Moral Reasoning and Moral Progress

Victor Kumar & Joshua May

Forthcoming in *The Oxford Handbook of Metaethics*,
D. Copp and C. Rosati (eds.), Oxford University Press.

Abstract: Can reasoning improve moral judgments and lead to moral progress? Pessimistic answers to this question are often based on caricatures of reasoning, weak scientific evidence, and flawed interpretations of solid evidence. In support of optimism, we discuss three forms of moral reasoning (principle reasoning, consistency reasoning, and social proof) that can spur progressive changes in attitudes and behavior on a variety of issues, such as charitable giving, gay rights, and meat consumption. We conclude that moral reasoning, particularly when embedded in social networks with mutual trust and respect, is integral to moral progress.

Word count: 6,957 (8,782 w/references)

Key words: moral judgment, post hoc rationalization, moral principles, consistency reasoning, social learning

1. Introduction

Megan Phelps-Roper grew up in Kansas as a member of the small but vocal Westboro Baptist church, infamous for its caustic denunciations of homosexuality, Jewish people, and the entire country for tolerating perceived sins. Church members have regularly picketed military funerals while defacing the American flag and holding signs displaying hate speech. Phelps-Roper left the church in 2012 after being convinced of its faults by people she befriended on Twitter. In a TED talk and various interviews, Phelps-Roper makes clear that key to her change in view was engaging with the thoughtful questions and comments that friends made on the social media platform. Friendliness and respect were necessary for Phelps-Roper to lower her guard and appreciate the inconsistencies in her worldview, but she contends that without the arguments she might never have changed her moral and political attitudes.

In general, if people are to make better moral judgments and engage in more virtuous behavior, it seems likely that moral reasoning must play a central role. Emotions and other psychological processes, by themselves, are as likely to distort moral judgment as improve it. The question we explore in this chapter is thus whether and how moral reasoning can improve moral judgments and even lead to moral progress.

We begin by considering the idea that moral reasoning is powerless to affect either moral judgment or behavior but offer reasons to think this pessimism, albeit grounded in scientific research, is overstated (Section 2). We then turn from defense to offense and discuss empirical and theoretical evidence for the power of moral reasoning. We focus on three forms—principle reasoning, consistency

reasoning, and social proof—and how they apply to several case studies in moral progress, such as charitable giving, gay rights, and meat consumption (Section 3). We conclude that there is concrete evidence of the power of moral reasoning, particularly when it is socially-embedded (Section 4). Arguments might rarely persuade immediately, but when encountered in friendly exchanges awash in mutual trust and respect the ideas simmer and, in the long-term, are integral to progressive social change.

2. Pessimism about Moral Reasoning

Everyday experience might suggest that moral reasoning is utterly impotent. With compelling evidence and arguments, you may have convinced others that stocks are better than bonds, that flossing doesn't prevent cavities, or that you're a descendant of Sojourner Truth. But have you ever convinced anyone to change their moral attitudes about abortion, affirmative action, or vaccine mandates?

Perhaps you've convinced yourself with moral reasoning. You might come to believe that the death penalty should be abolished, because you learned that false convictions are much higher than you had ever imagined. An empirical assumption has been corrected, but can reasoning do *more*, such as shift your moral values? Even if you have experienced a fundamental change in your moral outlook or behavior, for all you know it could have been due to emotional responses or simple habituation. It's hard to tell through introspection alone. Fortunately, there is research in the social sciences that investigates whether moral reasoning can change attitudes or behavior. We'll consider, but ultimately resist, pessimism that arises from two lines of research.

2a. Moral Judgment: Post Hoc Rationalization

One body of research seems to show that emotions drive moral judgment, while reasoning only comes later to rationalize the judgment to oneself and others. We'll argue that the research does not in fact support this position. To be clear, our aim is not to deny that emotions play any role in moral judgment (see May and Kumar 2018), but rather to challenge a popular view that privileges emotions and relegates reasoning to the role of post-hoc rationalization.

First, let's consider whether moral beliefs are more powerfully influenced by emotions, conceived as distinct from reasoning. A number of philosophers and scientists have argued for this sentimentalist view on the basis of empirical research (e.g., Nichols 2004; Prinz 2007). A prime source of support are studies that manipulate disgust in one group of participants but not another and then compare their moral evaluations of hypothetical dilemmas (e.g., Schnall, Haidt, Clore, & Jordan 2008). A series of experiments reported that merely priming people with disgust—e.g., using a foul smell or drinking a bitter beverage—made their moral judgments more harsh.

However, these studies haven't stood the test of time. The published studies report small and inconsistent effects that often weren't replicated in the unpublished literature (May 2014). A meta-analysis of both published and unpublished studies suggests that the effect is miniscule at best and likely a spurious result of publication bias (Landy & Goodwin 2015). In general, it seems that emotions affect moral judgment only when they are *integrated* with relevant reasoning, not incidental primes (Kumar 2017a; May 2018). Thus, disgust seems to affect moral judgment when people infer that the object of

evaluation is itself morally repellent—for example, when a person or action is regarded as dishonest or exploitative.

Of course, even if incidental emotions do not exert a significant effect on moral judgments, it doesn't follow that reasoning plays a role. In general, a wide range of research suggests that unconscious psychological processes underlie judgments and behavior. Reasons can be given, but they tend to be post-hoc rationalizations – devised to relieve pressure from an interlocutor, rather than accurately report the reasons that led one to form the judgment in the first place. In the domain of moral judgment, a classic piece of evidence in support of this hypothesis is the phenomenon of "moral dumbfounding." Researchers present participants with a "harmless" taboo violation, such as kissing between siblings or failing to keep a promise to someone who is now deceased. Participants condemn the violation but are dumbfounded when asked for reasons (Haidt, Koller, & Dias 1993). Hence, reasoning is not in the picture.

