disciplined. In short: tamer.

For all its potential to be conducive to human welfare and cooperativeness, the growth of groups can also be a destructive force. Jared Diamond (2005) vividly describes a commonly found dynamic:

> Those . . . collapses tend to follow somewhat similar courses constituting variations on them. Population growth forced people to adopt intensified means of agricultural production . . ., and to expand farming from the prime lands first chosen onto more marginal land, in order to feed the growing number of hungry mouths. Unsustainable practices led to environmental damage . . ., resulting in agriculturally marginal lands having to be abandoned again. Consequences for society included

food shortages, starvation, wars, or disease, and society lost some of the political, economic, and cultural complexity that it had developed at its peak.

(6)

As a general diagnosis of why quantitative shifts in group size can lead to social collapse, Tainter (1988) suggested that the increases in the complexity of social organization which are required to sustain ever larger populations are subject to decreasing marginal returns. Growing groups require more bureaucracy and information flow, more energy and food, more organizational structure, more control and surveillance and, typically, more defensive resilience. At a certain point, such societies must invest more and more resources to achieve gains of the same size. Tainter, it is worth mentioning, did not consider a society's return to lower levels of complexity to be an unambiguously bad thing: "*under a situation of declining marginal returns collapse may be the most appropriate response*" (198).

A third material mechanism driving moral progress is technological innovation. Jürgen Habermas, for instance, sketched the important role widely available newspapers and an associated culture of salons among educated citizens played in the early formation of a public sphere for civil discourse (1991). It seems plausible to suggest that it would have been considerably harder to mine the emancipatory potential of a newly established self-assured bourgeoisie without those technological underpinnings. In an influential paper, Greenwood et al. (2005; see also Cavalcanti & Tavares, 2008) describe the washing machine and other large household appliances as "engines of liberation", and indeed the actual success of the women's rights movement in terms of political participation and inclusion in the labor force is very closely aligned with the widespread adoption of such devices. The claim that the struggle for female equality was partly accelerated by technological innovation—which was, presumably, never supposed to have such an effect—is not to minimize or belittle the importance of social movements claiming long overdue civil and political participation. Political movements articulate their goals and fight for their legitimate interests; it is not a stain on their victory to also note its technological catalysts.

In his *It's Better Than It Looks*, Gregg Easterbrook (2018, p. 18ff.) shows how technological innovation, particularly in agriculture, significantly improved the input/output ratio of farming, thereby decreasing the need for land. This drastically reduces the incentive to go to war to acquire land. One should not forget that one of the main official justifications cited by the Nazis for invading Poland and later Russia was territorial: "Lebensraum" in the East was supposed to be the main spoil of war. Inventive technologies, together with specialization and trade, have made this incentive *de facto* obsolete.

Changes in the way energy is captured, increases in group size, and technological innovation all potentially lead to moral progress, in particular for the way in which they allow greater (individual and aggregate) well-being to go along with, and be facilitated by, increased human cooperativeness. These processes, however, do not always lead to unambiguous improvements, and are hence subject—as most forms of progress are—to their own dialectic of the enlightenment. Once the forces of technological innovation are unleashed, they often seem to develop a life of their own which can quickly spiral out of control; all of sudden, as in Goethe's *Sorcerer's Apprentice*, technology appears to develop its own interests: the machine, as Horkheimer suggested in his *Eclipse of Reason*, has dropped off the pilot, rushing blindly through space. Indeed, the earliest forms of technological innovation that would eventually make modern life possible have been described as "the worst mistake in the history of the human race" (Diamond, 1987). It seems that the main effect of the invention of sedentary agriculture over the course of the Neolithic Revolution was to make taxation possible—bags of harvested grain are eminently countable and thus seizable (Scott, 2017)—which enabled the formation of early states and forced carefree, adventurous hunter-gatherers into a life of bondage and servitude. The next 10,000 were a self-inflicted nightmare for us; only recently have some places started to catch up with pre-Neolithic levels of leisure and well-being (Widerquist & McCall, 2015).

Despite their inherently ambiguous character, new forms of energy capture, increases in group size, and technological innovation can engender, accelerate, or facilitate moral progress in a variety of ways. Since intellectuals are naturally biased in favor of overemphasizing the importance of intellectual processes in promoting social change, I tried to restore some necessary balance. A comprehensive account of the drivers of moral progress, however, must not ignore the equally crucial role epistemic mechanisms such as reasoning or the proliferation of knowledge and information have to play as well.

6.2   Social Integration: Functionalistic Mechanisms

The fundamental challenge for human societies is how to achieve sociality in the first place: why is there something social, rather than nothing? This is often called the problem of *social integration*. In principle, it is not unimaginable for there to occur no social integration at all, and for humanity to remain a mere assembly of individual entities which never manage to integrate their individual goals and intentions into a supraindividual structure.

An underappreciated driver of moral progress can be found in the *switch* to new forms of social integration. Such switches are rarely, if ever, normatively neutral; when social integration is achieved in novel ways, novel

moral expectations tend to arise, which end up biasing the process of cultural reproduction further downstream.

One of the main switches of this sort can be seen in the move from premodern to modern societies. Here is one version of the story of how distinctively modern societies came to be: in premodern societies, social integration is largely achieved on the basis of an unconscious and implicit reservoir of shared values, knowledge, practices, and norms. This conglomerate serves as a background that can, for the most part, be taken for granted: individual agents can avail themselves of an assumed consensus of which things are done, when, how, and why; this background rarely needs to be problematized and can be treated as given. Only when cooperation ceases to run smoothly does social integration have to be effortfully secured to restore undisturbed interactive success. With the switch to modern societies, however, the importance of such a stable set of shared knowledge, values, and meaning starts to dwindle such that more and more frequently, social integration can no longer be presumed and needs to be explicitly *achieved*. Increasingly, *language* takes over as the means of securing cooperation and cultural reproduction.

The key thing to file away about this switch is that it is not morally innocent. When communication increasingly has to step in to reduce the friction created by an erosion of the background support of the "life world", then social reproduction becomes dependent, for functional reasons, on mechanisms that are inherently more egalitarian and more flexible—that is, more progressive—than social integration that is dictated by authority and tradition. In the game of giving and asking for reasons, might and longevity don't count (as much), and neither do class, gender, or revelation. The increasing need for linguistic reproduction thus naturally engenders a self-reinforcing dynamic towards modern social *mores* (Heath, 2014c).

Groups also evolve along the axis of *designed* versus *spontaneous* social interaction. The invisible hand of the market remains the paradigm for uncoordinated, large-scale cooperation. Smaller groups such as tribes or bands can intentionally control their social roles or the distribution of resources; massive modern societies outstrip human capacities for central oversight. When this tipping point is reached, societies either regress to simpler forms of social organization or manage to implement social institutions that allow chains of cooperation to emerge despite the fact that no one is in charge.

Third, societies progress by renegotiating the ways in which the costs of people's actions "show up" in the social accounting: "Social evolution is partly a process of perceiving new externalities and devising institutions to internalize them" (Schmidtz, 2008, p. 206). The classical, but by no means the only, way in which this can be done is via individual property rights. Regimes of property take resources out of the ungoverned commons where

third-party effects can (seemingly) disappear by being transformed into a diffuse burden for the many. When people start staking claims, these previously invisible costs (and benefits, in the case of positive externalities) reenter the spotlight. Structures of central authority are another way of internalizing externalities because the Leviathan is uniquely positioned to encourage or discourage behavior through incentives and threats.

Finally, there are the grand narratives spun in 20th-century social theory which try to describe the developmental trajectory of societies, and in particular the rise of modernity, in terms of an increasing functional differentiation of social systems within society. From very early on in the development of modern social science, theorists such as Tönnies, Simmel, Durkheim, Weber, Parsons, or, more recently, Luhmann, Habermas, and Bourdieu, tried to characterize what is distinctive about modern society in terms of an evolution of increasingly separate social realms—value spheres, systems, fields—that follow their own internal logic. A main feature of premodern societies, on this account, is that the main social subsystems such as the law, politics, religion, art, science, or the economy are deeply intertwined. Rulers rule *dei gratia*, knowledge needs to be reconciled with what is politically expedient or religiously acceptable, and what is beautiful is also virtuous (and vice versa). Modern societies upend this state of literal confusion and allow for the evolution of separate social spheres that operate on the basis of their unique specific codes and functional demands.

6.3　Knowledge and Information: Epistemic Mechanisms

Moral progress is driven, to an important extent, by adjustments in the material basis of social reproduction. So far, so materialist. This concession does not entail, however, that more ethereal intellectual changes—scientific revolutions, new insights, more widely available information—do not have an equally important role to play. If people generally want to do the right thing, and come to form new moral beliefs regarding what is and isn't right, then the adoption of such new beliefs may eventually lead to social change in the direction of moral improvement.

One driver of moral shifts that clearly consists in improvements along a genuinely epistemic dimension is the removal of inconsistencies in our moral outlook. There are no true contradictions, so believing one can never amount to knowledge. Achieving consistency is thus always at least *necessary* for gaining knowledge, moral or otherwise.

Consistency reasoning seems to have an obvious contribution to make to moral progress as an engine of moral extension. Richmond Campbell and Victor Kumar (2012; see also Campbell, 2017), for instance, place a strong emphasis on how what they call "moral consistency reasoning" helps engender individual and social moral improvements. They recount

the story of Jan Baalsrud, a Norwegian resistance fighter seeking refuge from the Nazis in a family home in a rural village. The mother of the family, after realizing that there was no essential difference between the forlorn young man and her own son Marius, agreed to provide shelter and protection for Jan. The same dynamic frequently plays out at the social level: more and more, people come to realize that there is no essential difference between white and black people; consistency dictates to treat like-cases alike, which makes it difficult to defend an institution such as slavery (if, indeed, extending slavery to white people remains out of the question).

On the other hand, this apparent elective affinity with expansions of the moral circle towards more inclusive membership in the moral community cannot be driven by the removal of moral inconsistencies alone. For even if there is no relevant difference between my backyard and your backyard, that fact by itself does not tell me whether I should start mowing your lawn, or stop mowing either (Kahane, 2014).

In Campbell and Kumar's model, consistency reasoning represents a joint effort of System I and System II. The fact that there is no morally relevant difference between two cases—or individuals, or groups—is detected at the emotionally resonant System I level, the cognitive system in charge of quick, holistic, intuitive/automatic processing. I deem it monstrous not to save a child from drowning because it may ruin my new suit, but merely eleomosynary to give some money to charity to save a starving child's life. When considering the two cases, however, my intuitive system fails to detect any morally relevant difference between these two cases. This engages System II—the slow, conscious, and effortful type of cognitive processing—which works towards removing the inconsistency in whatever direction seems most appealing.

Campbell and Kumar appreciate this point:

> As we have already made clear, our two responses are not by themselves inconsistent. There is no contradiction in judging that one is morally obliged to save a nearby drowning child but not obliged to save a distant starving child even if the costs are comparable. The problem is, rather, that the two responses together seem morally indefensible because we see no morally relevant difference between them. The inconsistency arises out of our responses to the two cases in the context of our basic substantive moral commitments that inform us about what is morally relevant.
>
> (2012, p. 296)

However, this description of the process generates trouble, because it's not clear whether there is any genuine inconsistency to speak of in the first place. Note that it isn't logically or conceptually incoherent to suppose that

I should not, morally speaking, abandon my son to Nazi thugs while at the same time maintaining that there is nothing at all morally wrong with abandoning someone else's son to such a dire fate. Huemer (2016a) puts this point clearly:

> How could *reasoning* lead one from concern for the interests of other members of one's tribe to concern for the interests of individuals outside the tribe? Exactly what would the argument be? . . .
>
> 1. It is morally obligatory to consider the interests of other individuals within the tribe.
> 2. There is no morally significant difference between those in the tribe and those outside the tribe.
> 3. Therefore, it is also morally obligatory to consider the interests of other individuals outside the tribe.
>
>
> But where would premise (2) come from? Not logic alone; there is no incoherence in holding that tribal membership is morally significant— or more precisely, that the relation of *belonging to the same tribe* is a morally significant relationship between persons
>
> (2016a, p. 1997)

Huemer himself, of course, doesn't see anything wrong with this, because as a moral realist, he has no problem admitting that, when we recognize something as morally (ir)relevant, we acquire moral knowledge by embracing the objective moral facts via rational intuition. For the less realistically inclined, however, it becomes harder to defend consistency reasoning as a genuinely epistemic mechanism.

Moral progress should count as being epistemically driven if it is tied to processes of *rational learning*. Rational learning approaches to moral judgment and cognition have attracted increasing attention over the past years (Cushman et al., 2017). I have developed and defended such an account myself (Sauer, 2017b). In particular, the extent to which our moral beliefs are shaped by processes of rational learning seems to offer the most promising response to anti-rationalist challenges to moral judgment. If moral judgments are sensitive to and update on morally salient information, this can provide at least somewhat rational credentials. As far as mechanisms of moral evolution are concerned, describing moral judgment in terms of rational learning occupies something of a middle ground between mechanisms of transmission and retrieval and mechanisms that are themselves responsible for pushing the frontier of moral progress forward.

Rational learning theories have been used, for instance, to adjudicate between empiricist and (roughly speaking) nativist approaches to the cognitive roots of the relatively sophisticated moral principles our moral judgments seem to reflect. Consider, for instance, the fact that most people seem to draw a distinction between harms that were brought about as a means towards an end and harms that were brought about as a foreseen but not specifically intended side effect of an action. People are typically not able to consciously articulate this "doctrine of double effect" (Cushman et al., 2006) nor do they, it seems, have access to evidence that is both rich and unambiguous enough for them to retrieve the rule from experience (arguments of this sort are often called "poverty of the stimulus" arguments). This has led some to posit a hard wired "universal moral grammar" (Mikhail, 2007) to account for the patterns we find in ordinary moral judgment.

Others, such as Nichols et al. (2016), have shown that empiricists about moral judgment can explain how subtle moral principles can be acquired from scant evidence. The key to the solution consists in mechanisms of Bayesian learning, combined with sufficiently specific information to apply these mechanisms to. Generally speaking, people see a moral difference being doing something and merely allowing something to happen (or to continue). However, they are never really verbally instructed that there is such a principle, for instance during early childhood; nor is there sufficient evidence for children to extract this principle empirically from the observable evidence; it must, therefore, be innately prepared. Or so it seems—for as Nichols et al. have shown, such a principle *can* be learned from the available linguistic evidence if people automatically update on it according to simple rules of Bayesian learning.

Here is, roughly, how this works: as it turns out, since children are almost always told what they, specifically, ought to or ought not to *do*, rather than what to *allow* to happen or to be the case, they have reason to think that behavioral rules pertain to what actions rather than states of affairs more generally. From this, children can rationally infer that the rules likely govern specifically actions, just like they could infer that a repeatedly thrown (unloaded) die that doesn't turn up anything above the number 4 after being thrown ten times is probably a four-sided die. If it turns up a 5 at the 11th, the belief will be updated accordingly, namely to the next "smallest" hypothesis. Nichols et al. were able to show, based on linguistic and experimental evidence, that a variety of moral rules can be learned from limited information in just this way, and in surprisingly neat accordance with basic principles of Bayesian updating. The distinction between outcomes that are intended, foreseen, accidental, or not caused by any agent at all can be represented as a nested structure with intended outcomes as the narrowest in the space of hypotheses; if all verbal instructions

they receive, then, end up referring to actions rather than any of the larger hypotheses, learners will rationally conclude that this is what moral rules are about, thereby acquiring the distinctions between intending and foreseeing doing and allowing, and so forth.

The prospects of such rational learning accounts have also been demonstrated in the context of dual process theories of moral cognition. Such theories are frequently deployed to account for the contrast between so-called deontological and consequentialist moral judgment—roughly, judgments that gauge the moral worth of an action in terms of the outcomes it produces and moral judgments that are not exclusively outcome-oriented, emphasizing the inherent wrongness of some actions instead. Researchers such as Joshua Greene (2014) famously claim that this contrast tracks a morally irrelevant difference that is picked up in by alarm-like affective response (Berker, 2009; Sauer, 2012; Kahane et al., 2011). In so-called "Trolley" scenarios, people tend to endorse the consequentialist option only when such an option isn't emotionally conflicting or when they take the time and effort to override the temptations of non-consequentialism.

Railton (2017) and Cushman (2013), however, insist that this difference is more likely to be due to a difference between two different types of moral learning. In model-*free* learning, subjects evaluate a scenario on the basis of a cached (negative) response pairing a situation with an action they are broadly familiar with: pushing an innocent man to his death? No way! Model-based learning, on the other hand, operates on the basis of more or less fully articulated causal models of a situation which branch out into a model of the available options and the causal pathways linking them. When no cached response is available, such model-based processes will be activated, leading to a more outcome-oriented decision tree tailored to the specifics of the situation at hand: pulling a lever to redirect a trolley? Seems fine. What else happens down the tracks? The key thing to realize about both rational learning accounts is that they can vindicate the rationality of a variety of central distinctions whose moral relevance has been disputed.

The main problems for theories that couch the rationality of moral cognition in terms of learning are that in order for learning processes, however inherently rational they are, to render their outputs rational, the *inputs* these processes draw on must be independently justified. This is a limitation that no rational learning story can overcome on its own. What you end up learning is only as good as the information you can avail yourself of. The same mechanisms that allow you to improve your chess playing abilities also allow you to make your homophobic prejudices ever more fine-grained, subtle, and sophisticated. This does nothing to show that your original prejudice has anything at all to commend it. Rational learning processes update on morally relevant as well as morally *irrelevant* information (Gjesdal, 2018), which can lead to systematic distortions. Moreover,

once such overlearned responses are confronted with unexpected stimuli they are not prepared for, they can systematically misfire (Greene, 2017). Not all hope is lost, however, because if there is such a thing as moral progress, then we can expect this problem, though real, to be overcome with time: if our learning environment improves, then our learning outputs, though inevitably imperfect, will improve as well (Gjesdal, 2018, p. 470).

Moral learning doesn't always have to be slow and gradual. Sometimes, so-called "belief shocks" can rapidly relieve a society from the tyranny of regressive norms (Bicchieri, 2016, p. 44). When such a belief shock happens, a practice falls apart because people suddenly realize that few people want to follow the norm and are thus relieved from the trappings of "pluralistic ignorance". To maintain a social norm (of which moral norms are a proper subset), one needs two things: people need to believe that all (or most) of the other members of their group follow the norm and they need to believe that all (or most) of the other members of their group believe that one *ought* to follow the norm. When both of these beliefs are suddenly—and publicly—undermined, harmful social practices can rapidly come undone. One of the main reasons many people engage in contemptible practices such as genital cutting or foot-binding—rituals frequently performed against the protest of their uncorrupted natural disinclination to harm their wives or daughters—is because they believe that other people will do so and believe that one ought to. When it becomes publicly known that fewer people practice female genital mutilation or engage in drug abuse on campus, the pressure to conform can wither away. This is a fact that can be deliberately exploited by activist groups who can induce such beneficial belief shocks by making the respective information publicly available.

Sometimes, the epistemic dimension of moral progress simply consists in learning new facts by becoming acquainted with them. According to the contact hypothesis, for instance, real-life exposure is the best way of reducing discriminatory attitudes towards marginalized groups and of avoiding the pitfalls us/them thinking more generally. Prejudices, once acquired, tend to be reinforced by a biased processing of the encountered evidence:

> If much social prejudice is learned in these ways, then it should in principle be possible to unlearn it by changing people's experiential sample. . . . [T]he most effective ways of overcoming such bias involve "contact" processes that go beyond mere exposure of groups to one another, and include activities in which individuals from different groups co-participate in activities that have a common goal, draw upon the contributions of each, and involve taking the perspective of others. . . . In consequence, attitudes toward questions like gay marriage, and measures of implicit bias toward gays, have undergone a dramatic change—especially among the young, who have lived their entire lives in an atmosphere of more open recognition

of sexual orientation. . . . That a centuries-old form of prejudice—an "us" vs. "them" bias thought to be deeply rooted in our psyches—could undergo such rapid change is a tribute to the power of learning, even from "old" evidence that has been recategorized, if given a chance.

(Railton, 2017, p. 179f.)

Some processes of moral development are perhaps best described as *semi*-epistemic because they involve genuinely truth-conducive processes only indirectly, or indeed somewhat paradoxically. One case in point is rationalization. Despite its bad reputation, rationalization can be a progressive force. Ignoring some nit-picky definitional details, rationalization consists in providing an account of one's thoughts, motivations, or actions that purports to be justificatory and explanatory, but actually isn't. The reasons I cite, though perfectly good reasons in principle, aren't causally "effective" (Sauer, 2017b).

Now the good news about rationalization is that, like most forms of self-deception, it tends to be sincere (Greenspan, 2015 and see also Cushman, 2020). For rational agents, this can create useful downstream pressure to bring their actual thoughts, motivations, and actions in line with the confabulated reasons they cited for them:

> Offering sincere justifications of one's actions can contribute to moral progress by creating pressure to become consistent with those justifications, and hence to become better over time, even when the rationalized justifications misrepresent one's motivation. This benefit explains a significant way in which rational moral progress is possible even when actors rationalize their own actions.
>
> (Summers, 2017, p. 100)

This can occur at the social level as well. A socially prevalent purported reason for continuing the political disenfranchisement of women may have been that this allegedly makes them better off, "protecting" them from the burdens and hardship of democratic decision-making and public office, instead reserving for them the domestic comfort and simple pleasures of keeping a house and rearing the children. Now if it turns out that this arrangement doesn't actually end up protecting women, instead making them more rather than less vulnerable to abuse and the infringements of their rights and interests, the proffered reasons, though originally confabulated, can come back to haunt the advocates of gender inequality in unexpected ways, who now find themselves uncoercively coerced to follow the force of their arguments where it leads.

So far in identifying epistemic mechanisms of moral evolution, I have operated with a relatively narrow concept of the epistemic that focuses on paradigm instances of consistency, learning, belief, and reasoning (Tam,

2020; see also Stanley et al., 2018; Horne et al., 2015). However, we do not have to exclusively rely on such an "intellectualist" understanding of moral knowledge and its relation to social change. Elizabeth Anderson, for instance, has shown how practices of *contention* can, within a pragmatist understanding of rational moral updating, often be more effective at promoting the acceptance of true moral beliefs than the traditional, argument-focused approach of most Anglophone philosophy (Anderson, 2015, 2014).

These methods, according to Anderson, largely consist in uncovering abstract moral principles or in bringing particular intuitions about cases and more general moral principles into reflective equilibrium. Sometimes, they also involve "table turning" exercises such as the golden rule or the veil of ignorance. One big problem of this approach has to do with its reliance on outlandish thought experiments where nothing is at stake. Rather than responding to actual, pressing challenges arising from the conditions of real people deliberating and trying to live together, such forms of argumentation idly dabble in the *a priori* of what holds in all possible worlds, rather than our one actual one. As a result, they accomplish nothing and convince no one.

Moreover, and perhaps more importantly, one can see by looking at actual instances of progressive moral change such as the abolition of slavery that such academic forms of reasoning were oddly powerless in aiding the abolitionist cause. When abolitionists pointed to the golden rule, defenders of slavery showed how to reconcile it with social inequality; when abolitionists criticized slaves' chains, defenders reminded them of similar practices in the military; when abolitionists objected to slaves' dependency and disenfranchisement, defenders drew comparisons to women's rights (or lack thereof); when abolitionists complained about violence and abuse, defenders pointed out that many children were subjected to forced labor and physical punishment, too. In short: consistency reasoning and standard appeals to moral principles or rights frequently did next to nothing to decisively challenge the status quo.

Practices of contention did a better job:

> Contentious practices span a spectrum from pure moral argument at one end, to riots, war, and other violent acts on the other. Between pure argument and violence is a wide range of contentious activities that are more or less disruptive of habitual ways of life, from petitioning, publicity campaigns, theatrical performances, candlelight vigils, litigation, and political campaigns, to street demonstrations, boycotts, teach-ins, sit-ins, picketing, strikes, building occupations, and other forms of civil disobedience.
>
> (Anderson, 2015, p. 32)

These practices count as genuinely epistemic because they are uniquely successful at correcting morally pernicious biases against, in this case, black slaves. White abolitionists tended to make the mistake of trying to activate empathy and combat "hard-heartedness" (37), the idea being that a lack of sympathy must have been the main cause of slavery in the first place, and that by promoting such fellow feeling, the main obstacle to ending the plight of slaves could be removed. But empathy (or "pity"), it turned out once again, is fragile, easily exhausted, and so failed to do the trick. What proved utterly more effective was not to inspire contagious suffering, but to *demand* respect, *resist* oppression, and *seize* control. Moral recognition must be claimed, and in making such claims, people performatively demonstrate that they *merit* it.

Anderson's account of the epistemic clout of practices of contention shows that the boundaries between the epistemic and non-epistemic are fuzzy. In many cases, the process of gaining new insights will be deeply entangled with forces of social upheaval that make these insights heard.