Several problems afflict the research suggesting that reasoning is only post-hoc rationalization. First, the inability to give reasons does not show that reasoning is absent (Dwyer 2009; Railton 2014; May 2018; Nichols 2021: 40ff). Some taboo violations don't cause harm in a given case, but philosophers have suggested that people may be reasoning unconsciously that taboo violations *risk* harm (Jacobson 2012; Railton 2014). Even if two consenting adult siblings have protected intercourse just once and manage to avoid damaging their relationship with each other and with other family members, risking such harm is objectionable. Empirical evidence supports this interpretation by showing that judgments about taboo violations incorporate assessments of risk (Royzman, Kim, & Leeman 2015; Stanley, Yin, & Sinnott-Armstrong 2019).

Now, some theorists assume that reasoning is necessarily conscious, and thus participants should be able to access these purported reasons for their moral beliefs. But reasoning at its core involves rule-governed inference from one set of beliefs to another, which can be unconscious. Not all unconscious processes are reasoning, of course. The operations of the visual system don't necessarily amount to reasoning, because they appear to be transitions not among beliefs but rather "subpersonal" representations (May 2018: 9; May & Kumar 2018). Indeed, it is likely that much reasoning is unconscious (Harman, Mason, & Sinnott-Armstrong 2010). Some may be unconscious from the start, while other reasoning becomes unconscious through habituation.

Furthermore, even if reasoning follows, rather than precedes, a given moral judgment, it's possible that this reasoning influences *other* moral judgments downstream from it (Doris 2015: 147; Summers 2017; Cushman 2020). When people give reasons for their beliefs, they are often motivated to live up to those reasons in the future. This is especially true when the reasons are offered in social exchanges with friends and family, since people then feel responsible to their interlocutors. Still, post-hoc rationalization is often thought to stem from motivated reasoning. For example, many people want to continue eating meat and studies suggest that they find reasons to believe that animals don't suffer on factory farms (May and Kumar 2023). However, motivated reasoning can be epistemically rational in a social context. Studies have shown that people with different motives can exchange reasons that lead to new beliefs (Mercier and Sperber 2018). If a liberal and a conservative both offer rationalizations that are motivated to defend their existing political views, but they are both open to good reasoning and vigilant toward bad reasoning, then by exchanging reasons they can nevertheless live up to the best rationalization that arises in their discussion. Phelps-Roper was certainly motivated to defend her church's views, and her opponents on Twitter were motivated to defend their own worldview, but dialogue led to progress.

Other, more specific problems beset the argument that uses moral dumbfounding as evidence against the power of moral reasoning. First, it's possible that cases of moral dumbfounding are unusual.

People may not be able to give reasons for why taboo violations are wrong, but they generally have no problem explaining why harm or fairness violations are wrong. There is no mystery why it's wrong that one person assaulted another or took more than their fair share. So, reasoning may be absent only in a restricted subset of moral judgments. Second, researchers claim that participants are morally dumbfounded, but are they? When asked whether it is wrong for someone to be intimate with their sibling, participants can easily report that incest in general is wrong. Some may have reasoned consciously: "Incest is wrong; kissing a sibling is incestuous; so, kissing a sibling is wrong." Participants are "dumbfounded," perhaps, only because they interpret researchers as asking for a further reason why incest is wrong. But that is a deeper question, about the justification for a rule rather than a particular rule violation, one that even moral philosophers have a hard time answering (Mallon and Nichols 2011: 285; Nichols 2021: 40). Some participants might even believe in a religious justification for the rule but hesitate to offer it to researchers who they expect to be seeking a secular justification.

2b. Moral Behavior: Unexceptional Ethicists

So far, we have criticized the empirically-driven idea that moral judgments are guided by emotions and other putatively arational processes, while reasoning functions only to provide post-hoc rationalizations. Another basis for pessimism rests on empirical research suggesting more directly that moral reasoning hardly influences behavior.

Professional ethicists study moral arguments for a living, yet they don't necessarily behave more ethically than other people. One survey found that most fellow philosophers don't even rate ethicists as more moral than other philosophy professors or other people with a similar background (Schwitzgebel & Rust 2009). But colleagues only observe a small number of one's decisions. Many ethical choices—such as charitable donations and responding to student emails—aren't observed by friends or family, let alone colleagues. What if we go beyond peer ratings and directly measure such choices?

It's difficult to study the behavior of ethicists in natural settings, but thanks to the pioneering work of Eric Schwitzgebel and his collaborators, a series of creative studies makes a valiant attempt. The research compares the behavior of moral philosophy professors to either the general public or other academics. Ethicists in these studies weren't, for example, any more likely to vote, return library books, respond to student emails, donate to charity, or exhibit courtesy toward others at conferences (for review, see Rust & Schwitzgebel 2014; Schönegger & Wagner 2019). On all of these various measures, ethicists perform no better than people who aren't regularly steeped in moral arguments. (The one key exception is eating meat, which we'll turn to later.)

However, these studies don't necessarily show that moral reasoning is generally weak or impotent. On the one hand, Schwitzgebel and his collaborators do study a range of different behaviors. On the other hand, these behaviors aren't necessarily representative of the topics moral philosophers study most, which tend to involve public policy. Even if reasoning about ethics doesn't influence how often ethicists return library books and respond to student emails, we can't necessarily generalize and conclude that reasoning doesn't shape their attitudes toward abortion, affirmative action, and capital punishment.

Moreover, these studies focus only on professional ethicists, who aren't necessarily representative reasoners or the only models of good reasoning. Like lawyers, philosophers are highly skilled at articulating reasons for their views and finding holes in the opposition—slow "System 2" thinking on steroids. Haidt may well be right that *this* form of reasoning is liable to post-hoc rationalization. It's a

valuable skill in many contexts, such as courtrooms and debate competitions, but clever rationalizations don't capture all forms of reasoning and they aren't well-suited for changing behavior. Ordinary people, in contrast, might embrace more intuitive forms of reasoning as well, and ones that are amenable to being influenced by the wisdom of the crowd. If our concern is ultimately moral progress generally, not just among professional ethicists, then we should consider all types of moral reasoning and among a wider population.

3. Forms of Moral Reasoning

So far we have argued against pessimism about the power of moral reasoning. There is also positive scientific evidence that a wide range of moral attitudes and behaviors are driven by various forms of reasoning (Kennett & Fine 2009; Campbell and Kumar 2012; Sauer 2017; Kumar 2017b; May 2018; Nichols 2021; Kumar & Campbell 2022). We now shift from defense to offense by discussing three forms of moral reasoning that can change attitudes and behavior, even on controversial moral issues.

3a. Principle Reasoning

A common form of reasoning discussed by ethicists derives particular conclusions from general moral principles. Some people reason, for example, that in order to minimize overall harm vaccination against measles should be required of all children in public schools. In other cases, cost-benefit analyses take a back seat to general rules. For instance, many people believe that healthcare providers should be honest with their patients even if breaking bad news causes psychological distress. Here reasoning applies general moral principles or basic values to a particular case. Whether consequentialist or not (Greene 2014), principle reasoning can be an important mechanism of social change and moral progress. Still, one might wonder, can reasoning lead to more fundamental change, such as abandoning or revising a moral principle?

Some research suggests that beliefs in moral principles are sensitive to arguments, such as counter-examples and counter-arguments. For instance, although most people condemn sexual intercourse among siblings, even when it occurs among consenting adults, participants tend to find it more morally acceptable after reading an evolutionary argument meant to debunk our automatic aversion to all forms of incest (Paxton, Ungar, & Greene 2012). In another study, participants tended to strongly agree with the utilitarian principle "In the context of life or death situations, always take whatever means necessary to save the most lives." But their credence in this principle diminished after reading about the famous Transplant thought experiment in which one person is killed so his organs can be used to save five patients (Horne, Powell, & Hummel 2015).

Similarly, both controlled deliberation and automatic unconscious inference drive our resolutions of moral dilemmas that pit fundamental moral values against each other (Mikhail 2011; Greene 2014). Consider two famous trolley cases. In Switch, you're asked whether it's morally acceptable to pull a lever that will send a trolley down a sidetrack, running over and killing one person instead of five. In Push, you're asked about the appropriateness of pushing a large person off a footbridge to their death in order to prevent a trolley from running over and killing five people. Most people say that it's permissible to pull the switch but wrong to push the large person. Contrary to some early reports, judgments about such

sacrificial dilemmas are not simply emotional aversions to up-close and personal harm, like pushing. Rather, participants seem to be unconsciously applying rules, such as a prohibition on battery, prototypical violence, using another person as a means to one's ends, or some combination of these (Greene 2013; Mikhail 2014; Feltz & May 2017; May 2018: ch. 3).

Of course, abstract moral principles are open to interpretation, and our endorsement of them is malleable. Perhaps it is unsurprising that principle reasoning can nudge such attitudes around. But can reasoning yield attitude changes that produce different choices about *real-world moral issues*?

Consider the topic of charitable giving. Peter Singer and other ethicists have long argued that the average citizen in wealthy nations like America and Australia ought to donate more of their income to charitable organizations that help people in the world who are suffering from preventable illness, starvation, and death (Singer 1972). The arguments often turn on general moral principles like "if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it" (Singer 1972: 231). Members of the "effective altruism" movement have also made impassioned arguments in both academic journals and news media (MacAskill 2015). Yet most people, including ethics professors, seem happy to regularly purchase a pricey latte over a mosquito bed net that helps prevent a child in Uganda from contracting a deadly disease like malaria (Schwitzgebel & Rust 2014). Consumerism seems unmoved by consequentialism.

When people are moved toward philanthropy, the evidence might seem to suggest that what's effective are emotionally-evocative appeals. Better to use photographs and stories of particular individuals in need—so-called "identifiable victims"—rather than statistical information quantifying the suffering (e.g., Small, Loewenstein, & Slovic 2007). Nevertheless, people are clearly responsive to relevant reasons for or against donating. For example, a review of over 500 studies found that people are more likely to donate if the amount requested is relatively low and if they believe their donation will be effective (for review, see Bekkers & Wiepking 2011).

Some recent studies from several different labs have found that philosophical arguments, specifically, can motivate people to donate more of their hard-earned money to charitable organizations. The details differ across studies, but participants are randomly assigned to read either a philosophical argument (in favor of donating to charities that effectively help the global poor) or no moral argument (e.g., no text at all or a text about an entirely different topic). Some of the experiments also include conditions in which some participants instead read a more concrete appeal which takes the form of a narrative that describes a real child who benefited immensely from a charitable organization. All participants are then asked a series of questions, including how much of a bonus they are willing to donate if they are randomly selected to be one of the few winners (that is, a bonus amount earned on top of their base payment for participating). One series of studies found that people pledged to donate slightly more of the bonus when they read one of various philosophical arguments, as opposed to no text at all or a text on a topic unrelated to charitable giving (Lindauer, Mayorga, Greene, Slovic, Västfjäll, & Singer 2020; Buckland, Lindauer, Rodríguez-Arias, & Véliz 2021). Other experiments have also found that moral arguments can motivate slightly higher donations (Schwitzgebel & Cushman 2020), especially when the arguments are engaging like conventional narratives (Schwitzgebel, McVey, & May 2022).