It is clear that epistemic drivers of moral progress exist. The real question is how *effective* they are. In order for them to be effective, there must also be mechanisms translating these epistemic insights into real change of social practices. The important question is thus not about the existence of epistemic improvements along a morally salient axis, but about their *uptake* and implementation.

6.4   Crisis and Struggle: Social Movements

Moral change can also be driven by *social movements*. Herbert Schnädelbach once remarked that distinguishing the head from the body and separating the head from the body are two very different things. In the same non-homicidal spirit, I want to analytically distinguish social mechanisms of moral progress from epistemic or material processes, which is not to say that these are metaphysically *separate* processes that aren't deeply intertwined. The adoption of a new technology, for instance, has obvious epistemic, material, and social dimensions. These dimensions cannot be separated, but they can be distinguished: it remains theoretically useful to discuss specifically social aspects on their own terms.

Social change is often mediated by so-called "cultural attractors" (Sperber, 1996),[2] which are best seen as probabilistic filters that bias the copying process that ends ups patterning the distribution of cultural *know how* between two learning cohorts. Suppose that one group—call them "parents"—tries to get another group—call them "children"—to adopt two behavioral rules. One says that one must always wear a top hat during dinner; another says that one must never blow one's nose into the table cloth. Which one is more likely to get "picked up" by the following generation?

When we study the transmission of behavioral norms, those that tend to be supported by powerful affective reactions such as disgust are more likely to survive and spread (Nichols, 2004) and to be perceived as "fitting" (Sauer, 2011). Emotional support serves as a cultural attractor.

This "epidemiological" approach to studying cultural change can be applied to the moral as well. Liberal institutions, for instance, can be described in terms of cultural attraction (Cofnas, 2019; Hopster, 2020). It is not a stretch to suggest that broadly liberal attitudes and institutions such as equal rights and opportunities for everyone can serve as a cultural attractor for large groups of people. The promise to be treated with dignity and not be subjected to arbitrary discrimination or gratuitous violence is surely emotionally resonant enough to bias cultural reproduction in a liberal direction. Over time, more and more people will be drawn to norms and dispositions that favor empathy and equality over aggression and subjection, in particular the relatively marginalized and downtrodden. This is the kernel of truth about Nietzsche's "slave revolt" in morality.

Social movements occupy something of a middle ground between individual attitudes and institutions of the state. Like the former, social practices are often subtle and informal; like the latter, they are broad in scope and have many members. As such, social webs can play a particularly central role in maintaining injustice and oppression. Unfortunately, the patterns of meaning and salience supplied by culture sometimes figure in pernicious ideologies that justify social oppression and cloud its real origins with stabilizing illusions (Haslanger, 2017). Ideology critique, then, aims to remove this false consciousness by undermining its poor epistemic credentials and highlighting its role in justifying social injustice. Unfortunately, escaping an ideology takes more than removing the epistemic mistake due to first mover costs.

Needless to say, not everyone is fond of such attacks on their privileges, so burgeoning social movements are often perceived as signs of crisis. Crisis was of course Marx's primary explanation for why and how moral and political changes occur at all. Each historical period tends to be characterized, roughly, by a set of rights, customs, practices, and values combined with a particular configuration of material reproduction, that is, a way of running the economy. Time and again, the forces unleashed by the economy start to come into tension with the system of legal entitlements and social practices that accompanied the previous stage of material reproduction. The two no longer "fit" together, and all that is solid melts into air.

In many cases, the cultural crises that lead to such moral transformations are ushered in by social *struggle*. Such forms of mobilization by groups whose interests are not properly represented in social institutions play an important role in Railton's naturalistic account of moral realism (1986, p. 191ff.). Just as an individual's poorly informed or short-termish wants may

fail to reflect their more enlightened interests, so can the actual arrangements found in a society fail to reflect the overall interests of its members.

Most notably, perhaps, social struggle plays a foundational role in Axel Honneth's (1996) critical theory of recognition. Honneth notes that although concerns for physical well-being are of course far from trivial, most social movements follow a "moral grammar": their demands are couched in terms of a claim for increasing *recognition*, rather than mere material betterment. Disenfranchised individuals and excluded groups don't tend to just claim a bigger share of the pie. Instead, their motivations to spark conflict and their willingness to accept risks and sacrifices are driven by the moral goals of ending humiliation and overcoming disrespect. The normative infrastructure of this struggle for recognition can, according to Honneth, provide a template for sketching a vision of a society that embodies the conditions of a good life: collective practices in which the social spheres granting the basis forms of intersubjective recognition—love, respect, and solidarity—are accessible to everyone.

At the more violent end of the spectrum, social struggles can turn into *revolutions*. Jamieson (2017) notes that moral revolutions are often spearheaded by individuals with the perspective of what he refers to as a "privileged outsider": billionaires with an ascetic lifestyle campaigning against climate change or individuals from the more affluent strata of marginalized groups such as Martin Luther King.

Revolutions are arguably one of the most important mechanisms of moral evolution, yet from a normative perspective, they are surprisingly hard to make sense of. Christine Korsgaard, for instance, endorses Kant's paradoxical stance "[R]evolution is always wrong. Yet sometimes the good person finds she must rebel" (2008, p. 234). Allen Buchanan (2013) has shown that if we model revolutions—think of the Arab Spring—after widely accepted *jus in bello* principles, they are almost always confronted with the impossibility of reconciling the twin requirements of proportionality and reasonable chance of success: when revolutionaries refrain from perpetrating heinous and immoral acts as part of and as a means to their revolutionary ends, they are unlikely to succeed; to succeed, they must be inhumane. This dilemma holds especially sharply when the *jus ad bellum* conditions such as an obviously tyrannical government are clearly satisfied.

6.5   New Norms: Experiments in Living

If moral progress is largely an evolutionary process—that is, a matter of variation and selection of (successful) alternatives—then the development of such alternatives becomes a crucial part of social change for the better.

In almost all cases of important moral developments, the proposed changes had to be tried out in order to gain sufficient momentum and

eventually prevail. Such "experiments in living" are at the very least an indispensable complement to more abstract moral reasoning and (semi) violent struggle; frequently, such reasoning turns out to be so powerless as to hardly engender any significant change at all.

Let's return to an example we've used before. The abolition of slavery was not brought about by ethical arguments and novel moral insights. Here, it is useful to remind oneself of the fact that in the middle of the 19th century, people were far more used to the idea of forced labor in general than we are today. Indeed, as Anderson shows, almost all labor was, in one way or another, forced:

> About 95% of the world's population labored under one kind or another of involuntary servitude: if not outright slavery, then serfdom, debt peonage, apprenticeship, indenture, corvée, military impressment, penal servitude or other forms of coercion, such as coolie labor. Even the workers who were called "free" would not be considered free by today's standards.
>
> (Anderson, 2014, p. 15f.)

This was typically justified on grounds of efficiency, because it was presumed that if people weren't forced to work, they wouldn't; and if they didn't, social productivity would deteriorate to unacceptable levels.

Advocates of free labor thus faced an uphill battle, for it was unclear whether a society could survive at all if they relied on carrots rather than sticks to convince people of joining the work force. So before the abolition of slavery could start to seem realistic (to skeptics of abolitionism), it had to be tried out in a real-world experiment. Once such experiments in living—such as the production of sugar in the West Indies, where the productivity of Cuban slaves was compared to that of free Jamaicans—could be shown to succeed, abolitionists had produced new and powerful evidence not just for the desirability, but the very feasibility of their emancipatory vision.

Another reason why experiments in living are crucial for fostering moral progress is due to the tight grip of *status quo bias* (Bostrom & Ord, 2006). Our frequently irrational partiality towards whatever happens to be the case right now demands living proof that a suggested alteration of our way of life won't result in disaster. Many of these panicked prognoses would be funny if they weren't so consequential for a large amount of people:

> [I]t is worth noting that there are many cases in which conservatives have predicted dire consequences of de-moralization that have not occurred—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 thirty years ago.

(Buchanan & Powell, 2017, p. 131f.)

None of these apocalyptic consequences ever materialized, of course, but it takes actual "moral experimentation" (132) for people to realize this, and to escape the thrall of their poor imagination.

Publicly visible "norm entrepreneurs" (Sunstein, 2019) can assist in unleashing people's true preferences or indeed in creating entirely new ones: "Norm entrepreneurs draw attention to what they see as the stupidity, unnaturalness, intrusiveness, or ugliness of current norms". These pioneers can initiate self-perpetuating norm cascades in which they first convince others to abandon an unjustified rule; others, who don't join a movement until many others have already joined, will then be convinced as well. After a while, the entire norm is toppled.

I have said earlier that the idea of moral teleology—that social change is biased towards moral improvement—isn't committed to the idea that such improvements will take care of themselves. Moral progress doesn't bypass human action; rather, it unfolds through, with, and because of the involvement of individual and collective human agency.

This is why social movements are of such central importance to the trajectory of progress. Odious moral norms that deserve to be reevaluated often remain in place for far too long because people are caught in epistemic traps (they are unaware of the fact that many others are also inclined to reject those norms and would rather see them go) or because of first mover problems that discourage individual resistance which is likely to be ineffective yet personally costly. People may believe that their neighbors think it ethical to mutilate their daughters' genitalia, and that they expect them to do the same. In reality, the reciprocal expectation may be autopoietic, and may fall apart once it is revealed to the involved parents that everyone would rather spare their daughters this cruel and unnecessary treatment. Social movements essentially function to make such traps publicly visible, and to demonstrate that the people who suffer from them do not deserve to be mistreated. Here, Anderson (2014) draws on Charles Tilly's account of contentious politics to illustrate the four features that social movements typically need to have to successfully demand moral reform. Two of these features—worthiness and commitment—display the moral quality of the cause and those who campaign on its behalf; the other two—numbers and unity—are about the social importance of the cause, as evidenced by the number of people that are invested in it.

But moral progress doesn't always have to be engendered from the ground up, as it were, by channeling the forceful demands of the downtrodden and marginalized. It can come from everywhere in society, and not least from dissatisfied elites who would like to see their freedom unshackled from pointless and suffocating constraints.

Once certain liberal or proto-liberal regimes are established that allow for some plurality of lifestyles, tastes, preferences, orientations, values, and biographies, societies have moored themselves to a chain of benign feedback loops where increased elbow room for testing the livability of competing conceptions of the good can help societies discover normative innovations that replace the staleness of the status quo with the freshness of novel ways of human flourishing (Muldoon, 2015).

The result is a self-propelling dynamic which drives

> social change as a process of human development, which is producing increasingly humanistic societies that place growing emphasis on human freedom and self-expression. A massive body of cross-national data demonstrates that (1) socioeconomic modernization, (2) a cultural shift toward rising emphasis on self-expression values, and (3) democratization are all components of a single underlying process: human development. The underlying theme of this process is the broadening of human choice. Socioeconomic modernization reduces the external constraints on human choice by increasing people's material, cognitive, and social resources. This brings growing mass emphasis on self-expression values, which in turn lead to growing public demands for civil and political liberties, gender equality, and responsive government, helping to establish and sustain the institutions best suited to maximize human choice.
>
> (Inglehart & Welzel, 2005, p. 2)

Welzel buys into the idea that prosperity propels liberty:

> With widespread action resources, people emphasize emancipative values. Shared emancipative values create solidarities that encourage joint actions to assert and exercise freedoms. The solidarity experience generates common satisfaction. This reinforces the strong valuation of freedoms, creating another self-sustaining cycle. In this constellation, rulers are severely restricted in their institutional choices. They are under the pressure of public claims that are difficult to resist because these claims are put forward by capable and motivated people who act with the power of solidarity. Eventually, rulers must guarantee universal freedoms and are pressured to adhere to these guarantees.
>
> (Welzel, 2013, p. 53)

This dynamic holds with an important caveat: the society in question must have stumbled upon the "cycle of empowerment", where resources, values, and regimes interlock in serendipitous ways to generate positive feedback loops in the direction of ever further increases in freedom and autonomy.

This process is far from complete, and in many places, it is merely incipient or continues to be successfully suffocated by the powers that be. But it is probably hard to find a dimension of moral progress that appeals to more traditions in political philosophy, from libertarianism over republicanism, critical theory, poststructuralism, or egalitarianism, than a resistance to social unfreedom. On the trajectory towards emancipation, Welzel distinguishes three stages that societies typically have to pass through to achieve the highest levels of autonomy and access to opportunities (Welzel, 2013, p. 23). During the suffering stage, people's individual and political agency is stifled; poverty constrains their options, scarcity determines their values. In the struggling phase, some liberties begin to be unleashed; values follow suit, encouraging further empowerment; political regimes begin to respond, and start tolerating some level of citizen participation and individual autonomy. Finally, during the thriving stage, individuals enjoy expansive freedoms to shape their lives; political participation is encouraged and accessible to all; culturally, these societies put great emphasis on authenticity, self-expression, and individuality, and are comparatively wary of conformity and excessive reverence for authority and tradition.

Notes

1. For an exception, see Hermann (2019).
2. [An] attractor, as I have characterized it, is an abstract, statistical construct, like a mutation rate or a transformation probability. To say that there is an attractor is just to say that, in a given space of possibilities, transformation probabilities form a certain pattern: they tend to be biased so as to favor transformations in the direction of some specific point, and therefore cluster at and around that point (Sperber, 1996, pp. 111–112).

7 Unsocial Sociability

How Can Moral Progress Be Sustained?

Introduction

What does it take to promote and maintain moral progress? Expanding circle accounts of moral improvement hold that to bring about moral progress and avoid moral regress, we need to prevent the moral circle from tightening, and work towards expanding it further. In particular, this means that we need to make sure that the right social conditions are in place for people's inclusive attitudes to thrive, and for their xenophobic and discriminatory attitudes to be curbed. Other types of moral progress such as increases in people's welfare and freedom or the elimination of harmful and unjustified norms similarly require corresponding social conditions.

One of the main claims of this book has been that, when it comes to generating and promoting moral progress, people's psychological attitudes are not where the action is at all: moral progress happens in external institutions supplied by cultural structures and practices. The kernel of truth in evolutionary conservatism is that empathy and altruism *are* limited. There is simply no way to get people to care for everyone in strict proportion to their equal moral worth. This does not mean, however, that the prospects for moral progress are as grim as such conservatives suggest. The solution developed here is to stop further stretching the constraints of our evolved psychology, and to start bypassing them instead.

A big part of this chapter will be to identify institutional mechanisms that fit this profile, and to defend the importance of institutional solutions to moral and political problems in general. Here, another psychological obstacle looms large. Moral emotions have been set up for up close and personal face-to-face interactions among kith and kin (Greene, 2013). In larger societies, they can become morally powerless or even pernicious, leading to intergroup conflict and cycles of anti-social behavior. Stabilizing moral progress thus consists in finding a way for expanding the moral attitudes which regulate small-scale tribal societies to modern large-scale societies by building institutions whose normative infrastructure is essentially

incompatible with tribal morality, but necessary for realizing the enormous gains in welfare and quality of life which are nowadays considered without alternative. The institutional framework I will flesh out shows how to navigate this dilemma by harnessing the power of evolutionarily inherited moral intuitions for the sake of often counterintuitive institutional arrangements. Richerson and Boyd put it aptly when they frame the challenge in terms of "social innovations that make larger-scale society possible, but at the same time effectively simulate life in a tribal-scale society" (2005, pp. 231–235). Institutions, in a sense, are a con we pull on ourselves.

One deeply pessimistic message this entails is that we will never feel truly at home on the modern world. The world is improving, but a sense of alienation that perpetually energizes a yearning for a radical, nondescript alternative will remain with us.

This chapter has eleven sections. In the first two, I highlight two important aspects of institutions: how they constrain and enable. Section 7.1 will thus address the *normative* character of institutions; section 7.2 will address their power to *scaffold* human behavior. Section 7.3 makes the case for the importance of institutions by describing the distinction between *inclusive* and *extractive* institutions and the respective roles they play in promoting and hindering human welfare and freedom. The remaining sections present a typology of institutional mechanisms suitable to secure progressive moral gains: *bypassing* institutions (7.4) achieve expansions of the moral circle by economizing on moral motivation; *proxy* institutions allow people to discharge their less desirable tendencies without doing much harm (7.5); in section 7.6, I discuss *ameliorative* institutions and their potential for improving human traits and dispositions; *slow* institutions provide a counter to the bugs of intuitive cognition section 7.7; *reflexive* institutions draw on our knowledge of the flaws and biases of human thinking more generally and apply it to structure the social environment accordingly; in section 7.9, I will briefly describe how we can derive contentful moral principles from the distinct normative infrastructure of various institutions; finally, I briefly situate the institutional approach to sustaining moral progress within a general framework of cultural evolution and cumulative moral learning.[1]

7.1   Intelligent Design

We should not bludgeon our evolved dispositions into shape, but harness their power. The best way to achieve this is via clever institutional design (Heath, 2014b; Gallagher, 2013). Such smart institutional scaffolding doesn't waste time and resources on our nice attitudes, trying to make them more encompassing. Effective institutional kludges economize on

moral motivation, and achieve more desirable outcomes with the messy psychological cards we have been dealt. Out of the crooked timber of humanity, no straight thing was ever made—but what if we don't need to make a crooked thing straight to achieve the progressive results we want?

I have argued that the challenge from evolutionary conservatism can be countered, and that moral progress is more feasible than the limits of our altruistic and cooperative dispositions would make it appear. However, a further assumption that is frequently made is that even if moral progress is psychologically feasible, it is a matter of brute good luck whether it is *stable* (Wright, 2006; Tainter, 1988). This worry is overblown. When it comes to lasting moral improvements, individual psychologies are indeed too frail and susceptible to situational variation (Doris, 2002) to be left to their own devices. Moral attitudes require institutional support (Sterelny, 2007b; Levy, 2012), and institutional provisions can support the stability that individual minds cannot.

Institutions can be put to use both for promoting moral progress as well as for blocking moral regress. Moral regress happens when illegitimate forms of exclusion are reinstalled, when neutral traits or actions become remoralized, or when vulnerable people are singled out for ostracism and dehumanization. One of the main tasks of an account of progressive social change should be to look at institutions from this perspective. Moral progress happens when institutions facilitate (i) improvements in people's well-being through (ii) cooperative efforts in ways that are (iii) accessible to increasing numbers of people.

Let me be clear about the fact that other accounts of moral progress do not, of course, deny the importance of institutions for maintaining and promoting progressive gains (Buchanan & Powell, 2018). However, I do want to suggest that their importance for human cognition and behavior, and their conduciveness to moral progress, remain underestimated. This is one of the main gaps my account is supposed to fill. Along the way, I also want to clear the ground for the question whether there are certain elective affinities between institutional set-ups and the different *types* as well as *drivers* of progressive change I distinguished earlier.

Is the distinction between individualistic and institutional approaches to thinking about social change helpful to begin with? Virtually everyone agrees that both structural interventions as well as individual-level changes can be and often are at least of some importance for fostering lasting moral improvements. When it comes to reducing the pernicious force of racism to combat racial injustice, it is crucial that the racist attitudes held by individuals (such as implicit biases or explicit prejudicial beliefs) disappear *and* that the systemic factors that keep entrenching disadvantage (such as housing segregation or mass incarceration) are reformed. But structural prioritizers (see Madva (2016) for this helpful term)—such as myself—merely

insist that we should place *greater emphasis* on social practices and structures rather than individuals' psychologies in understanding and promoting moral progress.

I cannot do adequate justice to this complex and fraught debate here. It combines metaphysical (what are the fundamental ontological building blocks that need to change for society to change?), pragmatic (where should we target our interventions in trying to promote progressive change?), normative (which type of change, attitudinal or structural, is morally most desirable or appealing?), and explanatory (are individual or structural factors causally primary in engendering moral progress?) aspects.

Structural prioritizers often operate with counterfactual tests: if, say, our sexist attitudes were to magically disappear overnight, would gendered patterns of disadvantage vanish with them? If they would not, then systemic factors, supraindividual social practices and collective structures are what counts. But such tests are often neither here nor there, because individualists can simply turn the tables on them: if you were to achieve perfectly equal access to education and completely desegregated neighborhoods overnight, but individual hostilities were to persist, would racist or sexist structures not reappear after a while? It's a draw.

However, when it comes to understanding moral progress, the institutional level remains the more important one. For one thing, our best examples of moral progress almost always involve changes at the structural level. The abolition of slavery, female political participation, civil rights for minorities consist of, were made possible by, and are maintained through, institutional measures first and foremost. It is true that individual attitudes, such as the convictions of a small group of influential reformers, play an important role. But in virtue of being a social phenomenon, moral progress can only be made durable and widespread if systemic factors follow suit. Second, moral progress is about how to pile the next moral improvement onto the previous one so that *intra*generational improvements can be ratcheted up into *inter*generational progress. Individual minds are obviously part of this process, but they can't be the glue that holds it all together.

Institutions are the engines of moral learning and the vehicles of progressive reform. As people figure out new ways of living together, they figure out new ways of storing the respective information, norms, and skills in their external environment. This allows moral progress to accumulate over time. A focus on the cultural and institutional side of moral progress is thus a key element of the teleological account defended here in particular. As far as individual minds are concerned, each generation goes back too far to the moral drawing board.

That moral behavior would be secured institutionally in ways that need not be accessible to people's conscious awareness makes sense in light of

the fact that much of our conduct is shaped by hidden motives we are not aware of. This has implications for good institutional (re)design:

> Savvy institution designers must . . . identify *both* the surface goals to which people give lip service *and* the hidden goals that people are also trying to achieve. Designers can then search for arrangements that actually achieve the deeper goals while also serving the surface goals.
> (Simler & Hanson, 2017, p. 311)

The suggestion to bypass people's psychological dispositions in securing morally desirable behavior may seem unappealing from a distinctly moral point of view. Lily Frank (2020) imagines a "moral Shangri-La" in which an appropriate moral sensitivity in quality and quantity is secured via ecological engineering: ambient lights curb people's aggressive impulses, biomedical enhancement attunes their empathy towards disaster victims, and so on. In this world, Frank argues that moral struggle is essentially eliminated. This, she holds, removes an important source of moral progress. The scenario she imagines, however, would also activate many moral impulses that would motivate people do alleviate injustice all over the world. Struggle would still exist, but we would never struggle, so to speak, to engage in the right struggles.

When I say that moral progress is sustained through smart institutional design, I do not want to insinuate that successful institutions are literally designed (cf. van den Hoven et al., 2017). My point is, rather, that a workable institutional toolkit is very much beyond the scope of what even the smartest human designers can achieve. Institutional solutions to cooperative problems are almost never consciously intended and crafted; instead, such institutions (more or less) blindly evolve and are only ever incrementally improved upon by subsequent links in the chain of cumulative cultural transmission. To a large extent, moral progress is best promoted via non-intervention in the cumulative evolution of moral culture.

Consider the following thought experiment: would an individual from a different time, say the middle ages, perform better or worse in the Milgram obedience study? I suspect that their performance would, on average, be comparable to that of us (allegedly) authority-skeptical moderns. Likewise, imagine how an über-woke modern liberal, when transported with a time machine to a morally backward place, would likely revert to bigoted attitudes and abhorrent practices in a disconcertingly short amount of time. This illustrates that moral progress doesn't happen at the individual level, and that social change towards the better is not simply the aggregate of individual people becoming better. There is a genuinely social dimension to moral progress that cannot be reduced to the sum of individual moral improvement.

7.2   Storage and Retrieval: Mechanisms of Transmission

An aspect of moral evolution whose importance is often overlooked is that if moral progress is supposed to be possible, and especially if it is supposed to be biased in favor of moral improvement, societies do not just need to find ways of engendering moral change for the better. They also need to find ways of *entrenching* whatever changes they have undergone. Progressive gains, once achieved, need to be stored—somehow, somewhere—so that they can accumulate over time. Moreover, these means of storage must be such that the moral knowledge so compiled can then later be *retrieved*—absorbed, downloaded, acquired—by the following generation of moral torch-bearers.

In the next chapter, I will sketch an account of what I take to be the most important means of storing accumulated moral gains: institutions. Our step-motherly nature, however inadequate its provision of innate talents may otherwise have been, has bestowed upon as an ability to thrive in all sorts of environments from the buzzing streets of Mumbai to the coy wasteland of the Arctic. In large part, this is due to our capacity to offload moral knowledge onto external social practices. These end up functioning as a socially extended mind that facilitates cooperative success, at least potentially, even under the harshest and most ephemeral circumstances. Our extended moral mind allows us to store workable moral knowledge in our tangible environment.

What about retrieval and transmission? Infants are invariably *moral apprentices*. This entails, first, that—like Pac Man (Sripada & Stich, 2006)—they tend to pick up and internalize whatever behavioral norms, moral or not, are floating around in their environment. One way in which this happens is via explicit linguistic instruction:

> Children are born into a world rich with normatively appraised acts and agents. Other agents are persistent, inveterate moralizers. . . . [M]uch causal conversation is about others and their acts . . ., and much of this is normatively loaded. As soon as language comes online children are exposed to normative evaluations in stories, from their peers, from their parents' generation.
>
> (Sterelny, 2012, p. 163)

It would be misleading, however, to describe children merely as passive consumers of accumulated moral knowledge. In addition to that, they are also actively experimenting on what they can get away with, and indeed much of infant life consists in dipping one's toes in the normative water to see which behaviors are encouraged, tolerated, or sanctioned, which

explanations, justifications, excuses, criticisms, suggestions, or recommendations count, and for how much:

> [C]hildren explore, experiment on, and try to manipulate their social life. They try to influence others by moralizing themselves, and they are forced to respond to the attempts of others. Children develop in an environment saturated with local . . . normative evaluations. Thus a multitude of particular experiences, annotated with their moral status, acts as input to a pattern-recognition learning system. . . . Thus the general features of human social learning reappear in the development of moral cognition in children. Moral learning, too, is learning by doing, but in a structured and enriched environment.
>
> (Sterelny, 2012, p. 163f.)

The same mechanisms that allow knowledge and skills in general to be stored and retrieved—and thereby transmitted to the next generation—account for the accumulation of moral capital in a culturally enriched learning environment.

As with culturally accumulated knowledge and skills more generally, this transmission does not remain shackled to the glacial pace of generational change. For cultural learners like us, moral knowledge can spread much more rapidly. In primordial times, moral gains had to be secured via the selective pressures of relative reproductive success: new forms of cooperative social organization are not learned in the way individuals or groups learn during their own life time, but in a chewy process of evolutionary trial and error:

> When environmental conditions are positively but imperfectly correlated across generations, each generation acquires valuable information through learning that it cannot transmit genetically to the succeeding generation, because such information is not encoded in the germ line. In such environments, an animal could benefit from the transmission of information concerning the current state of the environment through some non-genetic information channel. Such information, called epigenetic by biologists, is quite common [. . .] and achieves its highest and most flexible form in cultural transmission in humans and to a considerably lesser extent in other primates [. . .]. Cultural transmission, also called social learning as opposed to individual learning, takes the form of vertical (parents to children), horizontal (peer to peer), and oblique (non-parental elder to younger) transfer of information.
>
> (Bowles & Gintis, 2011, p. 14f.)

This could also explain why recent moral progress, though still maddeningly slow, is accelerating: like most other technological developments, the accumulation of moral know-how seems to be subject to exponential returns.

For there to be any moral knowledge to be stored, retrieved, and improved upon, however, there must be mechanisms in place that would lead to moral changes for the better in the first place.

7.3   Norms and Practices

Theories of institutions come in two basic flavors that could be referred to as *normativist* and *externalist*. Normativist accounts of institutions stress the extent to which institutions are best described in terms of the norms and practices people use to navigate the social world; externalist accounts emphasize how institutions support certain behaviors by externally "embodying" certain social customs or practices. These two options are not mutually exclusive, and should be taken to highlight different important aspects of institutions and how they function: "Institutions are the humanly devised constraints that structure political, economic and social interaction" (North, 1991, p. 97).

Moreover, the question of what institutions are has two readings: a conceptual and a metaphysical one. Since I want to avoid both the trenches of conceptual analysis as much as the deep well of social ontology, I will discuss these questions only briefly. The main point of this section is to hone in on an account of institutions that will allow us to see what makes them so useful. Institutions, I will argue, are stable, externally supported practices that allow people to adjust their behavior against. This feature of stability is important because without it, institutions couldn't do the scaffolding that we need them to do in order to promote progressive ethical gains. I won't develop anything that comes close to a theory of institutions here, for reasons of both space and competence (to wit: lack of either). One way of approaching the nature of institutions, however—and not the worst way—is to proceed from examples.

Conventions are one. Famously, Lewis (1969) modeled conventions as solutions to arbitrary selection problems between symmetrical equilibria in coordination games. We both care about being, or at least appearing, polite, but neither of us care much about whether we shake hands or bow in order to do so. What we'd both like to avoid, however, is to mix strategies: we'd both like for both of us to either bow or shake hands. Either pair of strategies forms an equilibrium for us, and our task is merely to pick one. Conventions, then, are rules regarding how to behave that are common knowledge (I know that we are supposed to shake hands; I know that you know; I know that you know that I know; and vice versa). These interlocking shared expectations can solve the problem of how to settle for a successful greeting ritual. Once it is established, the situation is self-reinforcing. An institution is born.

Social facts are another example. Searle understands social facts in terms of constitutive rules (1995). Constitutive rules specify as what something "counts": they do not describe mere regularities in behavior, nor prescribe a certain conduct as normatively required, but determine which moves there are to make in the first place. Money, weddings, professional statuses, or passports are all instances of genuine social facts or practices whose very existence is determined by such constitutive rules.

In many cases, people's compliance with social rules will depend on their conditional expectations regarding how other people will likely behave and what other people are likely to deem normatively required (Bicchieri, 2005, 2016). Genuine social norms require not just that people all act in a certain way: many people simultaneously opening their umbrellas is usually due to rain, and not due to a flash mob. Likewise, mere customs lead to many people behaving in similar ways, but this is different from people behaving in similar ways because they (i) expect others to act this way, because they (ii) expect others to believe one ought to act this way, and because they (iii) have a conditional preference to act this way themselves, given (i) and (ii). This emphasis on conditional preferences and normative expectations is of particular importance when it comes to changing social norms, which is often best achieved by manipulating people's expectations regarding other people's actions and beliefs. My preference for binding my daughter's feet is dependent on what I expect others to do and expect of me; their absence makes it much easier for me to defect from this bizarre cruelty.

Institutional rules play a crucial role in resolving collective action problems. This can in itself be a form of moral progress, and it often happens independently of centralized intervention or the market-like aggregation of individual decisions via parcelized property structures. Famously, Elinor Ostrom (2015) has shown that there is a third way for resolving real-world prisoner's dilemmas/tragedies of the commons/collective action problems besides centralized coercion and full internalization of costs. In many cases, so-called "common pool resources" are best managed via local cooperative efforts. Turkish fishers in Alanya, for instance, draw on an intricate rotation system to maintain both egalitarian access to and sustainability of their fishing grounds.

In all of these cases—conventions, social facts, social norms, and collective action problems—institutional norms allow people to reap the benefits of cooperation, allowing society to deliver on its promise to provide, in Rawls' famous words, a cooperative venture for mutual advantage.

7.4 The Socially Extended Mind

Institutions do not just reside in the realm of abstract norms. They are concrete things, with concrete manifestations: institutional structures live in buildings, bridges, and books, and as such, they do not merely normatively constrain but also externally embody shared social practices.

The drawback of this is that institutions have a cruise ship's turn radius. There is an inherent tendency towards conservatism in institutional structures, which is why conservative writers from Burke over Hegel to Gehlen tend to love them, mostly for how they seem to allow them to defend a presumption in favor of the tried and true. I am aware of this danger—if indeed it is one; my task will be to flesh out the most progressive version of institutionalism I can come up with.

What, then, is the upside of letting social practices sediment into tangibility? If psychological and philosophical situationism has taught us anything, it's that internal psychological traits are typically too frail to support robust behavioral dispositions (Doris, 2002; Alfano, 2013). Some authors have therefore turned to external situation management as a solution for our imponderability; the case for institutions I have in mind advocates a generalized version of this solution, an added benefit of which is that inherited institutions whose normative infrastructure has been built up via cumulative improvements across generations do not have to deal with the bootstrapping problem of who's in charge of reliably implementing said situation management if human agents are always already too unreliable to function without them.

The main advance in conceptual engineering to help frame the role of social institutions for human life is due to Clark and Chalmers' "extended mind" hypothesis (1998). Clark and Chalmers argued that if something performs a cognitive function, we should not hesitate to classify it as a cognitive process regardless of whether or not it falls within the phrenologist's purview. External "devices" such as notebooks or spouses do not merely complement or supplant memory, mental arithmetic, or decision-making; a good deal of our cognitive operations is either so deeply coupled with artifacts and other features of the environment or simply occurs on such external platforms that a neat separation of the *forum internum* from the outside world becomes impossible.

I take no stand here on the—it seems to me largely terminological—issue of whether cognitive operations *genuinely* extend into the environment or whether purported examples for such functional extension are best reconstructed as fringe-y instances of otherwise ubiquitous cognitive-behavioral *scaffolding* (Sterelny, 2010). It would be implausible to deny the tremendous importance of the human *Mängelwesen's* ability to rely on external support to compensate for its hugely underpowered on-board equipment. This holds at the phylogenetic—think of gene/culture co-evolution (Henrich, 2015)—as well as at the ontogenetic—think of all the tools and devices you have used today since you woke up and what you would have achieved without them—level. Human history is the history of malleable learners confronting an increasingly rich cultural environment from which to literally grab or figuratively download all sorts of cognitively

indispensable tools, patterns, and practices. Together and over time, these tools have given us an upright posture, a useless jaw, an impractically large head, and (soon?) the quantum computer.

Increasingly, many authors note the possibility of enlisting aspects of the social environment to facilitate various cognitive-agential performances. Gallagher (2013) refers to them, somewhat awkwardly, as "mental institutions". According to his account, a mental institution

> 1. [i]ncludes cognitive practices that are produced in specific times and places. 2. Is activated in ways that extend our cognitive processes when we engage with them (that is, when we interact with, or are enactively coupled to them in the right way).
>
> (2013, p. 3)

The aforementioned conventional rules for greeting, established rules for resolving collective action problems, and constitutive rules for how to wed two people or make a promise can all serve as examples of such externalized social practices that offload cognitive operations into a shared social environment from which they can later be retrieved when necessary. Gallagher adds legal systems (think of constitutions) and property rights to the list, and we are free to add even more things such as museums, rituals, urban landscapes, and other distinctively social creations.

It is hard to overestimate the cognitive and practical importance of our socially calcified "extended will" (Heath & Anderson, 2010). Many, if not most, forms of individual and cooperative action would be impossible or at least sufficiently unlikely without the kind of scaffolding human agents receive from external social institutions. On the other hand, many things would become *too* likely to occur too unacceptably often: think of constitutions again, and how they protect democratic societies from the whims of unadulterated majoritarianism. In this way, institutions can shield social groups from the more harmful breakdowns of their collective self-control, much like an empty fridge or unshaved legs protect individuals from succumbing to more mundane temptations.

7.5   Institutions Rule

In engendering moral progress, targeting psychological dispositions is ineffective. Instead, we should target institutional mechanisms to structure our behavior in morally desirable ways.

One of the main errors of much of the Enlightenment tradition lies in its overestimation of the importance of internal psychological attitudes such as knowledge or benevolence. This psychologistic perspective, in turn, leads to a misguided focus on education as the main means of securing socially

progressive developments: if only people had all the information—which we can provide them—and if only they had the right moral principles—which we can instruct them about—surely moral progress will drop out, as it were for free. But individual psychologies cannot support robust moral gains; for that, a proper institutional architecture is required.

We can frame the importance of institutions in terms of the parametric/strategic distinction (Schmidtz, 2008, p. 148ff.). In criticizing what he refers to as the "Singer Principle"—that is, the claim that we are morally required to help others when nothing of comparable significance is at stake, which seems to entail that the affluent should give away most of their wealth to the global poor (or anyone in dire need)—Schmidtz emphasizes that when we give to charity, we do not commit a single isolated deed. What we really do is signal a strategy which other people use to anticipate our future behavior and adjust their own strategies against. Whether or not we agree with Schmidtz's ethical argument that the likely way in which people in need will respond to the expectation to be helped may make everyone worse off is beside the point here. The main issue is that externalized social practices—institutions—establish a background of for-granted behavioral rules which, by default and as a matter of course, is removed from deliberative second-guessing until the need arises:

> Institutions (hospitals, for example) serve the common good by leaving well enough alone—creating opportunities for mutual benefit, then trusting individuals to take advantage of them. That is how (even from a utilitarian perspective) institutions have a moral mandate to serve the common good that does not collapse into a mandate for ordinary moral agents to maximize utility.
>
> (Schmidtz, 2008, p. 152)

It would be a bad idea to permit doctors to cut up and harvest the organs of healthy patients not because it wouldn't maximize the good. It would, *for now*. But in a strategic world—our world—doing so would make it difficult to run hospitals in the first place.

The unique importance of institutions for moral progress is perhaps best illustrated through the examples of wealth and economic growth. This is connected to my discussion of different types of moral progress earlier. Wealth, broadly conceived, is tied to many morally significant aspects, such as health, opportunities, well-being, freedom, autonomy, or leisure, and thus shouldn't be dismissed as morally irrelevant. (Dismissing material wealth as morally insignificant is something only few can *afford*.) In particular, these two things are important because they allow us not to solve, but to simply eliminate more and more of the trade-offs that tend to keep moral philosophers busy (Cowen, 2018). Moreover, wealth and growth

can be used as *proxies* for moral progress, because societies do not grow wealthy unless they manage to build workable institutions that foster or rely on equality, political participation, cooperativeness, peace, stability, individual rights, creativity, and innovation, all of which are themselves deeply important measures for moral progress (Evans, 2017).

In explaining the origins of wealth, modern economics increasingly recognizes the crucial role of institutions (Rodrik et al., 2004). Acemoglu and Robinson (2013), for instance, press the case that only institutions can explain the sometimes staggering differences in wealth and development observed in otherwise identical places: people's prospects in North and South Korea or Nogales, Arizona and Nogales, and Mexico are very different.

Neither geography nor culture can account for these effects. Famously, Jared Diamond (1998) argued that most important differences in terms of wealth and development between societies can be explained by a number of trivial but hugely consequential accidents in geography or zoological distribution. Unlike the Americas or Africa, the Eurasian continent's main axis has an East-West direction, allowing a much faster and reliable spread of crops, livestock, and cultural tools. To add insult to injury, inhabitants of Eurasia can draw on a bountiful number of large animals such as pigs, sheep, or cattle. Africa, on the other hand, has Zebras, which turn out to be too dangerous and volatile to domesticate, or Elephants, which take more than a decade to grow and were only ever tamed, not domesticated.

Though such contingent differences in geography or megafauna can be important, they ultimately fail, according to Acemoglu and Robinson, to explain why, when such factors are held constant across conditions in a natural experiment, some places grew rich and others did not. Rather, these wildly different trajectories come down to what they refer to as *inclusive* and *extractive* social, political, and economic institutions (2013, p. 70ff.). All social institutions, they emphasize, are created and maintained *politically*:

> Inclusive . . . institutions . . . are those that allow and encourage participation by the great mass of people in economic activities that make the best use of their talents and skills and that enable individuals to make the choices they wish. To be inclusive, economic institutions must feature secure private property, and unbiased system of law, and a provision of public services that provides a level playing field in which people can exchange and contract; it also must permit the entry of new businesses and allow people to choose their careers.
>
> (74f.)

The key point here is that such institutions must be stable, equal, and open to all (or at last the vast majority of people) in a given region. Extractive institutions of the kind we see in North Korea or many African

countries today and everywhere for most of human history have none of the aforementioned features and are instead designed to benefit a tiny elite group of people at the expense of the rest. Such extractive institutions leave people destitute and insecure: private property is seized or taxed away, education consists of propaganda, significant swaths of people are systematically legally discriminated against, and the rest is coerced into lives and careers they have neither interest in nor talent for.

Institutions matter because in a strategic world populated by moderately intelligent people with at least some degree of foresight, they allow social groups to provide, harness, and channel existing incentives for the common good. As the example of economic wealth shows, these institutions can also be set up very badly, thereby enshrining stagnation and stalling progress. The question, then, is how to design institutions well, and what the profile of such well-designed institutions looks like.

Inclusive institutions generate increasing returns that lead to progress along other dimensions as well via a "virtuous circle" of positive feedback loops (from increasing wealth to increasing freedom to increasing division of labor and back again) (Boettke & Candela, 2017):

> In the course of our subsequent history we created novel social and physical environments exhibiting similar, or even greater, benefits of cooperation, among them the division of labor coordinated by market exchange and respect of rights of property, systems of production characterized by increasing returns to scale (irrigated agriculture, modern industry, information systems with network externalities), and warfare. The impressive scope of these modern forms of cooperation was facilitated by the emergence in the last seven millennia of governments capable of enforcing property rights and providing incentives for the self-interested to contribute to common projects.
>
> (Bowles & Gintis, 2011, p. 3f.)

Only such an institutional approach can explain moral progress in light of the fact that individual human nature did not recognizably improve:

7.6   Institutional Bypassing

Some enlightenment thinkers did acknowledge the importance of institutions for driving moral progress. Here, Kant stands out for recognizing—in one of his slightly more tongue-in-cheek moments—how superficially odious institutions can achieve progressive results through recruiting humans' "unsocial sociability":

> The means that nature employs in order to bring about the development of all of the predispositions of humans is their antagonism in society,

insofar as this antagonism ultimately becomes the cause of a law-governed organization of society. Here I take antagonism to mean the unsociable sociability of human beings, that is, their tendency to enter into society, a tendency connected, however, with a constant resistance that continually threatens to break up this society. This unsociable sociability is obviously part of human nature.

<div align="right">(Kant, 2006 [1784], p. 7f.)</div>

According to Kant, *adversarial* institutions such as status competition can lead to a perfection of talents that makes, via the positive externalities is produces, everyone better off despite encouraging a race to the bottom for the parties so competing. Smith's "invisible" hand and Hegel's "cunning of reason" stand in the same tradition of emphasizing the power of social forces to reach through individual agents' intentions to achieve supraindividual goals. These goals may not even occur to the people involved and may indeed contradict their consciously held intentions. If we want to look beyond psychological solutions to the feasibility of moral progress in general and the challenge from evolutionary conservatism in particular, institutions that bypass individual agents' mental states are the way to go.

In the following sections, I will work out the functional character of a variety of progressive institutions. Here, a small caveat is in order: one need not, as far as ones first-order normative judgments are concerned, agree that the institutions I describe are actually morally desirable. Rather, my aim is merely to illustrate how institutions can provide the cultural scaffolding required to make progressive moral gains sustainable. My argument concerns the structural profile of these institutions. It is possible to replace the examples I give with others one finds more morally or politically attractive.

Moreover, I want to place my discussion of those institutional profiles in the context of both the challenge from evolutionary conservatism and the overall teleological argument of this book. Remember that evolutionary conservatives emphasize the constraints imposed on the feasibility of progressive gains by pointing at features of our evolved psychology, such as hard-wired limitations of empathy, altruism, and our cooperative dispositions. Institutions show how these constraints, if they are real, can be successfully *bypassed*. Together with the driving mechanisms of moral progress described in the previous chapter, these institutions manage the accumulation of moral capital over time.

To illustrate, consider the following analogy: imagine an epistemological view that functions just like evolutionary conservatism does in the field of ethics. Such evolutionary conservatives in epistemology would be in some sense right to say that there are limits to knowledge and information processing, set by our evolved neural capacities. Our individual brains were not made to appreciate the vastness of the cosmos, or the mathematical

properties of prime numbers. Nothing in the environment of evolutionary adaptedness prepared us for gauging the speed of light or figuring out quantum weirdness. Therefore, this epistemological evolutionary conservative could suggest that many forms of knowledge are simply out of reach for simple primates like us.

Yet here we are, flying to the moon (or did we?), proving Fermat's Last Theorem, snapchatting, and vaccinating our children. How was this possible? How did we transcend the constraints of our evolved psychology, which seemed to be so formidable? The answer, of course, is that we didn't need enlarged individual epistemic capacities and bigger brains (by analogy: extended empathy or altruism) to achieve this. Evolutionary conservatism about knowledge on account of our unimpressive brains is false. To acquire more knowledge, we rely on cumulative cultural transmission and an epistemic division of labor as well as trustworthy institutions of knowledge-generation and information storage such as science. Our onboard epistemic capacities underwent little or no change over the past 10,000 years. Indeed, our superior individual intelligence does not seem to be the "secret of our success" at all (Henrich, 2015, pp. 8–22). If anything, this shows that our onboard epistemic capacities are not all that important. They are not what make us smart.

The same applies to our onboard *moral* capacities such as empathy. To bring about progressive moral change, we don't need to stretch our moral concern further. We need to work around it with clever institutional kludges. What are empathy and altruism for? For the most part, their moral value lies in the tendency to induce cooperative dispositions in agents. These cooperative dispositions plausibly face severe limitations, in no small part set by evolution. But with the right institutions, these constraints simply do not matter. We can design institutions that economize on our limited inclusivist attitudes. Adversarial institutions that indirectly induce cooperative behavior via staged competition are a prime example.

Consider the market as a paradigm case of such an institution. (Let me emphasize again that you do not have to be a fan of the market to see how it illustrates the power of institutional scaffolding.) People are only mildly and parochially cooperative. Evolutionary conservatives argue that this undermines the prospects of unlimited expansions of the cooperative moral circle. However, our psychological capacities—which we largely share with other primates—simply cannot account for current forms of human cooperativeness anyway. Modern humans are hypersocial in a way that cannot be attributed to changes in our psychological capacities because these developments are too recent. Cooperation in large-scale societies, if it had to reach through individual minds to work, would impose unbearably heavy motivational and epistemic burdens on individual people. It is almost impossible to motivate people to cooperate with strangers

on the other side of the planet out of the goodness of their hearts. There *are* evolutionary limits to human inclusiveness.

Does this mean that evolutionary conservatives are right, and large-scale cooperation is not a feasible political ideal? Of course not. There is an institutional arrangement—the market—that provides a workaround. It facilitates extensive chains of cooperation, and incentives to benefit others, without relying on the baker's, butcher's or brewer's benevolence. Joseph Heath (2014b) makes this point eloquently:

> When it comes to the limits of benevolence, Smith had what might be regarded as a common-sense view of the matter. The average person exhibits a certain degree of altruism, particularly toward family and friends, and can occasionally be motivated to help a stranger. But this disposition is rather limited and subject to exhaustion. When the limit is reached, the individual can be expected to act in a self-interested fashion. Cooperation, however, to the extent that it is vulnerable to free-rider problems, requires that individuals refrain from pursuing their self-interest. It therefore imposes a *motivational burden* upon individuals, which in turn generates a prima facie limit on the extent of cooperation, and therefore an upward bound on both the scale of a society and the level of social complexity.
>
> (211)

These are the limits of inclusivism all over again, but this time in an institutional context. Therefore, Heath continues,

> [A]n institutional arrangement, such as the market, which allows individuals to cooperate without tapping into these motives, is an important discovery. It expands the scope of cooperation by allowing society to economize on moral motivation, that is, to get more out of the level of moral motivation that it can plausibly expect (or non-tyrannically demand) from its members.
>
> (211)

It is frequently noted that literally millions of people cooperate to produce even the simplest items, such as pens (Satz, 2010). Markets produce this result of peaceful, mutually beneficial cooperation by working with, rather than against, the crooked timber of humanity. Market competition on the basis of decentrally generated price signals achieves this result by channeling people's existing motivations into socially desirable behavior. It is a form of staged competition through which some market participants (the competitors) are deliberately kept locked into a collective action problem to produce positive external effects for other market participants (the rest

of society). People cooperate independently of or even against their will, by acting within an institution whose overall rationale reaches through people's intentions to realize the common good.

The market is not the only example, of course. Religious institutions transform prudential concern for one's own welfare into moral concern for others' welfare via the threat of divine punishment. (Importantly—and this is something that was already appreciated by Max Weber—once a behavioral pattern is established, its religious underpinnings can slowly fade away without any loss in functionality.) In other cases, limited psychological resources can be expanded with technological solutions. A seemingly trivial but instructive example would be the website *airbnb*: trust is a limited resource. How many people would I trust enough to give them the keys to my house and allow them to stay there while I am away, use all my furniture, sleep in my bed, cook in my kitchen, and shower in my bathroom—one hundred? But modern technology has found a way to functionally "expand" my trust without changing anything at all about how my psychology works, and it is now possible for me to allow complete strangers into my home that I have never met and will never meet, all because clever technological design accomplishes a task that it would hardly be possible for my unaided mind to perform on its own.

7.7   Proxy Institutions

Bypassing institutions achieve desirable moral gains (such as extensions in the circle of people one is inclined to cooperate with) while economizing on moral motivation. They allow people to act *as if* they had the required moral attitudes, without actually requiring them. This arrangement is more sustainable than wanting to accomplish the same thing *intentione recta*.

At the same time, it is also highly counterintuitive, because the cooperative rationale of bypassing institutions is invisible from within, and remains hidden from those who participate—that's part of the point. Institutions are capable of sustaining moral progress because they don't have to bank on volatile psychological assets.

The point generalizes. The market is merely one example for how clever institutional design can render evolved psychological constraints irrelevant. Democratic institutions, for instance, also redirect otherwise morally dubious motives towards socially beneficial goals. In doing so, they serve as a *proxy*, and allow modern societies to channel the dialectic of authority and subversion into tolerable outcomes.