Moral reasoning can also have an impact on meat consumption. Researchers randomly assigned college students in an ethics class to study either an argument in favor of vegetarianism or instead an argument in favor of charitable giving (Schwitzgebel, Cokelet, and Singer 2020). The article students read on vegetarianism argues that eating meat from factory farms flouts a moral principle, namely that "it is wrong to cause pain unless there is a good enough reason" (Rachels 2004: 70). As payment for

participating in the study, students received meal vouchers that could be redeemed at the campus dining hall, which allowed the researchers to track which foods were purchased. Compared to the students who studied the argument in favor of charitable giving, students in the vegetarian group purchased slightly more meatless meals in the subsequent weeks. The effect is small but it's remarkable to find any reduction in meat consumption due to studying a moral argument (for a replication, see Schwitzgebel, Cokelet, and Singer 2023).

3b. Consistency Reasoning

In the previous section we argued that principle reasoning does have an impact on moral belief and behavior. It leads people to donate more money to charity, consume less meat, and reduce their confidence in moral principles subject to counter-arguments.

However, it's also true that principle reasoning is a source of rationalization. One reason is that principles can be open-ended and leave much room for interpretation. One can interpret a principle in a stringent way to condemn behavior that one already disapproves of, or one can interpret it in a relaxed way to excuse behavior that one has self-interested reasons to indulge in. Another reason that principle reasoning is recruited for rationalization is that multiple, conflicting principles can often be brought to bear on a situation. So, one can appeal to whichever principle supports one's views and ignore other principles (see e.g. Ditto, Pizarro, & Tannenbaum 2009). Yet another limitation of principle reasoning is that it has a tendency to be abstruse. Philosophers and legal experts can follow along, but most people are liable to lose the plot.

Another type of reasoning relies more on narrative and is widely employed: consistency reasoning (Campbell and Kumar 2012; Holyoak & Powell 2016). Instead of applying or challenging abstract principles, consistency reasoning involves marshaling intuitive responses to a given case and applying it to other cases, so long as there are no morally relevant differences between them. In short, consistency reasoning involves treating like cases alike (also described as "analogical reasoning" or "casuistry").

To illustrate, consistency reasoning seems to be at work in many people's choices to adopt a vegan or vegetarian diet. Many people believe that it would be wrong to support a system that treats dogs and cats with cruelty. But then it seems likewise wrong to support such treatment of pigs and cows on factory farms. There are obvious differences between pets and farm animals, but these differences do not seem morally relevant. For example, pigs seem to be no less cognitively sophisticated than dogs or cats. Nor do they have a diminished capacity to feel pain and pleasure. Appreciating the inconsistencies can ultimately undermine or debunk one's belief that eating meat from factory farms is morally acceptable (Kumar and May 2019; see also Kumar and Campbell 2012).

Everyday examples of consistency reasoning abound, but there is also empirical evidence of its impact on moral beliefs. Recall the two trolley cases discussed in the previous section. Ordinarily, participants consider only one of these cases on their own. In experiments where participants are presented with both cases, if they have judged that it is wrong to push the large person, they are then more likely to say that it is also wrong to pull the switch (Schwtizgebel & Cushman 2012; Liao, Wiegmann, Alexander, & Vong 2012). This is often described as an irrational "order effect," but it is likely that people are engaging in consistency reasoning: if it's wrong to sacrifice one for the sake of five in Push, then it's wrong to do so in Switch too (Horne & Livengood 2017; Barak-Corren, Tsay, Cushman, & Bazerman 2018).

Consistency reasoning is typically a social process. For example, people don't usually become vegans or vegetarians in a social vacuum: a trusted friend who knows about your sympathy for dogs might gently point out the similarities between, say, factory farms and the practice of dog fighting. Troubled and searching for a morally relevant difference, you might suggest that dogs are such intelligent creatures; she responds by telling a story that highlights the equally impressive intelligence of pigs. In a similar way, consistency reasoning is common in the moral education that parents offer to children (Campbell 2017). A child has just been caught hurting her classmate or stealing his toy and is unrepentant. The parent asks, "Well, but how would you like it if he did the same thing to you?" Eventually, most children realize that if it would be wrong for another child to harm her, it's also wrong for her to harm him.

Consistency reasoning has seemed to play a role in many progressive social movements for marginalized groups in the 20th and 21st centuries. Of course, many factors have been at play in the liberation of women and people of color, but one plausible factor is that people have reasoned via consistency: there are no morally relevant differences between races or genders that would justify subordination of Black people and women.

One of the most significant episodes of moral progress over the last half-century is the gay rights revolution (Kumar and Campbell 2022: 210-214). In the middle of the 20th century, it was much more likely that openly gay people would be ostracized from their families and communities. Of course, the gay rights revolution still has a long way to go, but significant progress has been made. Same-sex marriage has been legalized in many countries, such as America and Australia. And a massive study of implicit and explicit biases shows that anti-gay bias has dropped dramatically, declining by 33% just over the past decade (Charlesworth and Banaji 2019a, 2019b, in press).

Sexual orientation is different from certain other social identities in two important ways. First, it's relatively hidden. Someone's race is usually evident from their appearance, whereas their sexual orientation is not. Second, unlike race, gay people are interspersed throughout the population. In the U.S. and many other countries, by contrast, White and Black people experience high degrees of segregation. These two features have helped spur consistency reasoning (Kumar, Kodipady, & Young 2022). Many straight people have discovered that a close friend or family member, someone they already know and admire, is gay. And they have gone on to reason that if there's nothing morally objectionable about this gay person in their lives, there must not be anything wrong with the many other gay people who are strangers. This kind of reasoning is blocked sometimes, when people make special exceptions, usually on spurious grounds. But, in general, consistency reasoning has succeeded often enough to help contribute to moral progress.

3c. Social Proof

The final form of reasoning to discuss is an inherently social form or learning. Humans are experts at harnessing the knowledge of others by paying attention to what others do and copying them (Henrich 2015; Tankard & Paluck 2016; Bicchieri 2017; Kelly & Davis 2018). Of course, conformity can be motivated merely by a desire to maintain group affiliation or to avoid standing out from the crowd (compare Schwitzgebel 2019), but it can also be driven by a desire to form *accurate* beliefs (Cialdini & Goldstein 2004). For the latter, we prefer the term "social proof" (Cialdini 2009: 88) to emphasize that the behavior of others is serving as evidence about what to do.

To illustrate, imagine you're visiting a new city, walking its bustling streets while searching for a place to eat. On your walk, you see a group of people looking up toward the sky, which prompts you to look as well and catch a glimpse of a beautiful rare bird. Then you get to a street with many cafes and find yourself drawn to the busiest one while avoiding those without a patron in sight. Your motivation to look up and to dine at the popular eatery isn't just mindless mimicry; you're using the behavior of others as *evidence* of what's best to do. That people are looking up is evidence that there is something worth seeing; that people are flocking to a restaurant is evidence that the food and service are superb.

Numerous lab and field experiments from many different researchers and disciplines demonstrate how social proof works in human psychology. For example, the famous bystander effect in social psychology can be understood as exhibiting social proof in this sense. These studies on the bystander effect show that people are less likely to help a person apparently in need when many other people around do nothing (for review, see Latané and Nida 1981). What's often overlooked, however, is that helping increases dramatically when bystanders merely act startled by the sound of something crashing down on an apparent victim in another room. So, when bystanders exhibit no reaction whatsoever like expert Stoics, participants are likely taking this behavior as evidence that nothing serious really happened or that the person doesn't really need help (the "social influence" explanation). Indeed, the idea that participants are merely worried about looking foolish by helping ("audience inhibition") rests on the belief that no help is really needed, which is grounded in social proof.

Many studies demonstrating social proof directly measure changes in attitudes and behaviors regarding contemporary moral issues. Aid organizations have found that people in developing communities are more likely to curb harmful practices of child marriage and open defecation if they learn that others in the community are conforming to the new norm (Bicchieri 2017). Another group of researchers partnered with a hotel and tracked how often guests reuse their towels instead of having them laundered daily (Goldstein, Cialdini, & Griskevicius 2008). Some guests read the typical sign in their room which explained that reusing towels benefits the environment by using less water. Other guests, however, saw a sign that merely added "the majority of guests reuse their towels," which led twice as many people in this group to comply.

The behavior of others presumably shows where their own confidence lies. If most people merely say you ought to shun child marriage, that's some evidence that it's the right thing to do. But if most people still allow their young children to get married, then the evidence their beliefs provide is weak. If, on the other hand, most people *act* on that moral belief, then that serves as powerful evidence that prohibiting child marriage is not just a good idea but certain enough to act on, even if it requires personal sacrifices. Behavior, particularly costly behavior, is often better evidence than speech about what people really believe.

Reasoning is influenced most by individuals and groups one particularly trusts or respects. People are more likely to follow norms and behaviors displayed by their friends, family, and compatriots, for example (Tankard & Paluck 2016; Bicchieri 2017). Indeed, the closer others are to you in your social network, the more their behavior influences yours, including your disposition to vote, smoke, eat certain foods, or even get a divorce (Christakis & Fowler 2013). Studies of charitable giving have found that people are inspired to donate more after reading about moral exemplars in their community whom they can easily relate to, compared to famous historical figures from decades past, such as Rosa Parks or Nelson Mandela (Han, Workman, May, Scholtens, Dawson, Glenn, & Meindl 2022). Thus, social proof doesn't just work by providing information about consensus (wisdom of the crowd), but also by providing information about what people *like you* should do.

We are also influenced by the behavior of group members whom we regard as especially skilled or prestigious (Henrich 2015). One landmark field study demonstrated how popular teenagers can reduce bullying through their social influence (Paluck and Shepherd 2012). By administering surveys to students in an American high school, researchers identified the popular kids, and then randomly assigned some of these "social referents" to receive training on how to model anti-harassment (e.g., through speaking out, talking to peers, and selling wristbands). At the end of the school year, students with more direct social ties to these trend-setters were more inclined to perceive harassment as unacceptable and were less likely to be disciplined for conflicts with peers. In this way, social proof can get a foothold on social change even when a majority of individuals don't follow the new norm. Again, we must note that it's often an open question whether people are merely conforming, rather than treating others' behavior as evidence. That depends on the underlying psychological mechanism, which further research might be able to clarify, but many rational beliefs are based on indirect evidence of this kind (see, e.g., Levy 2022).

Social proof can also start small by indicating behavioral trends. Suppose you've considered whether to reduce your meat consumption but remain undecided. If you're uncertain about whether to become a vegetarian or reducetarian, you can be pushed over the edge by the knowledge that the diet is increasingly gaining acceptance in your community (for discussion, see May & Kumar 2022). One study approached customers who were waiting in line at a cafe on the campus of Stanford University (Sparkman & Walton 2017; see also Sparkman et al. 2020). Some of the customers read that 30% of Americans make an effort to reduce their meat consumption, which could naturally be read as indicating that a fixed minority of people are reducers (perhaps only animal rights activists, people with relevant medical conditions, and so on). Another group of customers read that 30% of Americans are *starting* to reduce their meat consumption, suggesting a general trend. Participants in this condition ordered slightly more meatless meals once they approached the cashier. Default meal options can also indicate social norms. In one series of experiments, conference attendees overwhelmingly chose the vegetarian lunch option when it was designated as the default (Hansen, Schilling, & Malthesen 2021; see also Campbell-Arvai et al. 2014; de Vaan et al. 2019). Conference goers might have been just lazily going with the default option, but as the researchers argue it's plausible that many participants viewed the default as a "normative signal" about what others are expected to choose.