As with bypassing institutions, proxy institutions disarm an evolutionary fuse. The hunger for power, prestige, and dominance evolved like the rest of our social attitudes (van Vugt et al., 2008; Haidt, 2012). Under normal conditions, these dispositions typically lead to socially harmful

attempts to benefit oneself, one's family, one's loyal allies, and to over-power and subject others.

But when the control of the monopoly in (legitimate) violence is tied to the ability to win elections, one's more Machiavellian inclinations have to become at least somewhat aligned with the interests of the majority. This arrangement is arguably far from perfect (Brennan, 2016; Somin, 2016; Caplan, 2011), but it illustrates that frequently, the best way of dealing with undesirable psychological dispositions (such as selfishness or the will to power) is not to bludgeon them into shape, but to harness their force.

The strongly egalitarian instincts we inherited from our tribal ancestors (Boehm, 1999) did not deal well with increases in group size:

> These [egalitarian, H. S.] mechanisms were insufficient, however, to pre-vent exploitation in the comparably recent and far more complex post-Agricultural societies—at least not until the advent of the rule of law and constitutional democracy, which are plausibly viewed as cultural innovations "designed" to check special interest adaptations and thus place limits on state and elite class power.
>
> (Buchanan & Powell, 2017, p. 122)

Despite its flaws, democratic institutions are arguably without rival when it comes to organizing the peaceful *transition* of power.

Within proxy institutions, people can discharge their undesirable tendencies without doing much harm. An example for this are norms of honor. Recently, it has been suggested that feelings of honor play an important role in moral revolutions (Appiah, 2011). While honor can also have objectionable features, it has proven remarkably useful in bringing about positive social change regarding slavery or women's rights (Kumar & Campbell, 2016). Appiah's examples for moral revolutions—the abandon-ment of the duel, foot-binding, slavery, and the establishment of political rights for women—are ultimately about class and social distinction. Origi-nally, dueling or foot-binding spread among the wealthy, as a form of con-spicuous consumption of harmful idiocy. Moral progress is then achieved because people get rid of such silly and pernicious practices because they are adopted by a wider class of people, thereby eliminating the practices' signaling benefits.

The basic idea here is that proxy institutions provide a "channel" for people's morally objectionable tendencies to be neutralized and shielded off from the rest of society. Groupishness can find a safe home in sports. Status competition can be redirected into a variety of fields. There is, unfortunately, no guarantee that proxy institutions do their job well, even when the idea behind them seems clever enough. Queen Victoria married off her voluminous progeny into Europe's other royal families. The idea, it

appears, was to harness the force of intrafamilial sympathy to create peace and prosperity for a whole continent. And it kind of worked—until it kind of didn't.

Correspondingly, there is neither a guarantee that bypassing institutions, or social practices with any other structural profile, will do their job well. Markets, for instance, are not guaranteed to counteract discrimination and oppression (Sunstein, 1991; neither do they, in and of themselves, *create* discrimination, either). Throughout this chapter, my claim is that institutions, warts and all, do a comparatively better job at securing and promoting progressive moral gains than non-institutional solutions. Perfect solutions do not exist anywhere.

At a more fundamental level, institutions of marriage *are* an excellent example for clever cultural design (Henrich et al., 2012). Monogamous marriage stands out. Here, it is worth noting that monogamy is the anthropological exception; polygyny is the norm. Polygynous marriage is highly inegalitarian both for women and for those men who face the threat of being completely shut out of sex and reproduction. For those men, violent anti-social behaviors such as rape or murder become viable strategies to increase their chances of becoming one of the women's monopolizers. This drains a society's resources because so much effort is put into protecting or unlocking, respectively, a needlessly unequal distribution. Normative monogamy, in turn, reduces the fierceness of sexual competition, allowing resources to be relocated from risky mate seeking to parental investment in the health and education of one's offspring. In doing so, institutions of monogamous marriage tap the force of jealousy and sexual desire to promote egalitarian and cooperative goals. Inclusive gains and improvements in well-being are achieved institutionally, without tampering in any way whatsoever with the psychological dispositions of those involved.

7.8 Ameliorative Institutions

I do not, again, mean to convey any substantive normative judgment about the aforementioned institutions. The examples of the market, democracy, religion, honor, or marriage merely serve to illustrate the *structure* of intelligent institutional architecture. They are supposed to show how moral gains can be stored in an external social environment, not necessarily how this environment is supposed to look like.

Once indirectly cooperative arrangements such as bypassing or proxy institutions are established, they benefit from feedback loops that tend to engender more directly cooperative dispositions further downstream. Consider the fact that evidence from behavioral economics suggests that people in market societies tend to make the "fairest" offers in economic games such as *Ultimatum* or *Dictator* (Henrich et al., 2010, p. 6; Brennan & Jaworski,

2015, p. 96ff.). Contrary to what unfettered individual self-interest would recommend, people from, for instance, the USA, tend to make relatively high and reject relatively low offers in such games. This effect is likely due to an increased familiarity with norms governing exchange between strangers. Familiarity with such norms, in turn, can foster virtues of trustworthiness and dependability (Anomaly, 2017).

Some institutional arrangements are thus not really about bypassing or channeling our psychological resources at all. Many institutions do not work around as much as *transform* our moral attitudes. This fact seems to cut in precisely the opposite direction of my argument against the progressive usefulness of psychological dispositions. Likewise, one could note that science *has* enlarged our minds by providing an extensive repertoire of information and technology to download from our culture and peers. If our attitudes can be transformed like this, then why isn't that how we should counter the conservative challenge?

It is of course true that social institutions can significantly alter and shape our psychological dispositions. However, my point about the merits of channeling or bypassing those dispositions altogether stands, because the chains of cooperation and mutual benefit between millions of people achieved by market institutions have not been brought about by such transformations, and do not depend on them. Rather, the most successful (in terms of both outcome and further scalability) modern institutions achieve certain moral gains—securing cooperation and reciprocal benefit between distant strangers—without the involvement of *any motivation whatsoever* to cooperate with or benefit others. Once such institutions are established, they can then create useful feedback loops which engender trust and other forms of social capital. Transformed dispositions play a welcome role, but the heavy lifting in this "expansion" of the circle of cooperation is done by an institutional arrangement that does indeed bypass, rather than modify, our dispositions. And the level of cooperativeness so achieved is clearly beyond the limits of what even the most optimistic assessment of our onboard psychological resources could consider feasible.

Institutions that more straightforwardly improve people's values and motivations offer welcome support, however. Such *ameliorative institutions* do not just bypass undesirable human dispositions or morally neutral dispositions, but actively promote people's positive traits. An example just mentioned would be a properly regulated market that fosters attitudes of trust and commitment.

In-group bias is one of the main sights where we would like to see such ameliorative effects. Getting rid of our hardwired groupishness seems neither (currently) possible nor indeed ultimately desirable, since most of our biases have benefits as well as costs, and eliminating them completely would throw out the baby with the bathwater by robbing us of the functional

flipside of our dysfunctional tendencies. This is one more reason favoring institutional over psychological (incl. pharmaceutical or genetic) solutions (Levy, 2012) makes sense. There is some evidence, however, that ameliorative debiasing works through a variety of approaches such as encouraging people to consider alternatives, delay judgment and decision-making (Lilienfeld et al., 2009) or to take the perspective of the out-group they are biased against (Galinsky & Moskowitz, 2000; see also Meidenbauer et al., 2018). That "more research is needed" seems trite and hackneyed; when it comes to debiasing techniques, however, it could hardly be more true. Unfortunately, most psychologists seem more interested in finding flaws than they are in finding remedies for those flaws.

Some institutions or social practices have an ameliorative effect via exposure: people become more accepting of gay rights or alternative lifestyles when they are exposed to them. Kumar and Campbell (2022) argue that social conditions can be differentially conducive to such exposure. Here, a distinction between the horizontal and vertical distribution of a trait matters: sexual orientation is basically randomly and horizontally distributed. If 5% of people (the precise numbers don't matter here) are gay, everyone is all but guaranteed to find out they have a gay uncle or friend that they really like. Once coming out starts to be less and less sanctioned, almost everyone will become aware of several gay people in their lives that they already like. This greatly accelerates the erosion of homophobia. Unfortunately, race is a trickier matter, because it is a "vertical" trait: it's not hidden, it's not random, and society is often already segregated into races. This makes it more difficult to create the cross-racial exposure that is so helpful in reducing animosity. One rarely finds out that once beloved uncle was secretly black the whole time.

7.9   Slow Institutions?

If all this sounds too good to be true, that's because often it is. The institutional approach to securing moral progress developed here faces a considerable bootstrapping problem: most of the solutions I describe are difficult to realize under currently existing conditions; in many cases, current societies are positively rigged against their successful implementation.

There is a basic asymmetry between progressive developments and their anti-progressive alternatives, which is that most of the latter are viscerally compelling and seemingly common-sensical even if they are actually confused and totally wrongheaded. Most progressive solutions to policy problems, on the other hand, are simply not accessible to intuitive cognition. The case for the provision of public goods or the evidence for implicit bias takes time to explain. When institutions such as the media don't put in the necessary effort it takes to explain why a certain option may be preferable

overall even if somewhat harder to understand, important policy levers remain misunderstood, ignored, and underappreciated—and thus unpulled.

The basic problem here is due to a different form of evolutionary conservatism, which doesn't have to do with the psychological constraints on expanding the circle of moral concern, but with the cognitive constraints on the processing of complex information that come with the fundamental architecture of the mind. Dual process theories of cognition hold that human cognition has a bifurcated set-up: one type of cognition is quick and efficient, but inflexible, impenetrable by top-down processes, and prone to error. This type of cognition is referred to as "System I" (Evans & Frankish, 2009). System II, on the other hand, is flexible, precise, and can handle complex novel problems. Because of that, however, it is also resource-expensive, computationally limited, and depletable.

Now, no matter how cerebral you fancy yourself to be, the fact of the matter is that 99% of everyone's cognition runs on the fuel of intuition. Intuitive cognition works perfectly fine for situations it has been equipped to deal with either by evolutionary, or by cultural, or by individual learning. Complex novel problems, on the other hand, remain intuitively intractable. Modern societies in particular constitute "hostile" environments for our automatic mind, which doesn't know how deal with counterintuitive issues such as advanced technology or life among millions of strangers more generally. At the same time, this means that our intuitive cognition will fall back on a variety of useless or harmful beliefs in response to the most pressing modern problems such as how to organize an economy, how to set up a functioning political system, or how to handle crime, conflict, or climate change. These beliefs are difficult to rein in, because to make things worse, the capacities of intuitive override our mind has at its disposal are limited, and intuitions tend not to come flagged as needing closer scrutiny (Stanovich, 2011).

In order to combat this problem, Heath (2014b) recommends the creation of "slow institutions" to counterbalance the bugs of our intuitive mind. Consider the fact that there is a fundamental difference in the respective newsworthiness of good and bad events. Bad events can happen instantly, good developments usually take a lot of time to unfold. This means that even when things fundamentally improve, there can be a perception that there is a constant stream of catastrophes. Over time, this perception turns into the impression that existing problems pile up and remain unsolved (as solved problems also tend to remain unreported), which can motivate the felt need for radical pseudosolutions such as Brexit or the closing of the Southern US border.

It is here that the bootstrapping problem of how to get on the track of good institutions looms especially large. Having such institutions would, of course, be a very good thing. But how do we establish them without

explaining why we need them first? And how do we get around the fact that explaining them will take a crash course in cognitive science and social theory?

Explaining the case for progressive social change, and the extent to which it has already happened, inevitably involves graphs, numbers, and difficult words which are either boring or incomprehensible or both to the vast majority of people. As Chris Rock once said in a different context, the need to make sense can sometimes be a disadvantage; various noble attempts to make such data regarding crime, poverty, the effects of immigration and commerce, global health, or security intuitively accessible do exist (Rosling, 2018), but they will never escape the fact that the nerds are bringing knives to the populists' gun fight.

7.10   Reflexive Institutions

The prospects of progress are always shaky. No set of institutions, however smart, can do anything about this basic fact. Then again, social regress isn't engineered by infallible evil geniuses, either. At the very least, defenders of progress can draw on the empirical evidence from evolutionary and social psychology to make sure certain regressive tendencies such as outgroup thinking and dehumanization are moderated (Bandura, 1999).

Reflexive institutions are institutions that specifically make use of this evidence. Livingstone Smith (2011) argues that racial or ethnic dehumanization may be due to forms of innate essentialist thinking, which is encouraged by certain forms of "generic" speech that could be institutionally regulated. Efforts at dehumanization exploit that we think of other human beings and the groups they belong to as distinct kinds with a metaphysically stable, ethnoracial core. This core—which is purported to lie beneath the observable surface of appearing like any other human—can then be reconceptualized as betraying an entity's real nature as subhuman soulless vermin. In a nice twist, Buchanan and Powell (2018, p. 227) note how the tendency to dehumanize outgroup members actually pays a subtle tribute to inclusivist moral progress, since dehumanization is only required to justify discrimination or genocide if one implicitly already acknowledges that, in principle, all humans deserve equal moral recognition. Institutionally speaking, the tendency to speak of "Jews", "Mexicans", "criminals", or "Hillbillies" in such generic terms could be culturally flagged as morally odious and discouraged accordingly.

When we remove personal information that is known to trigger pernicious discriminatory biases from applications, or when identifying details are left out of journal submissions to allow for anonymous peer review, we engage in reflexive institution building (Uhlmann & Cohen, 2005). The political separation of powers, constitutional constraints, and the general

system of checks and balances modern democracies have established can be seen in a similar context: as reflexively informed self-binding mechanisms that safeguard contemporary societies, however imperfectly, against the forces of irrationality that have the potential to undermine it from within.

Another example for this would be the smart regulation against so-called "fake news". Our mind is not transparent to itself. For instance, when asked about my beliefs regarding a subject matter, I cannot simply introspectively access my belief storage as if rummaging through a box of childhood memorabilia. Instead, my main evidence regarding what I believe about an issue is what I have said or written about it before. When we ascribe beliefs to ourselves, we have to use essentially the same mechanisms that we use to ascribe beliefs to other (Carruthers, 2009; Hall et al., 2012), foremost of which are external cues. These pathways can be hijacked by fake news. If I hear them often enough, it may over time become more difficult for me to filter them out of the raw material I use to decide what I actually do believe (Levy, 2017). This problem is further exacerbated by the fact that our mind functions in a more "Spinozist" rather than "Cartesian" fashion (Mandelbaum & Quilty-Dunn, 2015). Belief acceptance is the default, and largely works automatically; belief rejection, on the other hand, is the exception, and requires effort. What we hear, we tend to believe until we have reason not to and the resources to appreciate those reasons. And because beliefs are inferentially promiscuous, we also implicitly accept what follows from the beliefs so acquired. It is thus worth thinking about how to increase our own and our democracies' recalcitrance towards systematic misinformation, and how to regulate the formal and informal communication institutions and media to counterbalance the proliferation of bullshit, lies, and other reckless falsehoods (Rini, 2017).

7.11 Extracting Norms From Institutions

If morally progressive gains are best secured institutionally, and if the inner workings of institutions are not readily transparent to those swept up as participants in those practices, it becomes an interesting question how individual agents should relate, *within their stream of agential deliberation*, to those institutions the functionings of which may elude them.

Here, the distinction between consequentialism and deontology starts to play an unexpected Hegelian role. Institutions look consequentialist from the outside but deontological from the inside. Their overall rationale must, to a certain extent, remain opaque to those embroiled in them. The cunning of reason, then, consists in how this rationale reaches through individuals' intentions to fulfill the basic purpose of the institution via their unwilling and unwitting execution.

The basic thought is this: institutions have a "mixed" normative logic. Usually, they will look "consequentialist" (designed to promote some outcome, such as efficiency, or cooperation, or the production of knowledge, or collective decision-making, or the allocation of mates) from the outside (the "observer perspective"), but "deontological" from the inside (the "participant perspective"). Agents operating within those institutions need to treat the rules governing their institutional interactions as to a certain extent "categorically binding" and not up for *ad hoc* recalculation. Only when agents within institutions treat their normative infrastructure as a form of genuine deontic constraint can these institutions satisfy their underlying consequentialist rationale. This is the rational core in the mystical shell of Hegel's *Sittlichkeit*.

Consider the market again: the reason why markets are worth having in the first place is that they gravitate towards more efficiently allocated goods and more efficient uses of resources due to the way trade is sensitive to the information contained in competitively generated price signals (Satz, 2010, see also Hayek, 1945). This only works, however, if a certain set of extremely demanding conditions (such as full information that is freely available, absence of externalities and an exhaustive system of property rights, no transaction costs, and no monopolies) obtains. Since these conditions can never be fully realized in the actual world, we need to count on the agents that participate in those institutions to act on the basis of mid-level norms general compliance with which would promote the efficiency rationale of the market after all (Heath, 2014a).

These norms take the form of relatively simple and general deontic principles—don't lie to your customers, don't pollute, don't fix prices—which articulate the ground rules of the institution in a way that is so densely compiled as to make the underlying justification of those rules invisible to the naked eye of agential moral reflection. The internal ethical norms of sports, politics, or science are likewise sensitive to the opaque rationale that governs their infrastructure. It is not immediately clear *why* I shouldn't kneecap my opponents, accept bribes or fake data. Only the underlying institutional rationale of these practices—furthering human physical excellence, promoting the common good, generating knowledge—can explain why these domain-specific rules of proper conduct apply. In this way, it is possible to extract genuine moral norms—understood as norms that are supposed to figure in how moral agents deliberate about what to do—from supraindividual social practices.

Does the underlying rationale of an institution have to remain hidden to those acting within the institutions? Sometimes, but not always. Consider marriage: as a culture, we talk about marriage on terms of love, soulmates, the value of starting a family, and so on. It would probably undermine the practice quite considerably if people started talking about

it as an institutional solution for the distribution of mating partners. The same holds for democracy. According to the official version democracies embody the values of citizenship and community. Unofficially, they are about organizing the peaceful transition of power, and individual citizens with their inconsequential votes don't make a difference to this (Freiman). Some tribes determine where to hunt game by finding by looking at the flight patterns of birds. Here, the official story has it that, via those patterns, the Gods reveal where to hunt. In reality, the procedure works because it effectively randomizes. All of these unofficial rationales would probably undermine the functioning of their respective institutions at least a little bit. Consider the market one last time: the internal logic of the market is competitive. But the "ethos" that people need to act upon within the market should still be based on being a trustworthy merchant, offering good products for low prices, respecting the customer, and so on. When people start moving their competitive logic from the previously untransparent institutional level to the level of personal motivational orientations, the institutions can start to malfunction.

Note

1. This chapter draws on material previously published in: Sauer, H. (2019b). Butchering benevolence moral progress beyond the expanding circle. *Ethical Theory and Moral Practice*, 22(1), 153–167.

8 The Long March

Does Moral Progress Require Moral Facts?

Introduction

Perhaps there is moral progress. But is it progress towards the moral truth? The previous chapter concludes my substantive argument. I have argued that social change is biased in favor of moral improvement. But the form of progress we see is a progress without goals: cultural evolution generates moral innovations that both enjoy a competitive advantage in the game of intergenerational transmission of social practices *and* adequately track legitimate human interests. Various types of moral progress (the elimination of harmful norms, improvements in people's welfare, etc.) are driven by various mechanisms of moral progress (epistemic, cultural, functional, etc.) and stored in various institutional arrangements (bypassing institutions, proxy institutions, etc.). Over time, society becomes a morally better place.

From a metaethical point of view, moral teleology seems to be most easily married with moral realism. If there is moral progress, then it seems natural to assume that it is progress towards the moral truth. Over time, one may think that we improve the ratio of true to false moral beliefs, either by acquiring new true ones or by shedding old false ones or both. What makes our beliefs true—and this is the distinctively *realist* part—are objective moral facts that obtain independently of individual or collective minds. As moral progress unfolds, light dawns gradually upon a preexisting moral landscape.

Recent debates regarding moral progress have identified a number of important issues surrounding the topic of moral progress, such as what concept of moral progress one should work with, what moral progress consists in, whether moral progress, so characterized, is feasible in light of the evolutionarily inherited constraints on our psychology, or which institutions, if any, are required to maintain or promote it. As far as the meta-ethical issues surrounding moral progress are concerned, moral realism is frequently treated as the *default* position to make progress intelligible.

Moral progress, it seems, requires the existence of objective, mind-independent moral facts (Kitcher, 2011, p. 186ff.; Uttich et al., 2014).

The viability of moral realism must at the very least be conceded—or so it is alleged—for all this talk about moral progress to make sense. The basic idea behind this is a simple one, for if there is moral progress, it indeed seems fair to ask what it is progress towards. A natural answer to this question is that society progresses (if indeed it does) towards a greater realization of what is objectively morally required, and that the moral beliefs of those undergoing this long march come to further approximate the mind-independent truth.

I want to resist this *argument from moral progress to moral facts*. In order to do so, I will first explain the case for moral realism that is supposed to flow from the reality of moral progress. That there is such a thing as moral progress is an assumption which is granted rather than argued for in this debate, and I will grant it here as well. What is at issue for now is not whether there is any moral progress, or how best to describe and explain it, but whether the existence of moral progress requires the existence of moral facts. And this question can of course be addressed without making any commitment whatever to the actuality of moral progress.

In what follows, I will first lay out this "progressive" case for moral realism (section 8.1). I will briefly address and reject an important argument according to which cases of moral conversion only make sense in light of moral realism (section 8.2). I will then zoom in on what I take to be the most recent, and most promising, formulation of progressive realism, according to which specific recent trends in social development—such as increasing liberalization—call for a realist explanation (section 8.3). I will then proceed to discuss the three main alternatives to the realist account. According to the first, moral progress can be modeled after theoretical progress in the sciences, such that moral judgments can be likened to "ratified conjectures" of the kind we also find in, say, chemistry (section 8.4). The second has it that moral progress can be understood in functionalist terms, such that moral progress occurs when the circumstances allow morality to more fully or adequately perform its function (section 8.5). In section 8.6, I will discuss a proposal according to which moral progress should be measured in terms of improved problem-solving. These three approaches are supposed to provide an alternative to the progressive case for realism.

Ultimately, I will reject all of these accounts. The reason for this, in a nutshell, is that their explanandum does not exist. All of the aforementioned metaethical approaches, whether realist or not, start from the assumption that moral progress either involves or consists in moral *convergence*. The question,

then, is whether such convergence is best explained in realist or non-realist terms. I will show, however, that convergence does not have to be explained, because it is a myth: the idea that people depart from an original state of moral diversity, after which they—slowly and gradually—come to converge on a set of shared values such that this process of convergence calls out for an explanation is false, because there is no original state of moral diversity to begin with. Moral realism isn't a worse metaethical explanation for moral progress not because there are better, non-realist explanations for moral convergence, but because moral convergence isn't real. There is almost universal evaluative agreement at the fundamental metaethically relevant level.[1]

8.1 From Moral Progress to Moral Facts: The Simple Argument

There is a simple argument from moral progress to moral realism that goes something like this:

The Simple Argument

(1) There is moral progress.
(2) Moral progress requires (the existence of) moral facts. Therefore,
(3) There are moral facts.

This is a transcendental argument that establishes the existence of something whose existence is not independently known (moral facts) in terms of its being a condition of the possibility of something else (moral progress) whose existence is (per assumption) independently known.

There is an important ambiguity in the second premise in that "requires" can be given a *causal* or, alternatively, a merely *hermeneutical* reading. Item 2 in the list here could be taken to mean that the existence of moral facts must be assumed in light of the reality of moral progress because objective moral truths are required to *make sense* of social change as society-wide moral improvements. This is the hermeneutical reading. For this to be true, it is not necessary that said moral truths played any part in actually driving those improvements, as long as they provide the conceptual framework that allows us to evaluate them as improvements in hindsight.

The causal reading is different from this. Even if it is known that something exists, and that that something is helpful, or perhaps even essential, for rendering some other thing intelligible, it does not yet follow that said thing played a role in bringing the other thing about. Consider the laws of logic, or various mathematical truths. These may be useful or even inevitable for shedding light on how proper thinking or physical laws work without having any causal connection whatsoever to why thinking occurs,

or why nature works the way it does. Likewise, it is one thing to say that moral facts exist, and that they must exist in order for moral progress to be conceivable; it is another thing to say that the moral facts were *causally efficacious* in bringing moral progress about (Luco, 2019).

Since I am inclined to agree that there is moral progress, I will focus on the second premise. Moreover, for the purposes of this chapter, I will restrict my discussion to the weaker hermeneutical reading. This will make my job harder, however, because the weaker reading establishes a lighter burden of proof for the moral realist. For her argument to go through, she does not have to show that the moral facts made progress happen, only that without them, progress is hard to understand.

8.2   The Case of Conversion

Before I take a closer look at the realist case for moral facts on the basis of moral progress, I want to consider a slightly more sophisticated version of the simple argument presented in the previous section.

Some argue that we should think of moral progress at the level of society as requiring moral facts because of an analogy with what goes in cases of moral change at the level of the individual. What is usually described as moral progress when groups are concerned is sometimes described as conversion when we are talking about individuals. It would be implausible to suggest that moral conversion—understood as a more or less prompt and more or less drastic change in one's moral outlook—simply doesn't occur. The very possibility of such From-Saul-to-Paul-cases of individual moral reform, again, seems to show that something like moral realism must be correct.

This suggestion is supported, first and foremost, by the fact that such instances of conversion are typically experienced as being due to forces external to the moralizer. Experiences of conversion are rarely chosen, and often unwelcome. The fraud or the philanderer would often prefer to remain unconverted, continuing to reap the monetary or sexual benefits of their hitherto ways. Conversion is thus external in that it impinges and forces itself upon the agent as something unpleasant yet inescapable. This is, one may think, because mind-independent moral facts assert their authority over us—*nolens volens*, as it were.

What it means for episodes of conversion to be external has been clearly articulated by John McDowell (1998):

> The idea of conversion would function here as the idea of an intelligible shift in motivational orientation that is exactly *not* effected by inducing a person to discover, by practical reasoning controlled by existing motivations, some internal reasons that he did not previously realize he had.
>
> (102)

Conversions are evaluative shifts that make sense (they are "intelligible", and not merely caused by an iron rod that blew out a portion of one's brain) but cannot be arrived via inferences from already accepted propositional attitudes (including one's desires).

The resulting argument might go as follows:

The Argument From Conversion

(1) There are cases of moral conversion.
(2) The causes of such conversion are (experienced as) external to the converted.
(3) Conversion by forces (experienced as) external to the converted require moral facts. Therefore,
(4) There are moral facts.

A lot hangs, of course, on whether conversion that is *experienced* as originating from something external really does originate from something external. Given the frailty of our introspective powers, I doubt that this inference will typically be warranted.

But that is not my concern here; because even if experience of externality indicates objective externality, it remains unclear why the existence of such external sources of individual moral development should speak to the issue of moral realism or anti-realism at all. Waller (1992) puts this point vividly:

> Fundamental changes in values and moral principles do occur, and in at least some cases they are (as the moral conversion argument requires) experienced as changes compelled by forces outside of us. However, moral realism is not required to account for changes even at that level. To see how basic values might change (without moral realism), consider how one's basic values originate. . . . I may be able to trace the (independent, external) causes of my fundamental moral beliefs, and recognize that . . . a different causal process would have yielded different values, yet not waver in my moral commitments. . . . My belief that it is wrong to cause unnecessary suffering to any sentient creature . . . resulted from seeing *Bambi* at a particularly impressionable age; I would be a contended carnivore had I see *Rambo* instead.
>
> (133f.)

Moreover, even if experiencing something as external were evidence of its being external; and even if something's being external did require moral facts; the question would remain wide open in what sense, and why, cases of individual moral conversion should support extrapolation to the supraindividual

case of moral progress at the level of larger groups or societies. Indeed, it is unclear what it would even mean for a society-wide moral change be driven by external forces. External to what, and coming from where, and acting on what? This is especially true if one's reading of "external" is cashed out, as suggested earlier, in terms of what lies outside of an agent's subjective motivational set, because whole society's do not have those.

Finally, the biggest problem with the argument from conversion is that it cannot plausibly account, in realist terms, for conversion to the *worse* (Waller, 1992, p. 135f.). For which moral facts explain those cases of moral deterioration? The argument, in effect, presupposes that moral change will always occur for the better to justify its inference from moral progress to moral facts. If we drop that assumption, then the argument ceases to go through without assuming what it is supposed to show, namely that moral facts must be stipulated to account for moral progress.

8.3   A Realist Account of Moral Progress

Moral realism has a serious dialectical advantage because it seems uniquely able to explain why moral progress—or social change more generally—appears to have a *direction*. If there are mind-independent moral facts which (collections of) people can come to appreciate, then this would explain the appearance of goal-directedness. Moral facts have always been around, and as it gradually dawns upon us what they are, we are moving in their direction.

Moral realism also has a complementary disadvantage in that it seems to *predict* moral progress, and thus the prospects of defending it remain hostage to its occurrence. As I have explained earlier, my task here is to see why, if the existence of moral progress is granted, moral realism is supposed to be the most suitable metaethical candidate to account for it. Conversely, moral realism seems to stand and fall with evidence for moral progress. If there isn't any, then realists have to explain *this* fact. Other, non-realist metaethical accounts don't seem to share the same explanatory burden (see, however, Arruda, 2017).

If there are moral facts, how could they play a role in bringing about moral progress? Perhaps the most powerful recent argument for the link between realism and progress has been developed by Michael Huemer (2016a). Huemer offers a best explanation for recent trends towards social "liberalization", arguing that many important social developments of the past 300 years (or so) bear the marks of improvements in the cognitive appreciation of objective moral facts.

When it comes to identifying the explanandum of his theory, Huemer relies on evidence regarding decreases—often drastic ones—in the pervasiveness of war and murder, torture and execution, slavery, and racism,

sexism, or ableism, combined with corresponding increases in democratization, decolonization, and socially widespread intolerance towards the practices just mentioned. These developments, Huemer argues, are not random and contingent. Rather, they reflect a tendency of certain societies to move towards a coherent set of ethical ideals that he refers to as "liberalism". Liberalism, so characterized, "(1) recognizes the moral equality of persons, (2) promotes respect for the dignity of the individual, and (3) opposes gratuitous coercion and violence" (1987).

Pinker (2011) attributes such developments to a series of revolutions he labels the *pacification process*, the *civilizing process*, the *humanitarian revolution*, the *long peace*, the *new peace*, and the *rights revolutions*. Huemer, however, wants to go further and argue that these trends are epistemically driven. An important part of the explanation for why these things happened is that liberalism is the *objectively correct* moral outlook.

In order for this story to count as the "best" explanation of the data, we need to compare it to its rivals. Why can non-realist accounts, in particular evolutionary debunking arguments, not account for these changes, or not as plausibly? The evolutionary origins of our basic evaluative dispositions seem to provide grounds for moral skepticism (Street, 2006). There is no reason to think that if there are mind-independent moral facts, selective pressures would have enabled us to access them in any veridical way. The main problem, however, for any evolutionary explanation of the liberal trends Huemer talks about is their recency. For the most part, evolutionary developments remain shackled to a generational pace; selective pressures can unfold only by creating new beings, and seeing how they fare, which takes time. Three hundred years simply isn't long enough to explain changes as big as the ones mentioned by Huemer. This immunizes—partially, as we will see—these developments from the threat of epistemic debunking.

An important part of the realist account of moral progress is to supply a sketch of the causal pathway from moral facts to society-wide moral improvements. How do these improvements spread? Here, Huemer tells a story of piecemeal moral improvement, fostered by the reduction of biases and the step-by-step proliferation of such reductions. Liberalism, according to Huemer, results from overcoming biases—in group favoritism, gender discrimination, violent exclusion, and so on. Initially, everyone's moral beliefs are heavily biased. However, at any given point in time, some individuals will have a lower propensity for bias, or greater cognitive ability, relative to their contemporaries, to recognize them as arbitrary and unjustified. This cognitive elite, in turn, will typically occupy positions of comparatively greater social influence. More cognitively able individuals on average tend to be among the leaders, scholars, and preachers of a society. Over time, their influence will make sure that society develops in the direction of reduced biases, that is, liberalism. And since biases induce a skewed picture

of reality, reducing them will yield a more accurate and less distorted representation of the truth. Realism has been vindicated because it can uniquely account for the direction *and* content of recent moral revolutions.

Even if Huemer's story is correct, however, it doesn't seem able to escape certain GIGO worries. It would be an overstatement to suggest that "garbage in/garbage out" problems are generally lethal to those who hope to resist evolutionary debunking. Indeed, there do seem to be plausible attempts to explain how legitimate moral values could have emerged from more basic evolved evaluative dispositions:

> [I]t was likely important for our Pleistocene ancestors to understand the application of evaluative concepts in connection with relevant standards. They needed to make accurate evaluative judgments about *good* and *bad* dwelling places, or hunting partners, fighters, and mushrooms, and related normative judgments such as that one *ought* not to eat the little brown mushrooms or to fight with Big Oog. Moral judgments obviously go beyond these sorts of things, but just as in the other cases, they can be seen as an extension of such thinking. They still involve employing evaluative and normative concepts in connection with standards and ends, though now conceived as standards and ends defining *what it is to live well all things considered*, rather than just narrow standards of edibility or safety. . . . We discover the evil of racist voting laws, for example, by gaining empirical knowledge about the irrelevance of race to what matters to responsible voting, and by reflecting on the significance of such facts in light of ongoing experience of human life and the possibilities of good and harm it offers us, as part of forming a conception of what it is for human beings to live well. Why should this sort of intelligent extension of evolutionarily influenced evaluative judgment be thought any more problematic in principle than parallel extensions in other domains?
>
> (Fitzpatrick, 2015, p. 889f.)

The problem is that where Fitzpatrick focuses on the *gi*, Huemer focuses on the *go*: the former argument aims to show that the inputs for evaluative capacities of moral reasoning and generalization to operate upon don't have to be seen as epistemically worthless garbage; the latter argument wants to show that the outputs of moral reasoning can qualify as knowledge of objective facts regardless of the quality of the inputs, because the moral shifts cited by Huemer are thoroughly beyond the reach of evolutionary explanations. This is indeed what Huemer seems to suggest when he writes that the liberal trajectory of modern society appears to result from an all-around *rejection* of biases (thus leaving the garbage behind), rather than their purification (which would leave us with what is still, albeit very clean, garbage).

This strategy, however, risks scientific credibility because it would be implausible to suggest that *some* sort of opposition to violence, and *some* sort of egalitarian concern, and *some* sort of recognition of human worth have an evolutionary rationale, which was then further shaped and transformed by subsequent bias-reducing reasoning. Huemer is surely right to point out that there is no *logical* incoherence in restricting moral concern and the associated moral status merely to one's in-group; one does not have to make this claim, however, in order to argue that the trend of liberalization results, in part, from the removal of morally irrelevant biases from one's concern. But if having such concern is epistemic "garbage" to begin with, then purging it from bias won't do much to make it ultimately justified. Moreover, though Huemer also seems right to suggest that there is nothing logically contradictory in refusing to extend moral recognition to members of the out group, this is not where consistency reasoning exerts its force (Campbell & Kumar, 2012): the inconsistency lies in the fact that the features which supposedly ground moral status for in-group members—sentience, or autonomy, or reason, or any other aspect of personhood—turn out, empirically, to be found to the same degree in members of the outgroup. And if these features earn someone moral status, it then becomes hard to consistently justify why beings that evidently share these features should be denied moral status, even if there is nothing *logically* incoherent in supposing that only people with a certain skin color or gender deserve moral respect.

Two final problems with Huemer's realist account are that while evolutionary explanations cannot account for the *recency* of the social developments Huemer focuses on—evolution simply takes too long for it to explain the shift towards liberalism—Huemer's proposed explanation (removal of bias through reasoning) cannot provide an answer to it *either. Humans have been capable of reasoning for a very long time; influential people have* argued against bias for a long time; so why did these liberal shifts happen only now and not earlier? This suggests that the specific cause of these developments isn't epistemic at all. And indeed, as I discussed in Chapter 4, we now do have an evolutionary explanation of the shifts cited by Huemer as calling for a realist explanation, namely a cultural evolutionary one: liberalization was driven by the cultural evolution of WEIRD psychology: over time, the demolition of intensive kinship structures by the medieval church's marriage and family program has led to increases in analytic thinking, self-control, impersonal prosociality, and voluntarism as principles of social organization. We can tell that this explanation does not require the existence of moral facts in a straightforward way: Henrich develops his explanation without mentioning objective moral facts even once. Theories of cultural evolution have the conceptual resources to supply a non-realist explanation of progressive trends that draws on the forces of cultural attraction, socialization, economic growth, or solutions to cooperative problems (Hopster, 2020; Cofnas, 2019).

8.4 Anti-Realism: Moral and Scientific Progress, Functionalism, and Problem-Solving

Some have tried to maintain the possibility of moral progress without moral facts by likening moral to scientific progress. This seems baffling at first, for isn't scientific progress precisely the kind of thing that vindicates a domain's realist commitments? If this is so, then the analogy to progress in the sciences seems like one that metaethical anti-realists should avoid like the devil avoids holy water.

However, this impression may turn out to be misleading. Authors such as Catherine Wilson (2010) press precisely this point, and emphasize that they merely want to highlight the similarities between the dynamics of theory change in the sciences and the moral changes which even anti-realists ate inclined to welcome as progress. This is supposed to become possible by treating moral propositions as "ratified [theoretical] conjectures" (111), "analogous to 'Oxygen is the principle of combustion'" (99). Here, the key thing to realize is that, like scientific statements, moral propositions face the tribunal of experience in packs. A judgment such as *Late-term abortion is almost always wrong* would then serve as a "proxy" (100) for a more or less tightly connected web of inferentially structured and mutually support-giving or support-demanding beliefs about what a fetus is, what the effects of abortions on women, men, and society in general may be, how abortions are performed, and so on:

> Once the relevant conjectures have been ratified, we can truly claim that it was "always true" that the earth went round the sun and that the infanticidal prerogatives of Roman fathers were morally wrong. We can even claim, in the counterfactual mode, that they would have been true even if the earth had been annihilated by an asteroid before anyone came to know them. Further, epistemic agents may hold true moral beliefs that they do not know to be true—because they are unable to assist in their ratification and lack the right kind of causal connection to the actual ratifiers.
>
> (111)

But it seems unclear how, in the case of moral beliefs and their (possible or actual) truth, anything short of objective moral facts could be robust enough to support the counterfactuals just mentioned. The reason why we do not hesitate, in the non-moral sciences, to say things such as "it has always been true that oxygen explains combustion" is that we are confident that, in addition to the general theoretical conjectures that figure in that web of beliefs which beliefs in closer proximity to observation figure

in as well, there are also mind-independent natural facts that these observations answer to, however clumsily they may do so. And this, precisely, is what is supposed not to hold for the moral domain.

Once one starts looking beyond scientific or, more generally, epistemic endeavors altogether, it becomes obvious that approximation towards the truth isn't the only measure of progress. In many cases, we think there is progress when a device gets better at realizing a designated purpose, or when the performance of an organ is restored to previous levels after an illness. In both cases, it makes little or no sense to speak of an increase in truth or knowledge. Truths about what? Knowledge of what?

Rather, in those cases we judge progress in terms of how, and how well, a function is discharged. Here, the criteria for improvements are thoroughly non-alethic: an improvement in function can consist in achieving more with less, or achieving the same thing more elegantly, reliably, frequently, or accurately. Ever since its invention somewhere in the depths of prehistory, wheels have become much better, though not by becoming more true (whatever that would mean); instead, wheels now simply dwarf the performance of their ancient predecessors in all respects ranging from durability over friction to design.

It should be conceded right away that the assimilation of moral progress to functional improvements has struck many as obviously incorrect, and indeed morally objectionable in its own right:

> Whatever exactly moral progress amounts to, it certainly involves at least the following: the abolition of war and slavery, the reduction of poverty and class privilege, the extension of liberty, the empowerment of marginalized groups, and respect for animals and nature. If these concerns are not captured by an account of moral progress then I (and I think many of us) literally would not know what we are talking about when we talk about moral progress. In any case there is no reason to believe that a society that is morally progressive in these terms is on the way to the greater biological fitness of its members than one that is not. Nor is it plausible to suppose that a society whose members' genes are better represented in the next generation is one that we would regard as morally progressive (it may, for example, be a society that is characterized by a high incidence of rape).
>
> (Jamieson, 2017, p. 171)

Such proposals do not just seem moral odious, but methodologically suspicious: inevitably, they remain haunted by the specter of the naturalistic fallacy.

Most would also agree, however, that morality does have a function. There seems to be a broad consensus on what that function, or part of it, consists in: moral norms secure cooperation and prevent free-riding. They

were adaptive because they allowed cooperators to capture the associated gains resulting from reciprocal exchange (as in mutual back-scratching), risk-pooling (as in large game hunting) or economies of scale (as in everything that can only be done by more than one person).

In Kitcher's (2011) jargon, the function of morality is to curb what he refers to as "altruism failures":

> The tensions and fragilities of hominid (and chimpanzee) social life arise from the limited altruism of their participants. Altruism failures lead to conflict, to pain inflicted, to rough discipline, and lengthy peace-making. To the extent altruism failures can be avoided, life goes more smoothly, with increased opportunities for cooperation and, consequently, greater mutual benefits. Group members satisfy more of their desires and protest less.
>
> (222)

The good news for those unconvinced by moral realism is that functionalist descriptions of the phenomena provide criteria for progress that do not reduce to stories about an approximation towards the moral truth. Since it is possible to identify a function of morality, the prospects of understanding moral progress without moral facts are reanimated.

Another influential functionalist account of moral progress is due to Peter Railton (1986). Railton ties the possibility of moral progress even more closely to the function of moral norms, since on his account, progressive developments are often directly motivated by the alienation, discrimination, and disenfranchisement of marginalized groups. The (peaceful or violent) mobilization of previously not properly recognized groups, in turn, leads to social unrest and upheaval (191ff.). The attenuation of such conflicts and disturbances is paid for in the currency of progress: the demands of the downtrodden are satisfied, and order is restored.

Buchanan and Powell (2018, p. 77ff.) want to push back against functionalist accounts of moral progress. Recall that Kitcher holds that morality is about overcoming altruism failures. And for Railton, morality's progress is tied to (dys)functionality because disenfranchised groups mobilize, thereby threatening stability and social cohesion. Buchanan and Powell criticize both of these theories for their purported inability to account for subject-centered morality, that is, the existence of moral outlooks that afford moral status not on the basis of one's role as a strategically relevant cooperator, but on the basis of being a certain kind of thing such as a subject. Subject-centered morality thus recognizes the moral standing of very young infants, severely disabled people, future persons, non-human animals, or indeed all sentient beings.

According to Buchanan and Powell, functionalist accounts of moral progress thus face a tough choice: either, they refuse to classify the inclusivist shifts that led to the purely subject-centered attribution of moral status to the groups just mentioned as progress, which is normatively implausible; or they cannot account, in their desired functionalist terms, for what is arguably one of the most important forms of moral progress, which is theoretically unsatisfying.

Buchanan and Powell hold that Kitcher cannot explain the inclusivist anomaly because altruism always remains parochial, for simple reasons of adaptiveness. Railton, on the other hand, cannot account for it because many individuals or groups (the disabled, young children, etc.) cannot mobilize. Notice, however, how a combination of Kitcher's and Railton's argument *can* explain all that, because with Kitcher, we can argue that altruism for kin, even very young and/or disabled kin, will be motivationally strong enough that, with Railton, family members will mobilize *on behalf* of their non-strategically relevant inner circle.

Wilson (2010, p. 104) also doubts that functionalism can be the whole story. She writes:

> Philip Kitcher suggests that the function of morality is "the enhancement of social cohesion via the amplification of psychologically altruistic dispositions," together with the expansion of possible social roles for individuals to adopt. The chief difficulty with this characterization is that its second clause seems to have been added as an afterthought, on the grounds that social cohesion may well be served by rigid assignation to social roles in ways that come to be recognized as immoral. Yet while the expansion of social roles may be a criterion of moral progress and a worthy moral aim, it is difficult to see that it is the *function* of morality. (104)

Again, the problem is that functional improvements can be extensionally inadequate when it comes to capturing instances of moral progress. Functional improvements may constitute moral regress, and moral improvements may not be unambiguously functional. On the other hand, this reply underestimates the extent to which Kitcher's apparent *ad hoc* amendment regarding the emancipatory potential of more flexible roles can be given a functional explanation. Even if morality's function, in the abstract, is to secure cooperation, and even if discharging this function entailed, in the distal environment of our ancestors, parochial altruism, this would not mean that, as our somewhat younger ancestors began to dwell in larger protourban groups, the flexibility of role adoption could not have had any functional benefits.

A wholehearted rejection of functionalism would be implausible anyway. It would be implausible to deny that morality has a *de facto* function (like facilitating cooperation, among other things). And if that is so, it would then also be implausible to assert that the issue of whether or not morality gets better or worse at performing this function has *no bearing whatsoever* on the issue of moral progress. Surely, the functionalist account has *some* significance for progress. It is merely unclear how it does, and to what extent, and whether it is the whole story.

Some who want to resist the argument from moral progress to moral facts hope to model moral after scientific progress. Some think that the criteria for moral progress are functionalist. A third and final non-realist account of ethical progress I wish to consider is a *pragmatist* one. Amanda Roth (2012), for instance, agrees with many of the paradigm examples for moral progress mentioned in the literature—such as the success of feminism or the civil rights movement—but wants to explain these successes (as well as what makes them successes rather than failures) in terms if improvements in problem-solving.

One reason for eschewing a realist account of moral progress is due to its proclivity towards utopian thinking, which one may find independently unappealing. One problem with utopian thinking has to do with how it proliferates justifications of heinousness: if the holy land of ultimate moral perfection for everyone (who is left) is within arm's reach, then *any* action that promises to get us there is, just for that reason, defensible. Second is establishing too close connection between moral progress and a utopian end-state as its final goal risks moral skepticism. We do not, simply put, know what that end-state is like or how to get there, so we can never rule out that we are on the wrong track.

At the same time, Roth argues that the utopian temptation is quite understandable, for competing non-utopian accounts such as Moody-Adams' or Rorty's fail to explain what distinguishes genuine progress from mere change, and fall back on either an evolutionary story of arbitrary succession or an overly restrictive story that disallows the adoption of new values and ends. Progress as problem-solving is supposed to get around this issue.

What, according to Roth, is a "problem", and what do "solutions" look like such that they can be classified as progress? She writes:

> By "problem" I mean to pick out a certain sort of experience of ourselves, our beliefs, our values, and the world. Problems arise when we experience trouble, difficulty, or conflict—when there is disharmony amongst our empirical beliefs, our values, and our experience of living.
> (Roth, 2012, p. 8)

So far, so pragmatist—this is Heidegger's "obstinacy of the stuff", but this time in an ethical context. The important thing to realize about solutions to such cases in which problems emerge or first become apparent are not just instrumental discoveries of new ways of realizing one's otherwise fixed and unrevisable ends. Problem-solving is a much more holistic process in which everything is, in principle, up for grabs: "[A]ll of the various aspects of our system are revisable. Our goals, our values, our epistemic norms, our theoretical beliefs—all must face the test of experience" (13).

Let me mention two problems with the problem-solving account as an alternative to moral realism, one metaethical and one normative. For one thing, it is not clear why an analysis of moral progress in terms of problem-solving should be considered an alternative to realism at all. It does seem plausible to suggest that what counts as a problem and what does not is essentially relative to agents' desires, values, plans and goals, and in that sense, problems have a non-realist aspect to them. What counts as a solution to a problem, however, seems to have an irreducibly factual aspect as well: not all suggested solutions to even simple problems *work*, and indeed most usually do not; moreover, the distinction between the former and the latter category seems to be supplied by mind-independent criteria. Which ends individual or collective agents want to achieve may be up to them, but how to solve the problems standing in the way of their realization does not—it is determined by the facts.

For another thing—and this brings me to the normative problem—that a problem has been solved seems to be neither necessary nor sufficient for moral progress to have occurred. What problem has been solved, for instance, by replacing ancient readings of the *lex talionis*, according to which if you have raped my daughter, I acquire the equivalent right to rape *your* daughter, with the modern interpretation according to which appropriate sanctions should exclusively befall the perpetrator (Kitcher, 2011)? It is close to a criterion of adequacy for an account of moral progress that it be able to identify such a case as an ethical improvement, but which problem has been solved by this development, besides the problem that the older reading was morally wrong, seems unclear. At the very least, describing this change in terms of a problem that has been solved does not seem to capture what happened in the most fitting way. There can be moral progress that doesn't strictly speaking solve any problems.

It seems equally clear that there can be solutions to problems that don't constitute progress, let alone *moral* progress. Roth recognizes this when she discusses the "problem" of teen pregnancy and possible solutions to it. Some see teen pregnancy as a moral failing, some as a public health issue, and some as a matter of justice. Now one possible solution to the problem

of teen pregnancy, whatever it consists in, would be to get rid of it altogether by forced sterilization of all minors, male or female or both (just to be absolutely sure to leave no child behind, so to speak). Roth argues that cases like this show

> first, that a way of overcoming a problem counts as a real solution (and hence as progressive) only if it does not create more serious or intractable problems, and second, that what can count as a problem-solution depends importantly upon background values.
>
> (11)

The second point is well taken, for an anti-natalist's different values may welcome our sterilization program as tremendously progressive. But even those who would not advocate the last generation of humans to arrive quite so soon could doubt that progressive developments tend to solve more problems than they create on balance. This point holds in particular once we recognize that not all problems are moral problems, so that even if it could be shown that all instances of moral progress in the long run increase the net amount of solved *moral* problems (and even this could be questioned), it seems hard to believe that all instances of moral progress solve more problems, moral *and non-moral*, than they create.