Ultimately, social proof can provide myriad forms of evidence. If most people in your community are shunning child marriage, that's evidence of confident consensus on the issue. If an increasing number of your neighbors are buying solar panels, that's evidence that the technology has progressed enough to be affordable and reliable enough to warrant the investment (Frank 2020). If more of your friends are eating delicious vegetarian meals and sticking to it, that's evidence that the diet is feasible. Again, it's likely that people also often follow the crowd or prestigious individuals in order to gain the benefits of acceptance or avoid being the odd one out. But the behavior of others also powerfully informs our own reasoning about what to do.

4. Reasoning to Moral Progress

Our set of examples suggests that reasoning not only shapes moral attitudes and behavior but can lead to moral progress. The notion of "moral progress" comes with a lot of baggage. For some philosophers, it is wedded to putatively implausible ideas – that moral truth is objective, that societies are endlessly improving, that the moral arc of the universe "bends toward justice."

However, there is a way of understanding moral progress that is far less contentious. First of all, moral progress simply means improvement of society from a (broadly) moral point of view (Buchanan & Powell 2018). The improvement may be only partial and it may also be judged by standards that are ultimately either objective or subjective (e.g., realist or constructivist). Furthermore, moral progress need not be defined in terms of any final goals, such as a perfectionist theory of the good life or a political theory about the ideally just state. Instead, we can simply rely on uncontroversial cases of moral progress (Kumar & Campbell 2022). A wide range of metaethicists and normative ethicists would agree, for example, about paradigm examples of moral progress, including (but not limited to) the decline of prejudice, the amelioration of global poverty, the reduction of animal suffering, and the growth of environmental stewardship.

How is moral progress of this sort to be achieved? A natural thought is that our institutions must change. Laws must guarantee equality on the basis of sexuality and gender identity; government regulations must curb carbon emissions and the torturous conditions on factory farms. If reasoning plays a role here, it might only seem crucial for think tanks and policy makers. However, the line between individual and structural change is blurry and the two work together in tandem (Brownstein, Madva, & Kelly 2022; Hermann 2019). Particularly in democratic societies, lasting changes in legislation and leadership occur only if individual citizens are motivated to demand and accept them (Brownstein, Madva, & Kelly 2022). Moral judgments matter. And if individuals are to make better moral judgments, reasoning is key at all levels of society.

Yet some commentators maintain that moral reasoning plays a minor role in moral progress. In discussing the end of dueling, slavery, footbinding, and similar moral revolutions, Kwame Anthony Appiah maintains that moral arguments against these practices were "well known and clearly made a good deal before they came to an end," and so people were not "bowled over by new moral arguments" (xii).

As we've emphasized, however, moral arguments are not merely reasons written on a page, about which one consciously deliberates. Moral reasoning is a partly conscious and partly unconscious process that unfolds over time—often painfully long periods of time—before it dislodges deeply held beliefs and practices (Kumar and Campbell 2016). Phelps-Roper spent years on Twitter debating with others before abandoning an ideology of hate. She considered counter-examples to her principles and inconsistencies among them, and she only took these ideas seriously when they came from people who were kind and patient. Moral arguments are more like avocado trees than pepper plants: they can take many years to bear fruit. But fruit they can bear, in the right conditions.

Other theorists call for more action than argument. Nigel Pleasants maintains that "moral argument is not enough by itself" to drive moral revolutions but "requires certain social, material, and practical conditions to become effective" (2010: 177). He claims, in particular, that a key factor in the abolition of slavery was the availability of an alternative economic system of wage labor (see also Anderson 2014; Hopster et al. 2025).

However, feasible alternatives are often part of the reasoning process. Arguments do not exist in a vacuum, or at any rate they are most effective in social contexts. As Elizabeth Anderson says in relation to social movements, "deliberation makes use of reasons and arguments to find better alternatives to entrenched customs" (2014: 12). Social proof, for example, often works by providing concrete, trusted examples of viable moral alternatives through overt action that goes beyond words. The moral case against factory farming is well-known, but it interacts with other considerations relevant to most people's choices about their diet, including health, cultural significance, and tastiness. The availability of plant-

based alternatives to meat and popular vegan restaurants removes objections to the moral arguments against consuming animal products from factory farms (May & Kumar 2022).

Moral arguments and reasoning about them are unlikely to drive moral progress if we caricature these as reasons written down and considered in isolation. No doubt "pure moral argument" is insufficient for social change, as Anderson puts it (2014: 9; see also Rehren & Blunden 2024). Yet moral reasoning is often much more social, unconscious, and protracted than the standard caricature allows. People debate issues and share stories on social media, at parties, at protests, and around the dinner table. Over time, the ideas percolate, a series of stories strengthen a premise once thought to be weak, and witnessing first hand alternative approaches undermines assumptions about the status quo being indispensable. Ultimately, changes in individual attitudes can yield systemic changes through support for new leaders and legislation. A broader more socially-embedded conception of reasoning dovetails with structural changes to yield durable moral progress.

5. Conclusion

Humans care about forming beliefs for reasons. It's not generally acceptable to report that you have your convictions "just because" or "for no reason." We've seen that this isn't just performative: moral attitudes and behavior can be shaped by various forms of reasoning. People apply ethical principles or revise them in light of counter-examples. They identify and resolve inconsistencies among their existing moral beliefs. And they look to the behavior of others as evidence of what's feasible or worthwhile. These are just some forms moral reasoning can take, not an exhaustive list, but they each exhibit the power of reasoning to drive moral judgment and decision-making.