8.5   Moral Convergence

Realist explanations of moral convergence don't work. But that's not because non-realist explanations of moral convergence work. It's because convergence is not a real phenomenon. There is widespread moral agreement at the metaethically relevant fundamental level, so there is nothing for people to converge on.

Moral realism is the view that there are mind-independent facts about moral norms and values. Perhaps the most popular challenge to moral realism draws on the allegedly widespread phenomenon of moral disagreement. From Herodotus' *Histories* over Montaigne's *Of Cannibals* to early 20th-century anthropologists, the idea that intra- and intercultural evaluative diversity is somehow threatening to the existence of objective moral facts has held a lot of sway. This is known as the *argument from disagreement*.[2]

In what follows, I will argue that neither it is true that widespread disagreement challenges moral realism nor we find any widespread moral disagreement in the real world. What I will claim is that realism about any given domain, including the moral one, predicts widespread *dis*agreement,

and that what we find is widespread *agreement* about basic moral norms and values at the metaethically relevant level.

Why should moral convergence over time support realism? Roughly, the argument seems to go as follows:

The Argument From Convergence

1. *Diversity*: There is an initial state of significant moral disagreement.
2. *Convergence*: Over time, people converge on a shared set of values.
3. *IBE*: The best explanation of convergence involves mind-independent moral facts. Therefore,
4. *Moral realism*: There are mind-independent moral facts.

Extant anti-realist criticisms of this argument focus on *IBE*, and try to show that moral convergence is actually best explained in *non*-realist terms (Hopster, 2020). My claim, on the other hand, is that *Diversity* is false. People share a basic set of moral values, so there isn't any convergence at all.

What would it mean to find a lot or a little disagreement? It would be implausible to suggest that realism predicts *zero* disagreement—that is: no one ever disagrees with anyone else about any moral issue—or that non-realists predict *perfect* disagreement—that is, everyone always disagrees with everyone else about all moral issues. So what do realists and anti-realists predict about the quantity of moral disagreement? Here, I assume that there is *some* threshold of disagreement beyond which both parties to the debate would agree that realism becomes difficult to defend. Whatever that threshold is, that is the one I will rely on as well. I am not suggesting here that there is an agreed-upon quantity of disagreement within this debate. The point is, rather, that those who take sides in the debate surrounding the argument of from convergence must have some ballpark beyond zero and short of total disagreement in mind to engage with the issue at all. This ballpark is what I will assume as well. Of course, I will also argue that the very assumption that diversity is a threat to realism is actually a mistake, but as a characterization of the *de facto* dialectical situation, it seems appropriate.

Most authors who defend moral realism against the alleged threat from disagreement do so by invoking so-called *defusing explanations* that are supposed to show that what seems to be genuine *moral* disagreement actually turns out to be superficial disagreement about non-moral facts or something else, thus rendering the given case of *prima facie* disagreement non-fundamental and thereby unthreatening to the convergentist moral realist. Fewer authors decide to bite the bullet and

accept the existence of widespread moral disagreement, instead challenging the idea that moral realism is incompatible with disagreement in the first place. Divergentists such as Shafer-Landau (2003), for instance, claim that "disagreement poses no threat to realism of any stripe, and so, a fortiori, poses no threat to moral realism in particular" (228). As we will see, I agree with this statement, but aim to take it in an altogether different direction. I will argue that realism isn't threatened by moral disagreement because it actually predicts that moral disagreement will be widespread.

Here's why: whenever there are objective facts about a domain, we expect there to be lots of disagreement between a few experts on those facts and the ignorant masses. Consider the analogy with physics: we would be very surprised to find that pretty much everyone everywhere and at all times essentially agrees on the fundamental laws of the natural world. And rightly so: what we actually find is an enormous amount of disagreement about the basic structure of the cosmos, and that only a tiny group of people even *comprehends* what seem to be the best accounts of the workings of nature. At the very least, it should trouble realists that ethical experts have discovered very little, if anything, that isn't already in some shape or form known to the folk. The important "discoveries" ethicists have made (think: gender equality, the wrongness of slavery, the "innocent until proven guilty" principle) either draw on insights which are perfectly available to lay people or have been made at pretty much the same time that the rest of society has made them as well.[3] This is an oddity in its own right. Prima facie, realism often predicts disagreement.

Some authors come close to endorsing this view. McGrath (2008), for instance, also argues that moral realism does not require convergence. However, she, like most people, uses this compatibility claim as a line of *defense* of realism. She doesn't go one step further, saying that realism predicts disagreement and that we actually find lots of agreement: "It is *obvious* that there is a great deal of actual moral disagreement" (McGrath, 2008, p. 87, my emphasis). I will show that this isn't just not obvious but indeed that the empirical record suggests that there is virtually no fundamental moral disagreement at all.

Who is the main target of my argument? Most obviously, non-naturalist forms of realism seem to qualify. Realists such as Huemer (2008) hold that their intuitionist epistemology is a revisionary one. Competent moral judgment is difficult to do, and has the psychological deck stacked against it. This means that only few people will wind up appreciating the moral facts. These few, then, will disagree with the rest. It would be evidence against this position if it turned out that as far as fundamental moral beliefs are concerned, there is almost no revision, or

that the revisionary aspects of modern moral outlooks are everywhere explained by the discovery of non-moral facts. Naturalists also seem to qualify. Some versions of moral naturalism hold that the acquisition of moral knowledge should be modeled after the discovery of empirical facts (Tropman, 2012, 2014). Moral facts are *discovered*, and if this is so, then it seems plausible to suggest that not everyone discovered them automatically. Other types of realism may qualify as a target of my argument as well, though some may not.

As mentioned before, by far the most popular strategy for realists has been to show that cases of apparent moral disagreement can somehow be explained away in terms of disagreement about the non-moral facts, failures of interpretation, irrationality, special pleading, and so on. The lesson to take away from this first line of resistance is that how to zoom in on the "right kind" of disagreement is indeed a tricky issue (Sauer, 2018). The important thing to file away at this point is that virtually everyone agrees that relevant disagreements must not be "superficial". That is, they must not boil down to disagreements about the supervenience basis of moral facts. Other defusing explanations of purported moral disagreements must be blocked as well. The argument from agreement is compatible with this claim. It is indeed obvious that there is lots of non-fundamental moral disagreement.

What would it take for people to be in moral agreement? First, moral agreements would have to be fundamental, that is, not explainable by superficial agreement on the non-moral supervenience basis of moral truths. Suppose that there are two people who think that there should be no differences in average income between two given groups. However, one person actually thinks that income should be proportional to intelligence, while the other does not. Both people merely happen to agree that there are no significant differences in average intelligence between the members of both groups. This is merely superficial moral agreement.

Second, we would not expect *perfect* agreement: divergence in emphasis, such as the relative importance of promise-breaking or well-being, is allowed.

Third, the metaethically interesting level of disagreement concerns *basic* moral norms and values. People may hate each other's guts because one believes in socialism and the other in capitalism. Metaethically, this disagreement, however passionate, may turn out to be thoroughly irrelevant if both disagreeing parties care about the well-being of the poor, while simply disagreeing about which economic system has the best track record when it comes to realizing this value.

It is crucially important to have a steady grasp of what constitutes fundamental rather than superficial disagreement. Unfortunately, such as grasp is very difficult to come by.[4] Disagreement or agreement, it seems,

can be fundamental in one of two ways. A moral belief can be *epistemically* fundamental in virtue of being non-inferentially justified. On this notion, it seems plausible to suggest that there is little fundamental moral disagreement. However, since my argument is largely based on exploiting disanalogies, it now becomes a problem that the fundamental (so understood) non-moral beliefs do not encounter much disagreement either, because these non-inferentially justified beliefs mostly consist of observations. Or, one could understand fundamental moral beliefs in an explanatory sense, such that the fundamental moral truths are the ones that explain why the less fundamental moral truths hold. On that notion, there clearly is a sufficient amount of non-moral disagreement—the scientific experts have come up with explanations of the phenomena deeply at odds with what lay people believe. The problem, then, is that the same seems to hold for the moral domain. Here, the ethical experts have devised theories regarding what grounds rightness and wrongness which are in conflict with lay people's opinions on these matters. The fundamental moral truths would be adequately captured by esoteric theories such as utilitarianism or deontology.

What level of disagreement is there *at the level of moral theory*? For instance, there seems to be genuine disagreement between consequentialists and Kantians, or Kantians and virtue theorists, or virtue theorists and consequentialists.[5] However, it would be misleading to suggest that these theorists are either (a) insensitive to the moral concerns of the others or (b) endorse fundamentally different deontic outputs. Different moral theorists simply think that one theory provides a more elegant and simple account of whatever our moral duties are (about which they largely agree). Indeed, this point also ricochets, because it should be considered surprising that we only find this narrow space of options of three basic normative theories, and have found essentially only those for millennia. For why, if realism is true, would that be so?

If individuals and cultures actually agree about what morality demands, where does the *appearance* of disagreement—on which proponents as well as critics of the argument of disagreement tend to rely—come from? Here, I think the distinction between fundamental and superficial disagreement starts to play an unexpected role. Traditionally, this distinction has been used by defenders of realism to show that most apparent disagreements can actually be explained away in realism-friendly terms. The idea was that upon closer scrutiny, most candidates for genuine moral disagreement actually turn out to be disagreements about non-moral facts or simply misinterpretations of other cultures (Moody-Adams, 2009) and therefore fail to establish any relativistic conclusions. The suggestion I wish to make is that the distinction between superficial and fundamental

disagreements does not explain away disagreement, but explains the *appearance* of widespread moral disagreement. It is correct that people and cultures *seem* to disagree about morality all the time, but many, if not most, of these disagreements turn out to be superficial and defusable: this means that actually, there is a lot of *fundamental moral agreement* around. And if my argument from the previous section is sound, this spells trouble for realism.

This means that defusing explanations of moral disagreement come back to haunt the moral realist. Originally, these explanations were supposed to show that most cases of moral disagreement actually turn out to be superficial and thus unthreatening to the realist. And this is exactly right: most disagreements *are* superficial and non-fundamental. The examples mentioned earlier—pain is bad, cooperation is good, cheaters ought to be sanctioned, and morality is about harm, fairness, loyalty, respect, purity, and so forth—identify fundamental moral issues, and on those issues, there is overwhelming agreement among virtually all cultures and individuals. But this becomes a problem for moral realists once we realize that moral realism never actually required fundamental moral agreement in the first place. It requires *dis*agreement, so that the aforementioned defusing explanations end up infecting, rather than immunizing, moral realism.

8.6   The Fact of Moral Universalism

There is no principled way of classifying an instance of moral (dis)agreement as fundamental or superficial. In the end, I propose to let the empirical evidence speak largely for itself. Do the following examples for moral values strike us as being due to disagreement about a non-moral supervenience basis, or irrationality in the application of moral values? And if they do not, then by the standards deployed by virtually everyone in the debate, the examples mentioned in the following are sufficiently fundamental to count as metaethically relevant. What is the evidence for moral agreement, so understood?

(1) Basic Evaluative Dispositions

Selective pressures have bequeathed to us a set of hardwired basic evaluative dispositions. Street (2006) lists the following:

(1) The fact that something would promote one's survival is a reason in favor of it. SEP
(2) The fact that something would promote the interests of a family member is a reason to do it

(3) We have greater obligations to help our own children than we do to help complete strangers

(4) The fact that someone has treated one well is a reason to treat that person well in return

(5) The fact that someone is altruistic is a reason to admire, praise, and reward him or her

(6) The fact that someone has done one deliberate harm is a reason to shun that person or seek his or her punishment

(115)

Street frames her list in terms of what people have reason to do, but nothing hangs on this. For my purposes, it is enough to recognize that, for instance, human beings treat the fact that something would promote one's survival as something that favors doing it. Human beings have a basic disposition to value survival. Moreover, people overwhelmingly agree that pain is bad, cheaters should be punished, cooperation rewarded, and so forth. Let me emphasize already that there is widespread agreement between those people who could claim to be moral "experts" of some sort *and* laypeople on such basic evaluative facts. Because of their evolutionary rationale—the people whose ancestors didn't have the aforementioned dispositions simply aren't around anymore—these dispositions are likely to be deeply entrenched and thus widespread.

(2) Cooperative Strategies

Recent evidence suggests that cooperation and the various strategies that generate cooperative gains are universally considered morally good. Oliver Curry (2016), who explicitly sets up his "morality as cooperation theory" as a challenge to descriptive moral relativism, distinguishes seven such strategies:

1. Allocation of resources to kin
2. Coordination to mutual advantage
3. Social exchange
4. Hawkish traits (dominance as conflict resolution)
5. Dove-ish traits (submission as conflict resolution)
6. Division/Fairness
7. Possession/Property

This theory, according to which cooperation is the unifying thread within morality and allows us to distinguish genuinely moral concerns from others, has two main advantages. First, it is based on a more systematic

foundation than competing approaches such as moral foundations theory (Graham et al., 2009), whose list of moral foundations, according to Curry, is compiled in an "*ad hoc*" (10) fashion without much predictive power. Second, it is supported by impressive cross-cultural evidence. Curry and his colleagues looked at the ethnographic record of 60 societies specifically selected to be representative of all of humanity, including samples from all continents and forms of social organization, spanning over several centuries. The material they investigated was drawn from the so-called *Human Relations Area Files*, which they had independently coded for paragraphs pertaining to *Ethics* or *Norms*, yielding around 400 paragraphs from over 500 documents from the aforementioned 60 societies. They operationalized the cooperative strategies 1–7 (e.g., kinship was operationalized in terms of instances where [only] family or in group members were helped) and coded the strategies so operationalized in terms of whether they were considered good or bad (e.g., by being described as ethical, virtuous, and so forth). The strength of their results is hard to overstate: in 99.9% of cases, the aforementioned types of cooperation were considered morally good. Some items on this list, particularly list items 4 and 5, may not sound very moral to some. But the traits Curry has in mind here are close to traditional virtues such as courage and skill (list item 4) and humility and respect (list item 5). Conversely, the theory predicts that uncooperative strategies—"neglecting kin, betraying one's group, free-riding, cowardice, disrespect, unfairness and theft" (11) will be universally regarded as morally bad, which is very plausible. These findings provide very strong evidence that fundamental moral agreement is extremely widespread and robust.

(3) Cross-Cultural Values

Cross-cultural evidence suggests that all people all over the world basically agree on which values there are (Alfano, 2016). The Schwartz' (1994) Value Survey indicates that the following values are universally shared:

Power: Social status and prestige, control or dominance over people and resources

Achievement: Personal success through demonstrating competence according to social standards.

Hedonism: Pleasure and sensuous gratification for oneself.

Stimulation: Excitement, novelty, and challenge in life.

Self-direction: Independent thought and action-choosing, creating, exploring.

Universalism: Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature.

Benevolence: Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.

Tradition: Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide.

Conformity: Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.

Security: Safety, harmony, and stability of society, of relationships, and of self.

(Schwartz, 1994, p. 22)

Here, too, the material was sampled from 25,863 socioeconomically diverse adults from 44 countries from all continents. Respondents rated the values presented to them on a 9-point scale ranging from "of supreme importance" (=7) to "opposed to my values" (=–1). Here, the point is not that these values aren't emphasized differently. They are. But for the issue of whether there is fundamental moral disagreement, such differences in emphasis do not matter. What matters is that the aforementioned values are universally endorsed in all cultures included in the sample.

(4) Political Values

Political psychology suggests that people's moral beliefs can be traced back to a number of moral foundations, a general sensitivity to which is shared (Haidt, 2012). Haidt distinguishes six of them:

1. Care/harm
2. Fairness/cheating
3. Loyalty/betrayal
4. Authority/subversion
5. Sanctity/degradation
6. Liberty/oppression[6]

Virtually everyone is at least somewhat sensitive to the moral relevance of all six foundations, even though different people—and, in particular, different political groups—assign more (or less) importance to different foundations. Liberals famously emphasize the first two foundations at the expense of the others; conservatives have a more balanced appreciation of all six foundations (Sauer, 2015). This, however, doesn't support moral relativism; it merely explains why there are differences between basic political outlooks such as liberalism and conservatism. Whether or not Haidt's Moral Foundations Theory is the correct account of the origin of moral intuitions does not matter for my purposes. What matters here is that *something like* a theory of moral foundations is probably correct, regardless of whether it is Haidt's five (or six) moral foundations,

Rozin's CAD triad/Shweder's "Big Three", Rai and Fiske's (2011) four forms of relationship regulation or indeed Curry's seven pillars of morality as cooperation. In fact, recent evidence suggests that the number of basic moral foundations is even smaller (Schein & Gray, 2015). People across the political spectrum seem to agree that *harm* is actually the core of morality. If this is so, then the amount of evaluative agreement we find is even higher.

An interesting final data point for the idea that there isn't a lot of moral disagreement is this: consider the people with the most radically divergent moral views you can think of. Now consider how surprisingly non-radical these disagreements are. In particular, compare how radical they *could* be, and how radical disagreements can in fact become when we look at other domains about which we are confident realists. Moral disagreements often have the following flavor: the Maragoli think fairness requires that one give 25% in the Ultimatum Game, whereas US citizens believe that one ought to give around 45% (Henrich et al., 2010). Does this count as disagreement? Perhaps. But notice that in principle, it would not be inconceivable for one culture to think that harming others is wrong and that cooperation should be rewarded, for another culture to think that cooperation is wrong and that harming others is good, and for another culture to think that counting blades of grass is the best way to spend one's time or to jump up and down and scream at purple things (Street, 2006). But in reality, that is, in lived societies populated by people of flesh-and-blood, we *never find any disagreements anywhere near this strong*.

Now consider the radically divergent views people actually hold about realistic domains such as scientific truths, for instance about the origin of the cosmos. Some believe that it was created in six days by an omnipotent immaterial consciousness, some believe it had no origin but that the cosmos is a cyclical spiral, some hold that it was given birth to by a giant turtle, some people believe that it came into being around 13 billion years ago in a high-density even known as the Big Bang. There is almost no overlap between these accounts of the genealogy of the universe. The range of moral beliefs actually held by people both synchronically and diachronically is nowhere near this wide.

On the other hand, the empirically best-supported explanations given for patterns of disagreement almost always rely on differences in economic circumstances that explain differences in fairness intuitions or sharing dispositions. Even the most outlandish beliefs and practices, such as wife selling in 19th-century England or medieval trial by ordeal ultimately do not turn out to be based on fundamental evaluative divergence, but on limited information and rational incentives (Leeson, 2017).

8.7 Realism Requires Disagreement

One may think that it doesn't matter when we find moral agreement. Do we find an initial state of agreement, or a final state of agreement after a period of convergence over time? Either way, it seems, realism is vindicated.

I disagree. Realism about a given domain predicts considerable disagreement within that domain rather than agreement, consensus or convergence. Indeed, *agreement* should be considered suspicious, especially when there is a non-truth-tracking explanation for why there would be agreement even if there weren't any mind-independent facts to bring it about.

The alleged connection between agreement within a domain and realism about that domain is explicitly established by many advocates of moral realism. Here is a reconstruction by David Enoch (2011):

(1) In many discourses, there is wide-ranging agreement about the truths central to the relevant discourse.
(2) What best explains such wide-ranging agreement is that there are objective truths the discourse answers to, truths on which opinions gradually converge.
(3) Therefore, there are objective truths the relevant discourse answers to. (From (1) and (2), by inference to the best explanation.)

(196)

The best explanation, Enoch suggests, for wide-ranging agreement about a given domain is that this domain contains objective facts.

According to my argument, Enoch's premise (2) is false. What best explains agreement is precisely the *lack* of objective truths, because when there are such truths, *dis*agreement is to be expected. In general, the reason for this is that most of the interesting truths within a domain will be unobvious. Discovering them requires often painstaking inquiry and methodical reasoning, frequently conducted by professional investigators. Clearly, many philosophers think that this is precisely what isn't required for obtaining moral knowledge. I will argue in the following that this assumption is unwarranted. Given that most truths within a given realist domain should not be readily accessible, widespread agreement about the truths of that domain should make us suspicious.

The claim that realism doesn't predict agreement becomes especially plausible when there are other, non-truth tracking explanations for why people would come to converge on a set of domain-specific beliefs even if there were no mind-independent facts those beliefs answer to. For instance, if we find that the members of a community agree on certain beliefs because of the pernicious influence of some charismatic guru, the envisioned inference to the best explanation is blocked. In the case of

morality, such a non-truth-tracking explanation could, for instance, be supplied by selective pressures of biological or cultural evolution, more on which is discussed later.

The main reason for thinking that realism about a given domain predicts disagreement is that when we look at domains about which most people are confident realists—such as science—we find lots of disagreement. By this, I do not mean to suggest that scientific experts do not converge on the basic tenets of their respective disciplines. What I mean is that when comparing scientific experts and laypeople outside of a given discipline, we find that almost no one, globally speaking, believes the basic truths of science, be it physics, biology, economics, or cognitive science. Almost all of the foundational insights of these disciplines are only believed by an extremely tiny minority of experts or people who have received some formal training in them. All others believe these claims to be untrue or are, at the very least, thoroughly unaware of what these claims are in the first place.

Most scientific truths are deeply counterintuitive, and fairly recalcitrantly so. Counterintuitive claims are, by their very nature, unlikely to be believed by many people who haven't received some sort of training (or indeed indoctrination, as with the counterintuitive teachings of many cults and religions). Consider physics: there is nothing intuitive about the idea of inertia, the relativity of simultaneity, or the mysteries of quantum mechanics. Folk physics, on the other hand, is intuitively compelling but gets it all wrong (McCloskey et al., 1983). Consider biology: even today, the idea that natural selection (and other evolutionary pressures) instead of the vastly more viscerally appealing ideas of intelligent design or Lamarckianism remains deeply counterintuitive. Finally, consider economics: economists routinely complain about the fact that the public as well as elected officials fail to grasp the basic workings of the price mechanism, comparative advantage or the nature of public goods. That is because prices and global trade are strange, and difficult to comprehend (Kahneman et al., 1986).

Widespread agreement across the board is likely due to the fact that there are no counterintuitive truths in a domain for experts to discover over time, and for non-experts to remain unaware of. This is to say that realism about that domain is probably false.

There are several possible responses to my earlier claim that realism requires disagreement. For one thing, realists can reject *invidious comparisons* like the analogy between scientific and moral knowledge, so that the disagreement we find with regard to the former ceases to matter for the lack of disagreement we find in the latter. For another, realists can invite *comforting comparisons*, by seeking an analogy between everyday non-scientific knowledge and moral knowledge, such that the agreement we find in the former, combined with our inclination to construe this domain

realistically, supports realism about the latter. Third, realists could argue that the *a priori* character of moral knowledge explains how widespread agreement can be compatible with realism. Fourth, realists could object that the argument from disagreement is committed to the purportedly implausible idea of moral expertise. Fifth, realists can launch a counterattack, arguing that the pattern of moral agreement and disagreement we find is exactly what we would expect of moral realism were true. I will take up these options in turn.

Scientific and Moral Knowledge

One may be inclined to argue that the analogy with science is flawed, as it stacks the deck against those who would want to defend realism from the argument from agreement. But notice that the analogy with science is one that is frequently sought by realists themselves (Huemer, 2005). It then seems dialectically unfair to exploit this analogy when it seems to help realism, but reject it when it turns out not to.[7]

However, there *is* a legitimate point here, which is that what should be compared to what to make a relevant analogy (or disanalogy) work is a difficult question. One could say that even physicists and laypeople agree about a lot, namely observational reports (e.g., the red light is blinking, this feather falls at a certain speed in a vacuum under the current experimental conditions). The experts will merely be better able to explain, in more fundamental terms (such as the equivalence of inert and heavy mass), why these observations occur. Considering this analogy, should beliefs fundamental values be compared to observational reports—about which the experts and laypeople purportedly agree—or fundamental explanatory principles—about which they do not? The second option seems preferable because it compares like to like. And if we compare fundamental and unifying principles to each other, we find lots of disagreement between experts and laypeople in the sciences but hardly any disagreement between ethical experts and ethical laypeople.

This point brings us back to the earlier discussion of how to draw the distinction between fundamental and superficial (dis)agreements. Earlier, I suggested sticking to a criterion that disqualifies moral disagreements that can be defused as mere disagreement about a non-moral supervenience base. But if that is the notion of fundamentality assumed here, then the analogy with disagreement in physics seems to break down, because there is no supervenience base for physical properties besides those same physical properties. However, the analogy between ethics and physics (or other domains for that matter) remains instructive, because in both cases, we scrutinize the level of fundamental explanatory truths (the basic laws of nature/the basic moral values) for patterns of agreement and disagreement.

In the case of ethics, we arrive at that level by subtracting non-moral and other defusable sources of disagreement; in the case of physics, we do so by looking at the "Quinean" core of theory construction.

To be sure, ethical experts such as moral philosophers defend lots of strange and counterintuitive views—such as the morality of open borders, or infanticide, or human enhancement—but they do not defend them because they endorse different values. They merely have a better, or at least different, appreciation of what these values entail and how to implement them.

Everyday Knowledge and Moral Knowledge

There is another analogy, namely one with knowledge of everyday truths, which purports to show that realism doesn't predict disagreement. If realism about everyday midsized objects is true, we would expect a lot of disagreement about them. But we don't see a lot of disagreement about everyday midsized objects. Therefore, realism about nearby midsized objects is false. Clearly, the first premise is false. At the very least, this parody shows that one needs additional premises to get from realism to expected disagreement which might not be plausible in the moral case.[8] Again, however, most contemporary moral realists—remember the targets of my argument narrowed down earlier—do not think that moral facts are like facts about ordinary midsize objects. They believe that acquiring moral knowledge requires empirical or rational investigation of the kind which is not necessary to gain access to mundane facts about apples and oranges.

Easy Moral Knowledge

One could also hold that moral truths are *a priori*. Michael Smith, for instance, writes:

> It is agreed on nearly all sides that moral knowledge is relatively a priori, at least in the following sense: if you equip people with a full description of the circumstances in which someone acts, then they can figure out whether the person acted rightly or wrongly just by thinking about the case at hand.
>
> (2000, p. 203)

Many others hold similar views: "Morality does not require beliefs that are not known to all moral agents" (Gert, 2004, p. 90). Or: "this book . . . contains no new information about what kinds of actions morality prohibits, requires, discourages, encourages, or allows. Anyone who is intelligent enough to read this book already has all of this information" (3).

Or: "[i]t seems implausible to say that it would take a 'moral genius' to see through the wrongness of chattel slavery" (Guerrero, 2007, p. 71). Typically, there isn't much disagreement on conceptual trivialities.

However, most realists continue to claim that moral truths are unobvious, so at the very least, this reply to my argument does not work for those theories. Moreover, even though I have argued that we do find widespread moral agreement, there is at least *some* moral disagreement. But if the "easy knowledge" reply is correct, then we would not expect even this much disagreement. There is no disagreement on whether bachelors are married or unmarried. The reply backfires, so realists should seek no comfort in it.

Moral Expertise

My argument so far seems to rely, at least implicitly, on the claim that there could be moral expertise, which is controversial.[9] It may turn out that all the heavy lifting in the argument from disagreement is actually being done by the claim that there can be no one who fundamentally knows better about morality than other people. So it's never been about disagreement in the first place, and always about the idea that no one's fundamental moral views are superior. This would surprise no one who has ever taught relativism to undergraduates, who seem to take offense with non-relativism for precisely this reason: that it seems disrespectful or intolerant to describe someone else's moral views as mistaken.

Counterattack

Some realists may claim that we find a certain pattern—a mix, as it were—of agreement and disagreement, and that the pattern we find can comfortably be accounted for by realism. Consider Brink (1984). Brink may argue first that there's more agreement than Mackie and others allow, and that the rest can be explained by appealing to distorting factors via the usual defusing explanations. Some moral truths are easy and others are tough to know, just like some truths about the natural world are easily accessible and others aren't. This predicts a pattern of agreement and disagreement, which the realist can maintain is what we observe.[10] However, this line of reply overlooks the fact that what we find is strong agreement about the *basic principles* of morality. If the analogy with science works, then we should not expect this (notice the huge disagreement between physicists and the folk on the fundamental laws of nature). So the pattern we in fact observe remains problematic for realism.

Here are three final points. Disagreements about how to draw the circle of moral concern seem to be of fundamental importance. But are they? Is it fundamental disagreement in the metaethically interesting sense that some people think that certain groups (Jewish people, people of color, and other minorities) do not deserve full moral consideration? I suggest that it is not, because these aren't cases of fundamental disagreements about *basic values* (i.e., some people thinking suffering is bad and others don't), but merely about membership. Criteria of membership are among the *normatively* most urgent cases of disagreement. Metaethically, they are less interesting.

Finally, it is likely that almost all the convergence we actually observe isn't domain-specific. It occurs because people's circumstances come to resemble each other, or because people converge on the non-moral facts, or for other reasons. Look at what explains the developments realists like to explain in terms of domain-specific moral convergence, that is, developments towards liberalization, or increasing resistance to violence and discrimination (see Pinker, 2011; Huemer, 2016a). In almost all of these cases, these increases in agreement are not due to a better appreciation of mind-independent moral facts, but other, non-rational (though perhaps very desirable) forces such as upheavals in social organization.

Notes

1. This chapter draws on material previously published in: Sauer, H. (2019a). The argument from agreement: How universal values undermine moral realism. *Ratio*, *32*(4), 339–352.
2. An excellent overview of the most promising versions of the argument from disagreement can be found in Enoch (2009). Useful starting points for the modern debate are Harman (1975) and Brink (1984). Realist perspectives on disagreement can be found in Huemer (2005) and Shafer-Landau (2003). For the recent empirical revival of the argument from disagreement, see Wong (2006), Prinz (2007), and Doris and Plakias (2008; cf. Fitzpatrick, 2014; Meyers, 2013; Fraser & Hauser, 2010; Sneddon, 2009; Leiter, 2008).
3. Thanks to Michael Huemer for helpful comments on this issue.
4. The following discussion draws heavily on comments I received from Michael Huemer.
5. Thanks to Nicholas Laskowski for helping me phrase this point.
6. Haidt's *The Righteous Mind* still operates within his earlier five moral foundations framework. The liberty/oppression foundation was added more recently, see Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012).
7. In a recent paper, Finur Dellsén (2017) shows that whether or not we should expect agreement or disagreement on a given issue, and what the epistemic implications of it are, is frequently a counterintuitive question. Dellsen argues that even disagreement *between the* experts, and not just between experts and laypeople, supports realism.
8. Thanks to AK for discussion on this point.
9. Thanks to JW.
10. Thanks to AK.

Conclusion

Social change is biased in favor of moral improvement. This has been the main thesis of this book.

The arguments developed in the previous pages are based on a curious mix of optimism and pessimism. They are optimistic, in that I wanted to make the case that we are on a trajectory of moral improvement that is likely to continue. Our norms and values have improved—not everywhere equally, not for everyone equally, and not as fast as one could have hoped—and there is ample potential for further improvement.

There is, however, a serious dose of pessimism in this line of thought as well, because if I am right, and the way in which progress happens is via a path of slow cultural evolution over the course of which various mechanisms of social change amalgamize to advance the moral frontier by storing progressive gains in sophisticated social institutions, then moral progress is also hard to accomplish, morally acceptable societies are difficult to build, and progressive gains can never be taken for granted.

Another way of putting my central contention is that emanicaptive dynamics are autocatalytic: they generate the fuel they run on. As societies grow and switch to more adaptive structures of social cooperation, they are pushed towards further liberalization of their norms and values, more inclusive criteria of full membership, more well-being, and more freedom, which in turn increases their potential for technological innovation, economic prosperity, and political attractiveness. Via a chain of virtuous feedback loops, these developments unleash further progressive changes, and so on.

Why aim for progress, so understood? One obvious answer is that, other things being equal, we want things to become better rather than worse. A related but deeper concern has to do with the underappreciated possibility that there may be serious, large-scale moral wrongs happening all the time, without us knowing about them at all. Only progressive developments, over the course of which social practices and perceptions

radically shift, can uncover them. Williams (2015) calls these "ongoing moral catastrophes":

> Literally hundreds of generations have thought that they had the right moral values. Two thousand years ago, the Romans—the imperialistic, crucifying, slave-owning Romans—were congratulating themselves on being civilized, because unlike the barbarians they had abolished human sacrifice. This was genuine progress, but what they did not realize was that thousands of years' additional progress remained to be made. We are in the same position: we know how much progress is embodied in our values, but not how much progress remains to be made in the future. This, then, is the . . . [w]: most cultures have turned out to have major blind spots in their moral beliefs, and we are in much the same epistemic situation as they are, so we will probably also turn out to have major moral blind spots.
>
> (Williams, 2015, p. 3)

If abortion really is murder, or if meat consumption really is gravely wrong, or if large-scale incarceration really is socially debilitating, then we need to rectify such wrongs. But these are candidates for "moral catastrophes" that we already consider. The best way to avert—and this is the important qualification—thus far *unknown* moral catastrophes is to encourage and foster further moral progress. Moral progress, after all, is the process by which we become aware, and get rid of, our most heinous ways.

In a recent article in *The Atlantic*, Patrick Collison and Tyler Cowen call for the establishment of a new field of "progress studies".[1] They argue that social and moral progress—by which they mean "the combination of economic, technological, scientific, cultural, and organizational advancement that has transformed our lives and raised standards of living over the past couple of centuries" doesn't take care of itself. Moreover, we know that some places, or social conditions, or institutional arrangements, are orders of magnitude better at generating significant ameliorative gains than others. Germany dominated scientific inquiry at the turn of the previous century; Silicon Valley now dominates the evolution of digital technology. But why? Progress is needed not only for uncovering, and stamping out, egregious unknown wrongs, but also for discovering avenues for large potential but thus far unrealized gains.

David Estlund invites us to play the "unbelievable moral progress game" (Estlund, 2019). He wonders: "What unbelievable moral achievements might humanity witness a century or two from now on?" (260). The question is designed to be unfair, because if you think you have an answer, you probably just lack imagination, or you're not thinking radically enough. Perhaps, a century from now, everyone will be a vegetarian and we will

look back with shame and horror and astonishment on the carnivorous ways of our past; perhaps we will have moved beyond a society where the color of our skin, or the region of the world we were born in, will in any way be correlated with how our lives go. But are these instances of moral progress "unbelievable" enough? Or do we need to go further? Will we one day legalize all drugs? Allow children to vote? Can we imagine a world without scarcity, without work? Can we ever achieve, and be satisfied with, real equality, and eliminate our desire for status and competition? Should we aim to create a world without suffering? And how? By painlessly killing wild predators? By painlessly euthanizing ourselves, phasing out our own existence until there is no more frustration, no more fear, no more pain?

I take no stance here on which of the aforementioned examples are important, and what the morally desirable trajectory of social change will be. Time, I suppose, will tell. Moreover, it remains unclear how such forms of progress will be implemented. Maybe direct interventions via genetic engineering or biochemical enhancement are the way forward?

The problem is that virtually all such interventions are either weak or morally ambiguous or both (Anomaly, 2020). But how will such new forms of progress be facilitated and implemented instead? The answer to this question, I speculate, also hints at some reasons for why the moral progress game is so hard to play. If the arguments in this book are correct, then the reason why it is so hard to imagine radically new forms of progressive change is not that we cannot anticipate how our values will develop—because at the end of the day, they hardly do. The main problem is that moral progress, and the direction it takes, largely depends on what novel institutions evolve and manage to stick around that will allow intelligent, norm-sensitive cultural learners—us—to solve the problems of today and tackle the problems of tomorrow.

But what are the institutions of the future? It is very difficult, if not impossible, to predict what they will look like. Some people think about alternative institutional arrangements, and they almost always strain credulity. Should criminals be put in charge of law enforcement? Should doctors pay when their patients die?[2] Should people with kinks be allowed to bring their BDSM masters to work?[3]

If the cultural evolution account of moral progress developed here is correct, it should be clear that figuring out future institutions ahead of time would be difficult (Feldmann, 2004). We can't anticipate the evolution of social practices. Nevertheless, a possible idea for a recipe for creating future institutions is this: identify a thus far untapped force of human nature (e.g., aggressiveness) and think about how this force could be harnessed for the common good, like markets harness selfishness or marriage harnesses jealousy. For instance, the rise of deep fakes and fake news create a social environment that makes new epistemological institutions necessary. This

has happened before: the ability to deceive created a niche for improved theory of mind in early humans. Democratic institutions with competing parties created the need for journalism and reporting. Likewise, deep fakes and fake news create a phenomenon that could lead to new institutions of fact-checking and epistemic vigilance.

This is merely one possible example. The powers of reason are too limited to figure out in advance how this story will unfold. Or perhaps the way it unfolds will tell us what the powers of reason are in the first place? If that's true, then the unfolding of this story will itself be the unfolding of reason.

Notes

1. www.theatlantic.com/science/archive/2019/07/we-need-new-science-progress/594946/
2. http://mason.gmu.edu/~rhanson/altinst.html
3. https://juliagalef.com/2017/08/23/unpopular-ideas-about-social-norms/

Bibliography

Acemoglu, D., & Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Books.

Alexander, M. (2020). *The New Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

Alfano, M. (2013). *Character as moral fiction*. Cambridge University Press.

Alfano, M. (2016). *Moral psychology: An introduction*. Polity.

Anderson, E. (2014). Social movements, experiments in living, and moral progress: Case studies from Britain's abolition of slavery. https://kuscholarworks.ku.edu/bitstream/handle/1808/14787/Anderson_Social_Movements.pdf

Anderson, E. (2015, November). Moral bias and corrective practices: A pragmatist perspective. *Proceedings and Addresses of the APA*, *89*, 21–47.

Anomaly, J. (2017). Trust, trade, and moral progress: How market exchange promotes trustworthiness. *Social Philosophy and Policy*, *34*(2), 89–107.

Anomaly, J. (2020). *Creating future people: The ethics of genetic enhancement*. Routledge.

Appiah, K. A. (2011). *The honor code: How moral revolutions happen*. WW Norton & Company.

Arruda, C. T. (2017). The varieties of moral improvement, or why Metaethical constructivism must explain moral Progress. *Ethical Theory and Moral Practice*, *20*(1), 17–38.

Baker, R. (2019). *The structure of moral revolutions: Studies of changes in the morality of abortion, death, and the bioethics revolution*. MIT Press.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209.

Barber, A. (2013). Science's immunity to moral refutation. *Australasian Journal of Philosophy*, *91*(4), 633–653.

Baxter, A. J., Scott, K. M., Ferrari, A. J., Norman, R. E., Vos, T., & Whiteford, H. A. (2014). Challenging the myth of an "epidemic" of common mental disorders: Trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depression and Anxiety*, *31*(6), 506–516.

Beckert, S. (2014). *Empire of cotton: A global history*. Vintage Books.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs*, *37*(4), 293–329.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Birch, J. (2017). *The philosophy of social evolution*. Oxford University Press.

Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.

Boehm, Ch. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.

Boettke, P. J., & Candela, R. (2017). The liberty of progress: Increasing returns, institutions, and entrepreneurship. *Social Philosophy and Policy*, *34*(2), 136–163.

Bogardus, T. (2016). Only all naturalists should worry about only one evolutionary debunking argument. *Ethics*, *126*(3), 636–661.

Bostrom, N., & Ord, T. (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics*, *116*(4), 656–679.

Boudry, M. (2019). *Waarom de wereld niet naar de knoppen gaat*. Polis.

Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press.

Bregman, R. (2020). *Humankind. A hopeful history*. Bloomsbury.

Brennan, J. F. (2014). *Why not capitalism?*. Routledge.

Brennan, J. F. (2016). *Against democracy*. Princeton University Press.

Brennan, J. F. (2020). *Why it's ok to want to be rich*. Routledge.

Brennan, J. F., & Jaworski, P. (2015). *Markets without limits: Moral virtues and commercial interests*. Routledge.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good science. In M. H. Appley (Ed.), *Adaptation level theory: A symposium* (pp. 287–302). Academic Press.

Brink, D. O. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy*, *62*(2), 111–125.

Buchanan, A. (2013). The ethics of revolution and its implications for the ethics of intervention. *Philosophy & Public Affairs*, *41*(4), 291–323.

Buchanan, A., & Powell, R. (2017). De-moralization as emancipation: Liberty, progress, and the evolution of invalid moral norms. *Social Philosophy and Policy*, *34*(2), 108–135.

Buchanan, A., & Powell, R. (2018). *The evolution of moral progress: A biocultural theory*. Oxford University Press.

Buddeberg, E. (2019). Is there moral progress? *Analyse & Kritik*, *41*(2), 195–204.

Campbell, R. (2017). Learning from moral inconsistency. *Cognition*, *167*, 46–57.

Campbell, R., & Kumar, V. (2012). Moral reasoning on the ground. *Ethics*, *122*(2), 273–312.

Camus, A. (1991 [1951]). *The rebel. An essay on man in revolt*. Vintage International.

Caplan, B. (2011). *The myth of the rational voter: Why democracies choose bad policies*. Princeton University Press.

Carr, D. (2017). Historical teleology: The grand illusion. *History and Theory*, *56*(2), 307–317.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121–138.

Carter, I. (2011). Respect and the basis of equality. *Ethics*, *121*(3), 538–571.

Chambers, C. (2017). *Against marriage: An egalitarian defence of the marriage-free state*. Oxford University Press.

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222–1226.

Christakis, N. (2019). *Blueprint. The evolutionary origins of a good society*. Little, Brown Spark.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Clark, G. (2007). *A farewell to alms. A brief economic history of the world*. Princeton University Press.

Cofnas, N. (2019). A debunking explanation for moral progress. *Philosophical Studies*, 1–21.

Cohen, D., & Nisbett, R. E. (1996). *Culture of honor: The psychology of violence in the South*. Westview Press.

Cohen, G. A. (2009). *Why not socialism?*. Princeton University Press.

Cohen, J. (1997). The arc of the moral universe. *Philosophy & Public Affairs*, *26*(2), 91–134.

Cohen, M. A., & Zenko, M. (2019). *Clear and present safety. The world has never been better and why that matters to Americans*. Yale University Press.

Cole, M. A., Rayner, A. J., & Bates, J. M. (1997). The environmental Kuznets curve: An empirical analysis. *Environment and Development Economics*, 401–416.

Colombo, M., Bucher, L., & Inbar, Y. (2016). Explanatory judgment, moral offense and value-free science. *Review of Philosophy and Psychology*, *7*(4), 743–763.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Cowen, T. (2011). *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better*. Penguin.

Cowen, T. (2018). *Stubborn attachments. A vision of a society of free, prosperous, and responsible individuals*. Stripe Press.

Crimston, D., Hornsey, M. J., Bain, P. G., & Bastian, B. (2018). Toward a psychology of moral expansiveness. *Current Directions in Psychological Science*, *27*(1), 14–19.

Cuneo, T., & Shafer-Landau, R. (2014). The moral fixed points: New directions for moral nonnaturalism. *Philosophical Studies*, *171*(3), 399–443.

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In *The evolution of morality* (pp. 27–51). Springer International Publishing.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292.

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, *43*, E28. doi:10.1017/S0140525X19001730.

Cushman, F., Kumar, V., & Railton, P. (2017). Moral learning: Psychological and philosophical perspectives. *Cognition*, *167*, 1–10.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.

Deaton, A. (2013). *The great escape: Health, wealth, and the origins of inequality*. Princeton University Press.

de Lazari-Radek, K., & Singer, P. (2012). The objectivity of ethics and the unity of practical reason. *Ethics*, *123*(1), 9–31.

Dellsén, F. (2017). When expert disagreement supports the consensus. *Australasian Journal of Philosophy*, 1–15.

Deneen, P. J. (2018). *Why liberalism failed*. Yale University Press.

de V. Cavalcanti, T. V., & Tavares, J. (2008). Assessing the "engines of liberation": Home appliances and female labor force participation. *The Review of Economics and Statistics*, *90*(1), 81–88.

Diamond, J. M. (1987). The worst mistake in human history. *Discover*, *8*(5), 64–66.

Diamond, J. M. (1998). *Guns, germs and steel: A short history of everybody for the last 13,000 years*. Random House.

Diamond, J. M. (2005). *Collapse: How societies choose to fail or succeed*. Penguin.

Dixon, B. (2005). Achieving moral progress despite moral regress. *Social Philosophy Today*, *21*, 157–172.

Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A political theory of animal rights*. Oxford University Press.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.

Doris, J. M. (2009). Genealogy and evidence: Prinz on the history of morals. *Analysis*, *69*(4), 704–713.

Doris, J. M., & Murphy, D. (2007). From My Lai to Abu Ghraib: The moral psychology of atrocity. *Midwest Studies in Philosophy*, *31*, 25–55.

Doris, J. M., & Plakias, A. (2008). How to argue about disagreement: Evaluative diversity and moral realism. In W. Sinnott-Armstrong (Ed.), *Moral psychology volume 2, the cognitive science of morality: Intuition and diversity* (pp. 303–331). MIT Press.

Dornes, M. (2016). *Macht der Kapitalismus Depressiv? Über seelische Gesundheit und Krankheit in modernen Gesellschaften*. Fischer.

Douthat, R. (2020). *The decadent society: How we became the victims of our own success*. Avid Reader Press.

Duffy, B. (2018). *The perils of perception. Why we're wrong about nearly everything*. Atlantic Books.

Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493.

Easterbrook, G. (2018). *It's better than it looks. Reasons for optimism in an age of fear*. PublicAffairs.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth* (pp. 89–125). Academic Press.

Egonsson, D. (2013). Moral progress. *International Encyclopedia of Ethics*, 1–9.

Enoch, D. (2009). How is moral disagreement a problem for realism? *The Journal of Ethics*, *13*(1), 15–50.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Enoch, D. (2014). A defense of moral deference. *The Journal of Philosophy*, *111*(5), 229–258.

Estlund, D. (2019). *Utopophobia: On the limits (if any) of political philosophy*. Princeton University Press.

Evans, J. S. B. (2017). A working definition of moral progress. *Ethical Theory and Moral Practice*, *20*(1), 75–92.

Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond*. Oxford University Press.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of Personality and Social Psychology*, *117*(1), 50.

Feldmann, H. (2004). Political implications of cultural evolution. *International Journal of Social Economics*, *31*(11/12), 1089–1108. doi:10.1108/03068290410561195.

Fitzpatrick, S. (2014). Moral realism, moral disagreement, and moral psychology. *Philosophical Papers*, *43*(2), 161–190.

FitzPatrick, W. J. (2015). Debunking evolutionary debunking of ethical realism. *Philosophical Studies*, *172*(4), 883–904.

FitzPatrick, W. J. (2019). Moral progress for evolved rational creatures. *Analyse & Kritik*, *41*(2), 217–238.

Fleischacker, S. (2009). *A short history of distributive justice*. Harvard University Press.

Frank, L. E. (2020). What do we have to lose? Offloading through moral technologies: Moral struggle and progress. *Science and Engineering Ethics*, *26*(1), 369–385.

Frankfurt, H. (1987). Equality as a moral ideal. *Ethics*, *98*(1), 21–43.

Fraser, B., & Hauser, M. (2010). The argument from disagreement and the role of cross-cultural empirical data. *Mind and Language*, *25*(5), 541–560.

Freyenhagen, F. (2013). *Adorno's practical philosophy: Living less wrongly*. Cambridge University Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Fricker, M. (2013). Styles of moral relativism. A critical family tree. In R. Crisp (Ed.), *The Oxford handbook of the history of ethics* (pp. 793–818). Oxford University Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708.

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, *25*, 4–12.

Gert, B. (2004). *Common morality: Deciding what to do*. Oxford University Press.

Gert, J. (2015). Disgust, moral disgust, and morality. *Journal of Moral Philosophy*, *12*(1), 33–54.

Gill, M. B. (2007). Moral rationalism vs. moral sentimentalism: Is morality more like math or beauty? *Philosophy Compass*, *2*(1), 16–30.

Gjesdal, A. (2018). Moral learning, rationality, and the unreliability of affect. *Australasian Journal of Philosophy*, *96*(3), 460–473.

Glover, J. (1999). *Humanity. A moral history of the 20th century*. Yale University Press.

Godlovitch, S. (1998). Morally we roll along: (Optimistic reflections) on moral progress. *Journal of Applied Philosophy*, *15*(3), 271–286.

Goldin, I. (2018). The limitations of Steven Pinker's optimism. *Nature*, *554*(7693).

Graeber, D. (2018). *Bullshit jobs. A theory*. Simon & Schuster.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029.

Graham, J., Waytz, A., Meindl, P., Iyer, R., & Young, L. (2017). Centripetal and centrifugal forces in the moral circle: Competing constraints on moral learning. *Cognition*, *167*, 58–65.

Gray, J. (2004). *Heresies. Progress and other illusions*. Granta.

Greene, J. D. (2008). The secret joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology. Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development*. MIT Press.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Atlantic Books.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (Neuro) science matters for ethics. *Ethics*, *124*(4), 695–726.

Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*, *167*, 66–77.

Greenspan, P. (2015). Confabulating the truth: In defense of "defensive" moral reasoning. *The Journal of Ethics*, *19*(2), 105–123.

Greenwood, J., Seshadri, A., & Yorukoglu, M. (2005). Engines of liberation. *The Review of Economic Studies*, *72*(1), 109–133.