In addition, the case against moral reasoning is weak. Moral reasoning is more than just conscious deliberation and is not primarily post hoc rationalization. Studies suggesting that incidental emotions drive moral judgment have been overblown. Although emotions do play a role in shaping our moral attitudes, they typically do so by integrating with our capacities for learning and inference. Recognizing the many forms that moral reasoning can take also shows that we shouldn't expect professional ethicists to be experts at all forms of it. They might be particularly skilled at consciously reasoning about moral principles, but that is just one form of reasoning. Ethicists are human too, so they are also susceptible to post hoc rationalization and forms of motivated reasoning that help justify existing moral beliefs, rather than changing them. In this light, it is not especially surprising that moral philosophers do not always behave better than other academics or the general public.

Of course, one's own values and interests can bias all forms of reasoning. Wishful thinking and rationalization can yield moral stagnation by concocting reasons to support one's existing moral beliefs. Moral regress is also a risk when arguments are developed and spread not because they are more defensible but because they support policies that benefit oneself or one's group. Even social proof can perpetuate the status quo, as most people follow the crowd or only gain inspiration from trendsetters in their own moral tribe.

Accordingly, moral reasoning is more likely to be progressive when we converse with people who share different ideas and experiences, particularly when we do so with mutual trust and respect. With an open mind, both conscious and intuitive reasoning can reveal new insights and inconsistencies that over time improve moral judgment. One's deepest moral beliefs might rarely change, but socially-embedded moral reasoning can promote the kind of intellectual humility that penetrates echo chambers, tempers extremism, and yields productive conversations.

Acknowledgements

For helpful feedback on this chapter, we thank David Copp, Joshua Greene, Hanno Sauer, Eric Schwitzgebel, and the students in Josh May's Moral Progress seminar. May's work on this project was made possible through the support of a grant from the John Templeton Foundation (Academic Cross-Training Fellowship 61581). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the foundation.

References

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

Barak-Corren, N., Tsay, C. J., Cushman, F., & Bazerman, M. H. (2018). If you're going to do wrong, at least do it right: Considering two moral dilemmas at the same time promotes moral consistency. *Management Science*, 64(4), 1528-1540.

Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and voluntary sector quarterly*, 40(5), 924-973.

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Brownstein, M., Kelly, D., & Madva, A. (2022). Individualism, Structuralism, and Climate Change. *Environmental Communication*, *16*(2), 269-288.

Buchanan, A., & Powell, R. (2018). *The evolution of moral progress: A biocultural theory*. Oxford University Press.

Buckland, L., Lindauer, M., Rodríguez-Arias, D., & Véliz, C. (2021). Testing the motivational strength of positive and negative duty arguments regarding global poverty. *Review of Philosophy and Psychology*, 1-19. DOI: 10.1007/s13164-021-00555-4

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.

Campbell-Arvai, V., Arvai, J., & Kalof, L. (2014). Motivating sustainable food choices: The role of nudges, value orientation, and information provision. *Environment and Behavior* 46(4): 453-475.

Charlesworth, T. E. S., & Banaji, M. R. (2019a). Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016. *Psychological Science*, *30*(2), 174–192.

Charlesworth, T. E. S., & Banaji, M. R. (2019b). Gender in Science, Technology, Engineering, and Mathematics: Issues, Causes, Solutions. *The Journal of Neuroscience*, *39*(37), 7228–7243.

Charlesworth, T.E.S., & Banaji, M.R. (in press). Patterns of Implicit and Explicit Attitudes II. Long-term change and stability, regardless of group membership.

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine*, 32(4), 556-577.

Cialdini, R. (2009). *Influence: The Psychology of Persuasion* (Revised Ed.). HarperCollins.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591-621.

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, *43*.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of learning and motivation* 50: 307-338.
Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. Oxford University Press.
Dwyer, S. (2009). Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment. *Mind & Language*, *24*(3), 274-296.
Feltz, A., & May, J. (2017). The means/side-effect distinction in moral cognition: A meta-analysis. *Cognition*, 166, 314-327.
Frank, R. H. (2020). *Under the Influence: Shaping the Social Forces That Guide Our Choices*. Princeton University Press.
Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, 35(3), 472-482.
Greene, J. D. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin.
Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, 124(4), 695-726.
Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of personality and social psychology*, *65*(4), 613.
Han, H., Workman, C. I., May, J., Scholtens, P. Dawson, K. J., Glenn, A. L., Meindl, P. (2022). Which Moral Exemplars Inspire Prosociality? *Philosophical Psychology* 35(7): 943-970.
Hansen, P. G., Schilling, M., & Malthesen, M. S. (2021). Nudging healthy and sustainable food choices: three randomized controlled field experiments using a vegetarian lunch-default as a normative signal. *Journal of Public Health*, 43(2), 392-397.
Harman, G., Mason, K., & Sinnott-Armstrong, W. 2010. "Moral Reasoning." In *The Moral Psychology Handbook*, ed. by J. M. Doris and the Moral Psychology Research Group, pp. 206–45. Oxford University Press.
Henrich, J. (2015). *The Secret of Our Success*. Princeton University Press.
Hermann, J. (2019). The dynamics of moral progress. *Ratio* 38, 139–12.
Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, 142(11), 1179.
Hopster, J. K. G., Arora, C., Blunden, C., Eriksen, C., Frank, L. E., Hermann, J. S., ... & Steinert, S. (2025). Pistols, pills, pork and ploughs: the structure of technomoral revolutions. *Inquiry* 68(2): 264-296.
Horne, Z., Powell, D., & Hummel, J. (2015). A single counterexample leads to moral belief revision. *Cognitive science*, 39(8), 1950-1964.
Horne, Z., & Livengood, J. (2017). Ordering effects, updating effects, and the specter of global skepticism. *Synthese*, *194*(4), 1189-1218.
Jacobson, D. (2014). Moral dumbfounding and moral stupefaction. In *Oxford Studies in Normative Ethics*, ed. by M. Timmons, pp. 289-315. Oxford University Press.
Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy*, 35(1), 54-76.
Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? *Ethical theory and moral practice*, 12(1), 77-96.
Kumar, V. (2017a). Foul behavior. *Philosophers Imprint*, 17: 1-17.
Kumar, V. (2017b). Moral vindications. *Cognition*, 176, 124-34.
Kumar, V., & Campbell, R. (2012). On the normative significance of experimental moral psychology. *Philosophical Psychology*, 25, 310-30.
Kumar, V., & Campbell, R. (2016). Honor and moral revolution. *Ethical Theory and Moral Practice*, 19, 147-59.
Kumar, V., & Campbell, R. (2022). *A Better Ape: The Evolution of the Moral Mind and How it Made us Human*. Oxford University Press.
Kumar, V., Kodipady, A., & Young, L. (2022). A psychological account of the unique decline in anti-gay attitudes. Under review.