Guerrero, A. A. (2007). Don't know, don't kill: Moral ignorance, culpability, and caution. *Philosophical Studies*, *136*(1), 59–97.

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press.

Haidt, J. (2012). *The righteous mind. Why good people are divided by religion and politics*. Penguin.

Haig, D. (2020). *From Darwin to Derrida. Selfish genes, social selves, and the meanings of life*. MIT Press.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE*, *7*(9), e45457.

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology*, *68*, 155–186.

Harman, G. (1975). Moral relativism defended. *The Philosophical Review*, *84*(2), 3–22.

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*(1), 1–17.

Haslanger, S. (2017). Culture and critique. *Aristotelian Society Supplementary*, *91*(1), 149–173.

Haybron, D. M. (2007). Do we know how happy we are? On some limits of affective introspection and recall. *Nous*, *41*(3), 394–428.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, *35*(4), 519–530.

Heath, J. (2009). Three evolutionary precursors to morality. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, *48*(4), 717–752.

Heath, J. (2014a). *Morality, competition, and the firm: The market failures approach to business ethics*. Oxford University Press.

Heath, J. (2014b). *Enlightenment 2.0*. HarperCollins Canada.

Heath, J. (2014c). Rebooting discourse ethics. *Philosophy & Social Criticism*, *40*(9), 829–866.

Heath, J. (forthcoming a). *On the scalability of cooperative structures*. Remarks on GA Cohen, Why not socialism. Unpublished Manuscript.

Heath, J. (forthcoming b). In defense of stigmatization.

Heath, J., & Anderson, J. (2010). Procrastination and the extended will. In C. Andreou & M. D. White (Eds.), *The thief of time*. Oxford University Press.

Hegel, G. W. F. (1975 [1830]). *Lectures on the philosophy of world history. Introduction: Reason in history*. Cambridge University Press.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Henrich, J. (2020). *The weirdest people in the world. How the west became psychologically peculiar and particularly prosperous*. Allen Lane.

Henrich, J., Boyd, R., & Richerson, P. J. (2012). The puzzle of monogamous marriage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 657–669.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Herman, A. (1997). *The idea of decline in Western history*. The Free Press.

Hermann, J. (2017). Possibilities of moral progress in the face of evolution. *Ethical Theory and Moral Practice*, *20*(1), 39–54.

Hermann, J. (2019). The dynamics of moral progress. *Ratio*, *32*(4), 300–311.

Hickel, J. (2016). The true extent of global poverty and hunger: Questioning the good news narrative of the Millennium Development Goals. *Third World Quarterly*, *37*(5), 749–767.

Honneth, A. (1996). *The struggle for recognition: The moral grammar of social conflicts*. MIT Press.

Hopster, J. (2020). Explaining historical moral convergence: The empirical case against realist intuitionism. *Philosophical Studies*, *177*(5), 1255–1273.

Horkheimer, M., & Adorno, Th. W. (1998 [1944]). *Dialektik der Aufklärung. Philosophische Fragmente*. Fischer.

Horne, Z., Powell, D., & Hummel, J. (2015). A single counterexample leads to moral belief revision. *Cognitive science*, *39*(8), 1950–1964.

Huemer, M. (2005). *Ethical intuitionism*. Palgrave Macmillan.

Huemer, M. (2008). Revisionary intuitionism. *Social Philosophy and Policy*, *25*(1), 368–392.

Huemer, M. (2013). *The problem of political authority*. Palgrave MacMillan.

Huemer, M. (2016a). A liberal realist answer to debunking skeptics: The empirical case for realism. *Philosophical Studies*, *173*(7), 1983–2010.

Huemer, M. (2016b). Confessions of a utopophobe. *Social Philosophy & Policy*, *33*(1–2), 214.

Huemer, M. (2019). *Dialogues on ethical vegetarianism*. Routledge.

Husi, S. (2017). Why we (almost certainly) are not moral equals. *The Journal of Ethics*, *21*(4), 375–401.

Hutchison, A. (2014). The Whanganui river as a legal person. *Alternative Law Journal*, *39*(3), 179–182.

Inglehart, R. (2018). *Cultural evolution. People's motivations are changing, and reshaping the world*. Cambridge University Press.

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge University Press.

Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, *7*(8), e42366.

Jamieson, D. (2002). *Morality's progress: Essays on humans, other animals, and the rest of nature*. Oxford University Press.

Jamieson, D. (2017). Slavery, carbon, and moral progress. *Ethical Theory and Moral Practice*, *20*(1), 169–183.

Jansari, A., & Parkin, A. J. (1996). Things that go bump in your life: Explaining the reminiscence bump in autobiographical memory. *Psychology and Aging*, *11*(1), 85.

Jones, C. W., & Paris, C. (2018). It's the end of the world and they know it: How dystopian fiction shapes political attitudes. *Perspectives on Politics*, *16*(4), 969–989.

Kadri, S. (2007). *The trial: Four thousand years of courtroom drama*. Random House.

Kahan, D. M. (1999). The progressive appropriation of disgust. *The Passions of Law*, *63*, 63–65.

Kahane, G. (2013). Must metaethical realism make a semantic claim? *Journal of Moral Philosophy*, *10*(2), 148–178.

Kahane, G. (2014). Evolution and impartiality. *Ethics*, *124*(2), 327–341.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2011). The neural basis of intuitive and counterintuitive moral judgement. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728–741.

Kant, I. (2006 [1784]). Idea for a universal history from a cosmopolitan perspective. In *Toward perpetual peace and other writings on politics, peace, and history*. Yale University Press.

Kass, L. R. (1997). The wisdom of repugnance: Why we should ban the cloning of humans. *Valparaiso University Law Review*, *32*, 679.

Kelly, D. (2011). *Yuck!: The nature and moral significance of disgust*. MIT Press.

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy*, *35*(1), 54–76.

Kelly, D., & Hoburg, P. (2017). A tale of two processes: On Joseph Henrich's the secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter. *Philosophical Psychology*, *30*(6), 832–848.

Kitcher, P. (2011). *The ethical project*. Harvard University Press.

Kitcher, P. (2017). Social progress. *Social Philosophy & Policy*, *34*(2), 46.

Kleingeld, P. (2011). *Kant and cosmopolitanism: The philosophical ideal of world citizenship*. Cambridge University Press.

Kling, A. (2016). Cultural intelligence. *National Affairs*, 27, 150–163.

Knapp, C. (2003). De-moralizing disgustingness. *Philosophy and Phenomenological Research*, 66(2), 253–278.

Korsgaard, C. M. (2008). Taking the law into our own hands: Kant on the right to revolution. In A. Reath, B. Herman, & C. M. Korsgaard (Eds.), *Reclaiming the history of ethics essays for John Rawls* (pp. 297–328). Cambridge University Press.

Korsgaard, C. M. (2018). *Fellow creatures: Our obligations to the other animals*. Oxford University Press.

Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The problem with morality: Impeding progress and increasing divides. *Social Issues and Policy Review*, 12(1), 218–257.

Kumar, V. (2017). Foul behavior. *Philosophers' Imprint*, 17.

Kumar, V., & Campbell, R. (2016). Honor and moral revolution. *Ethical Theory and Moral Practice*, 19(1), 147–159.

Kumar, V., & Campbell, R. (2022). *A better ape. The evolution of the moral mind and how it made us human*. Oxford University Press.

Landy, J. F., & Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science*, 10(4), 518–536.

Leeson, P. T. (2013). Gypsy law. *Public Choice*, 155(3–4), 273–292.

Leeson, P. T. (2017). *WTF?!: An economic tour of the weird*. Stanford University Press.

Leiter, B. (2008). Against convergent moral realism: The respective roles of philosophical argument and empirical evidence. In W. Sinnott-Armstrong (Ed.), *Moral psychology volume 2, the cognitive science of morality: Intuition and diversity* (pp. 333–337). MIT Press.

Leiter, B. (2015). Why Marxism still does not need normative theory. *Analyse & Kritik*, 37(1–2), 23–50.

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, 360(6396), 1465–1467.

Levy, J. T. (2016). There's no such thing as ideal theory. *Social Philosophy and Policy*, 33(1–2), 312–333.

Levy, J. T. (2020). Contra politanism. *European Journal of Political Theory*, 19(2), 162–183.

Levy, N. (2011). Culture by nature. *Philosophical Explorations*, 14(3), 237–248.

Levy, N. (2012). Ecological engineering: Reshaping our environments to achieve our goals. *Philosophy & Technology*, 25(4), 589–604.

Levy, N. (2015). Less blame, less crime? The practical implications of moral responsibility skepticism. *Journal of Practical Ethics*, 3(2).

Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective*, 6(8), 20–36.

Levy, N., & Alfano, M. (2020). Knowledge from vice: Deeply social epistemology. *Mind*, 129(515), 887–915.

Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*(4), 390–398.

Luco, A. (2019). How moral facts cause moral progress. *Journal of the American Philosophical Association*, *5*(4), 429–448.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin Books.

Macklin, R. (1977). Moral progress. *Ethics*, *87*(4), 370–382.

Madva, A. (2016). A plea for anti-anti-individualism: How oversimple psychology misleads social policy. *Ergo, an Open Access Journal of Philosophy*, *3*.

Makkreel, R. A. (1992). Purposiveness in history: Its status after Kant, Hegel, Dilthey and Habermas. *Philosophy and Social Criticism*, *18*(3–4), 221–234.

Mandelbaum, E., & Quilty-Dunn, J. (2015). Believing without Reason: or: Why Liberals shouldn't watch Fox News. *The Harvard Review of Philosophy*, *22*, 42–52.

Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, *108*(28), 11375–11380.

May, J. (2014). Does disgust influence moral judgment? *Australasian Journal of Philosophy*, *92*(1), 125–141.

McCloskey, D. N. (2006). *The bourgeois virtues: Ethics for an age of commerce*. University of Chicago Press.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 636.

McCullough, M. E. (2020). *The kindness of strangers. How a selfish ape invented a new moral code*. OneWorld.

McDowell, J. (1998). *Mind, value, and reality*. Harvard University Press.

McGrath, S. (2008). Moral disagreement and moral expertise. *Oxford Studies in Metaethics*, *3*(87), 87–108.

McGrath, S. (2011). Skepticism about moral expertise as a puzzle for moral realism. *The Journal of Philosophy*, *108*(3), 111–137.

Medin, D. L., & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological review*, *111*(4), 960.

Meidenbauer, K. L., Cowell, J. M., & Decety, J. (2018). Children's neural processing of moral scenarios provides insight into the formation and reduction of ingroup biases. *Developmental Science*, *21*(6), e12676.

Meyers, C. D. (2013). Defending moral realism from empirical evidence of disagreement. *Social Theory and Practice*, *39*(3), 373–396.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152.

Milanovic, B. (2010). *The haves and the have-nots: A brief and idiosyncratic history of global inequality*. Basic Books (AZ).

Moody-Adams, M. M. (1999). The idea of moral progress. *Metaphilosophy*, *30*(3), 168–185.

Moody-Adams, M. M. (2009). *Fieldwork in familiar places: Morality, culture, and philosophy*. Harvard University Press.

Morris, I. (2015). *Foragers, farmers, and fossil fuels: How human values evolve*. Princeton University Press.

Muldoon, R. (2015). Expanding the justificatory framework of Mill's experiments in living. *Utilitas*, 27(2), 179–194.

Murray, C. (2012). *Coming apart. The state of White America, 1960–2010.* Crown Forum.

Nagel, T. (2012). *Mind and cosmos: Why the materialist neo-Darwinian conception of nature is almost certainly false.* Oxford University Press.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment.* Oxford University Press.

Nichols, S. (2014). Process debunking and ethics. *Ethics*, 124(4), 727–749.

Nichols, S., Kumar, S., Lopez, T., Ayars, A., & Chan, H. Y. (2016). Rational learners and moral rules. *Mind & Language*, 31(5), 530–554.

Nisbet, R. (1980). *History of the idea of progress.* Basic Books.

Norberg, J. (2016). *Progress. Ten reasons to look forward to the future.* Oneworld.

Norberg, J. (2020). *Open. The story of human progress.* Atlantic Books.

Nordhaus, W. D. (1996). Do real-output and real-wage measures capture reality? The history of lighting suggests not. In *The economics of new goods* (pp. 27–70). University of Chicago Press.

Norlock, K. J. (2018). Perpetual struggle. *Hypatia*, 34(1), 6–19.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1), 97–112.

Nussbaum, M. C. (1998). "Whether from reason or prejudice": Taking money for bodily services. *The Journal of Legal Studies*, 27(S2), 693–723.

Nussbaum, M. C. (2007). On moral progress: A response to Richard Rorty. *University of Chicago Law Review*, 74(3), 939–960.

Nussbaum, M. C. (2009). *Hiding from humanity: Disgust, shame, and the law.* Princeton University Press.

Oakley, A. (2015). *Sex, gender and society.* Ashgate Publishing, Ltd.

O'Connor, C. (2019). *The origins of unfairness: Social categories and cultural evolution.* Oxford University Press.

Ostrom, E. (2015). *Governing the commons.* Cambridge University Press.

Parfit, D. (1997). Equality and priority. *Ratio*, 10(3), 202–221.

Parker, Th. (1853). *Ten sermons of religion.* Crosky, Nichols, and Company.

Persson, I., & Savulescu, J. (2019). The evolution of moral progress and biomedical moral enhancement. *Bioethics*, 33(7), 814–819.

Pettit, P. (1997). *Republicanism: A theory of freedom and government.* Clarendon Press.

Pickett, K., & Wilkinson, R. (2010). *The spirit level: Why equality is better for everyone.* Penguin UK.

Pinker, S. (2011). *The better angels of our nature. The decline of violence in history and its causes.* Allen Lane.

Pinker, S. (2018). *Enlightenment now. The case for reason, science, humanism, and progress.* Viking.

Plakias, A. (2013). The good and the gross. *Ethical Theory and Moral Practice*, 16(2), 261–278.

Pleasants, N. (2018). The structure of moral revolutions. *Social Theory and Practice*, 44(4), 567–592.

Prinz, J. (2007). *The emotional construction of morals.* Oxford University Press.

Prinz, J. (2011). Against empathy. *The Southern Journal of Philosophy*, 49, 214–233.

Protzko, J., & Schooler, J. W. (2019). Kids these days: Why the youth of today seem lacking. *Science Advances*, *5*(10).

Putnam, R. D. (2000). *Bowling alone. The collapse and revival of American community*. Simon & Schuster.

Rachels, R., & Rachels, S. (2015). *The elements of moral philosophy*. McGraw-Hill.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57.

Railton, P. (1986). Moral realism. *Philosophical Review*, *95*(2), 163–207.

Railton, P. (2017). Moral learning: Conceptual foundations and normative relevance. *Cognition*, *167*, 172–190.

Reyes, J. W. (2015). Lead exposure and behavior: Effects on antisocial and risky behavior among children and adolescents. *Economic Inquiry*, *53*(3), 1580–1605.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.

Richter, D., & Berger, K. (2013). Nehmen psychische Störungen zu? *Psychiatrische Praxis*, *40*(4), 176–182.

Richter, D., Berger, K., & Reker, T. (2008). Nehmen psychische Störungen zu? Eine systematische Literaturübersicht. *Psychiatrische Praxis*, *35*(7), 321–330.

Ridley, M. (2011). *The rational optimist*. Fourth Estate.

Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, *27*(2), E-43.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, *9*(2), 131–165.

Rønnow-Rasmussen, T. (2017). On locating value in making moral Progress. *Ethical Theory and Moral Practice*, *20*(1), 137–152.

Rose, D., Schaffer, J., & Tobia, K. (2020). Folk teleology drives persistence judgments. *Synthese*, *197*(12), 5491–5509.

Rosenberg, A. (2019). *How history gets things wrong: The neuroscience of our addiction to stories*. MIT Press.

Rosling, H. (2018). *Factfulness. Ten reasons we're wrong about the world — and why things are better than you think*. Sceptre.

Roth, A. (2012). Ethical progress as problem-resolving. *Journal of Political Philosophy*, *20*(4), 384–406.

Roth, M. P. (2014). *An eye for an eye: A global history of crime and punishment*. Reaktion Books.

Rowland, R. (2017), The significance of significant fundamental moral disagreement[1]. *Noûs*, *51*, 802–831.

Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science*, *8*(2), 67–73.

Satz, D. (2010). *Why some things should not be for sale: The moral limits of markets*. Oxford University Press.

Sauer, H. (2011). The appropriateness of emotions. Moral judgment, moral emotions, and the conflation problem. *Ethical Perspectives*, *18*(1).

Sauer, H. (2012). Morally irrelevant factors: What's left of the dual process-model of moral cognition? *Philosophical Psychology*, *25*(6), 783–811.

Sauer, H. (2015). Can't we all disagree more constructively? Moral foundations, moral reasoning, and political disagreement. *Neuroethics*, *8*(2), 153–169.

Sauer, H. (2017a). No excuses for moral realism. *Philosophical Studies*, 1–26.

Sauer, H. (2017b). *Moral judgments as educated intuitions: A rationalist theory of moral judgment*. MIT Press.

Sauer, H. (2018). *Debunking arguments in ethics*. Cambridge University Press.

Sauer, H. (2019a). The argument from agreement: How universal values undermine moral realism. *Ratio*, *32*(4), 339–352.

Sauer, H. (2019b). Butchering benevolence moral progress beyond the expanding circle. *Ethical Theory and Moral Practice*, *22*(1), 153–167.

Scanlon, T. (2018). *Why does inequality matter?*. Oxford University Press.

Scheidel, W. (2018). *The great leveler: Violence and the history of inequality from the stone age to the twenty-first century*. Princeton University Press.

Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*(8), 1147–1163.

Schinkel, A., & de Ruyter, D. J. (2017). Individual moral development and moral progress. *Ethical Theory and Moral Practice*, *20*(1), 121–136.

Schmidtz, D. (2008). Separateness, suffering, and moral theory. In D. Schmidtz (Ed.), *Person, polis, planet. Essays in applied philosophy* (pp. 145–165). Oxford University Press.

Schröder, M. (2018). *Warum es uns noch nie so gut ging, und wir trotzdem ständig von Krisen reden*. Benevento.

Schulz, J., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018). *The origins of WEIRD psychology*. Available at SSRN 3201031.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45.

Scott, J. C. (2017). *Against the grain: A deep history of the earliest states*. Yale University Press.

Searle, J. (1995). *The construction of social reality*. Penguin.

Sen, A. (2009). *The idea of justice*. Harvard University Press.

Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford University Press.

Shellenberger, M. (2020). *Apocalypse never. Why environmental alarmism hurts us all*. HarperCollins.

Shermer, M. (2015). *The Moral Arc. How science makes us better people*. St. Martin's Press.

Simler, K., & Hanson, R. (2017). *The elephant in the brain: Hidden motives in everyday life*. Oxford University Press.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics*, *9*(3–4), 331–352.

Singer, P. (2011a). *Practical ethics*. Cambridge University Press.

Singer, P. (2011b). *The expanding circle: Ethics, evolution, and moral progress*. Princeton University Press.

Smith, A. (2004 [1759]). *The theory of moral sentiments*. Cambridge University Press.

Smith, D. L. (2011). *Less than human: Why we demean, enslave, and exterminate others*. St. Martin's Press.

Smith, M. (2000). Moral realism. In H. LaFollette (Ed.), *The Blackwell guide to ethical theory*. Blackwell.

Sneddon, A. (2009). Normative ethics and the prospects of an empirical contribution to assessment of moral disagreement and moral realism. *The Journal of Value Inquiry*, *43*(4), 447–455.

Somin, I. (2016). *Democracy and political ignorance: Why smaller government is smarter*. Stanford University Press.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Cambridge University Press.

Sripada, C. S., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition*. Oxford University Press.

Stanley, M. L., Dougherty, A. M., Yang, B. W., Henne, P., & De Brigard, F. (2018). Reasons probably won't change your mind: The role of reasons in revising moral decisions. *Journal of Experimental Psychology: General*, *147*(7), 962.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Sterelny, K. (2007a). Snafus: An evolutionary perspective. *Biological Theory*, *2*(3), 317–328.

Sterelny, K. (2007b). Social intelligence, human intelligence and niche construction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1480), 719–730.

Sterelny, K. (2010). Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465–481.

Sterelny, K. (2012). *The evolved apprentice*. MIT Press.

Sterelny, K. (2019). Evolutionary foundations for a theory of moral progress? *Analyse & Kritik*, *41*(2), 205–216.

Stern, N. (2006). *The economics of climate change*. Cambridge University Press.

Stevenson, B., & Wolfers, J. (2008). *Economic growth and subjective well-being: Reassessing the Easterlin paradox* (No. w14282). National Bureau of Economic Research.

Stokes, P. (2017). Towards a new epistemology of moral progress. *European Journal of Philosophy*, *25*(4), 1824–1843.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, *127*(1), 109–166.

Summers, J. S. (2017). Rationalizing our way into moral progress. *Ethical Theory and Moral Practice*, *19*(5), 93–104.

Sunstein, C. R. (1991). Why markets don't stop discrimination. *Social Philosophy and Policy*, *8*(2), 22–37.

Sunstein, C. R. (2019). *How change happens*. MIT Press.

Surprenant, C. W., & Brennan, J. (2019). *Injustice for all: How financial incentives corrupted and can fix the US criminal justice system*. Routledge.

Tainter, J. (1988). *The collapse of complex societies*. Cambridge University Press.

Tam, A. (2020). Why moral reasoning is insufficient for moral progress. *Journal of Political Philosophy*, *28*(1), 73–96.

Teixeira, R. (2017). *The optimistic leftist. Why the 21st century will be better than you think*. St. Martin's Press.

Tersman, F. (2017), Debunking and disagreement. *Noûs*, *51*, 754–774.

Tiberius, V. (2014). *Moral psychology: A contemporary introduction*. Routledge.

Timmons, M. (2012). *Moral theory: An introduction*. Rowman & Littlefield.

Tropman, E. (2012). Can Cornell moral realism adequately account for moral knowledge? *Theoria*, *78*(1), 26–46.

Tropman, E. (2014). Why Cornell moral realism cannot provide an adequate account of moral knowledge. *Theoria*, *80*, 184–190.

Trüper, H., Chakrabarty, D., & Subrahmanyam, S. (Eds.). (2015). *Historical teleologies in the modern world*. Bloomsbury.

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, *120*(1), 65.

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*(6), 474–480.

Uttich, K., Tsai, G., & Lombrozo, T. (2014). Exploring metaethical commitments: Moral objectivity and moral progress. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 188–208). Bloomsbury Academic.

van Bavel, B. (2018). *De Onzichtbare Hand. Hoe markteconomieën opkomen en neergaan*. Prometheus.

van den Hoven, J., Miller, S., & Pogge, T. (Eds.). (2017). *Designing in ethics*. Cambridge University Press.

van der Vossen, B., & Brennan, J. (2018). *In defense of openness: Why global freedom is the humane solution to global poverty*. Oxford University Press.

van Vugt, R. H., & Kaiser, R. B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, *63*(3), 182–196.

Västfjäll, D., Slovic, P., & Mayorga, M. (2015). Pseudoinefficacy: Negative feelings from children who cannot be helped reduce warm glow for children who can be helped. *Frontiers in Psychology*, *6*, 616.

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PLoS ONE*, *9*(6), e100115.

Veenhoven, R., & Vergunst, F. (2014). The Easterlin illusion: Economic growth does go with greater happiness. *International Journal of Happiness and Development*, *1*(4), 311–343.

Vogler, C. (2002). *Reasonably vicious*. Harvard University Press.

Waller, B. N. (1992). Moral conversion without moral realism. *The Southern Journal of Philosophy*, *30*(3), 129–137.

Waytz, A., Iyer, R., Young, L., Haidt, J., & Graham, J. (2019). Ideological differences in the expanse of the moral circle. *Nature Communications*, *10*(1), 1–12.

Weber, M. (2004 [1919]). Science as a vocation. In M. Weber (Ed.), *The vocation lectures*. Hackett.

Welzel, C. (2013). *Freedom rising*. Cambridge University Press.

Whitman, D. (1998). *The optimism gap: The I'm ok—they're not syndrome and the myth of American decline*. Walker and Company.

Widerquist, K., & McCall, G. (2015). Myths about the state of nature and the reality of stateless societies. *Analyse & Kritik*, *37*(1–2), 233–258.

Williams, E. G. (2015). The possibility of an ongoing moral catastrophe. *Ethical Theory and Moral Practice*, *18*(5), 971–982.

Wilson, C. (2010). Moral progress without moral realism. *Philosophical Papers*, *39*(1), 97–116.

Wolff, J., & De-Shalit, A. (2007). *Disadvantage*. Oxford University Press.

Wong, D. (2006). *Natural moralities. A defense of pluralistic relativism*. Oxford University Press.

Wright, R. (2006). *A short history of progress*. Canongate Books.

Index