Kumar, V. & May, J. (2019). "How to Debunk Moral Beliefs." *Methodology and Moral Philosophy*, eds. Jussi Suikkanen & Antti Kauppinen. Routledge.
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.
Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological bulletin*, 89(2), 308-324.
Levy, N. (2022). *Bad beliefs: Why they happen to good people*. Oxford University Press.
Lindauer, M., Mayorga, M., Greene, J. D., Slovic, P., Västfjäll, D., & Singer, P. (2020). Comparing the effect of rational and emotional appeals on donation behavior. *Judgment and Decision making*, 15(3).
Mallon, R., & Nichols, S. (2011). Dual processes and moral rules. *Emotion Review*, *3*(3), 284-285.
May, J. (2014). Does disgust influence moral judgment? *Australasian Journal of Philosophy*, *92*(1), 125-141.
May, J. (2018). *Regard for reason in the moral mind*. Oxford University Press.
May, J. & Kumar, V. (2018). Moral Reasoning and Emotion. *The Routledge Handbook of Moral Epistemology*, eds. K. Jones, M. Timmons, & A. Zimmerman. Routledge.
May, J. & Kumar, V. (2023). Harnessing Moral Psychology to Reduce Meat Consumption. *Journal of the American Philosophical Association*. DOI: 10.1017/apa.2022.2
MacAskill, W. (2015). *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing.
Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge University Press.
Mikhail, J. (2014). Any animal whatever? Harmful battery and its elements as building blocks of moral cognition. *Ethics*, 124(4), 750-786.
Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.
Nichols, S. (2021). *Rational rules: Towards a theory of moral learning*. Oxford University Press.
Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: a field experiment on collective norms and harassment behavior in a school social network. *Journal of personality and social psychology*, 103(6), 899-915.
Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive science*, 36(1), 163-177.
Prinz, J. (2007). *The emotional construction of morals*. Oxford University Press.
Railton, P. (2014). The affective dog and its rational tale: Intuition and attunement. *Ethics*, *124*(4), 813-859.
Rehren, P., & Blunden, C. (2024). Let's not get ahead of ourselves: we have no idea if moral reasoning causes moral progress. *Philosophical Explorations* 27(3): 351-369.
Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: unraveling the moral dumbfounding effect. *Judgment & Decision Making*, *10*(4).
Rust, J., & Schwitzgebel, E. (2014). The moral behavior of ethicists and the power of reason. H. Sarkissian and J. C. Wright (eds.) *Advances in experimental moral psychology*, pp. 191-208.
Sauer, H. (2017). *Moral judgments as educated intuitions*. MIT Press.
Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and social psychology bulletin*, *34*(8), 1096-1109.
Schönegger, P. & Wagner, J. (2019). The moral behavior of ethics professors: A replication-extension in German-speaking countries. *Philosophical Psychology*, 32, 532-559.
Schwitzgebel, E. (2019). Aiming for moral mediocrity. *Res Philosophica*, 96(3), 347-368.
Schwitzgebel, E., Cokelet, B., & Singer, P. (2020). Do ethics classes influence student behavior? Case study: Teaching the ethics of eating meat. *Cognition* 203: 104397.

Schwitzgebel, E., Cokelet, B., & Singer, P. (2023). Students eat less meat after studying meat ethics. *Review of philosophy and psychology*. DOI: 10.1007/s13164-021-00583-0

Schwitzgebel, E. & Cushman, F. (2020). Contest winner! A philosophical argument that effectively convinces research participants to donate to charity. Blog post at The Splintered Mind (May 23), URL: http://schwitzsplinters.blogspot.com/2020/06/contest-winner-philosophical-argument.html

Schwitzgebel, E., McVey, C., & May, J. (2022). Engaging Charitable Giving: The Motivational Force of Narrative versus Philosophical Argument. *Philosophical Psychology*. DOI: 10.1080/09515089.2022.2088340

Schwitzgebel, E., & Rust, J. (2009). The moral behaviour of ethicists: Peer opinion. *Mind*, 118(472), 1043-1059.

Schwitzgebel, E., & Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, 27(3), 293-327.

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143-153.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & public affairs* 1(3), 229-243.

Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological science*, 28(11), 1663-1674.

Stanley, M. L., Yin, S., & Sinnott-Armstrong, W. (2019). A reason-based explanation for moral dumbfounding. *Judgment and Decision Making*, 4(2), 120-129.

Summers, J. S. (2017). Rationalizing our way into moral Progress. *Ethical theory and moral practice*, *20*(1), 93-104.

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, 10(1), 181-211.

de Vaan, J. M., van Steen, T., & Müller, B. C. (2019). Meat on the menu? How the menu structure can stimulate vegetarian choices in restaurants. *Journal of Applied Social Psychology* 49(12): 755-766.