

[cambridge.org/bbs](https://www.cambridge.org/bbs)

Joshua May

Department of Philosophy, University of Alabama, Birmingham, Philosophy Department, Birmingham, AL 35294-1260.

joshmay@uab.edu<https://www.uab.edu/cas/philosophy/people/faculty-directory/josh-may>

Précis

Cite this article: May J. (2019) Précis of *Regard for Reason in the Moral Mind*. *Behavioral and Brain Sciences* **42**, e146: 1–60. doi:10.1017/S0140525X18002108

Précis accepted: 27 July 2018
Précis online: 7 August 2018
Commentaries accepted: 7 December 2018

Keywords:

debunking arguments; moral judgment; moral motivation; moral psychology; moral skepticism; rationalism; rationalization; sentimentalism; virtue

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 8) and an Author's Response (p. 46). See [bbsonline.org](https://www.bbsonline.org) for more information.

Abstract

Regard for Reason in the Moral Mind argues that a careful examination of the scientific literature reveals a foundational role for reasoning in moral thought and action. Grounding moral psychology in reason then paves the way for a defense of moral knowledge and virtue against a variety of empirical challenges, such as debunking arguments and situationist critiques. The book attempts to provide a corrective to current trends in moral psychology, which celebrate emotion over reason and generate pessimism about the psychological mechanisms underlying commonsense morality. Ultimately, there is rationality in ethics not just despite but in virtue of the neurobiological and evolutionary materials that shape moral cognition and motivation.

1. Optimistic rationalism

The past few decades have seen an explosion of scientific research on how we form our moral judgments and act on them (or fail to so act). What conclusions can we draw from all of the blood, sweat, and grant money?

If you ask most philosophers and scientists working both within and outside the field of moral psychology, you will likely hear something like the following. It turns out that Hume was right: Emotions are the star of the show, whereas reason (conceived as distinct from emotion) is a mere slave to the passions. Moreover, most people are lucky if they can squeeze some well-founded moral decisions out of their hominid brains, which are riddled with unconscious biases, swayed by arbitrary features of their circumstances, and constrained by antiquated heuristics that no longer track morally relevant factors.

This description of the received view is oversimplified, of course, but it is not far off. Jonathan Haidt (2003), for example, speaks of an “affect revolution,” which apparently explains the “rationalist delusion” (Haidt 2012) that reason plays a foundational role in moral cognition. Of course, such champions of sentimentalism do not themselves always conceive of this as pessimistic (compare to, e.g., D’Arms & Jacobson 2014; Nichols 2004), but it is easy to do so. After all, if reason merely serves the passions, morality is ultimately founded on non-rational or arational feelings. Indeed, some theorists explicitly track the evolutionary and psychological origins of moral psychology in order to raise doubts about the possibility of moral knowledge (e.g., Joyce 2006) or virtuous motivation (e.g., Batson 2016). Others allow reason the power to lead us toward moral progress, but the picture remains revisionary and pessimistic. Commonsense morality, we are told, must be jettisoned in favor of a counter-intuitive moral system, such as strict utilitarianism, which counsels us to always promote the greater good and implies that the ends always justify the means (e.g., Greene 2013; Singer 2005).

In *Regard for Reason in the Moral Mind* (May 2018), I suggest that this is all wrong. A careful examination of the science reveals that reasoning plays an integral role in ordinary moral thought and action. Moreover, this makes moral knowledge and proper moral motivation achievable without the need to substantially reject or revise our basic modes of moral deliberation, such as valuing more than the consequences of an action. Hence, I dub the view defended in the book *optimistic rationalism* and oppose it to a variety of philosophical theories, including sentimentalism, psychological egoism, Humeanism, and moral skepticism. Below I elaborate on some of the intricacies of my view and, importantly, summarize some of the main arguments for it that appear in the book.

First, a note on labels and the structure of the discussion. I divide up the moral mind into two key elements: moral cognition and moral motivation. For each element in turn, I consider, first, the primarily *empirical* questions about what drives them – for example, emotion, reason, arbitrary factors, and evolutionary pressures. Next, I examine *normative* questions about the status of each element – for example, are moral cognition and motivation deeply flawed, given how they work and what influences them? I generally use the honorific “moral knowledge” or at least “justified moral belief” to mark when moral cognition goes well. When moral

© Cambridge University Press 2019

CAMBRIDGE
UNIVERSITY PRESS

motivation goes well, I generally speak of “virtuous motivation” or “acting for the right reasons.”

A decidedly optimistic theme will emerge. Skeptical arguments require an empirical premise positing various influences on our moral minds, but the arguments also require a normative premise stating that these influences are morally irrelevant, arbitrary, extraneous, or otherwise problematic. I argue, however, that it is rather difficult to maintain both of these premises at once, at least when leveling wide-ranging critiques of our moral minds.

2. Moral cognition: Sources

2.1. Emotion

A multitude of studies seemingly suggest that emotions alone affect moral judgment, not merely because they can affect inference by, say, directing our attention. I start with reconsidering the popular studies (Ch. 2), before going on to adduce evidence of moral inference (Ch. 3).

There are three main lines of evidence in favor of sentimentalism, and most of the evidence focuses on the emotion of disgust. First, feelings may seem necessary for conceiving of a norm as distinctively moral rather than a mere convention. For example, the norm against sexual harassment at work seems a matter of ethics, whereas the norm against wearing pajamas to work a mere matter of social propriety. Shaun Nichols (2002) has argued that we treat moral norms as distinctive partly because we have strong feelings toward violations of them. However, there is far too much weight placed on the moral/conventional distinction as diagnostic of moral judgment. Even if people rate a norm as slightly more like a convention when they lack strong feelings toward it, that is not enough to demonstrate that norms are genuinely moral only if we have such feelings. Moreover, the relevant studies fail to manipulate emotions as a variable and are difficult to replicate in some circumstances (see, e.g., Royzman et al. 2009).

Second, sentimentalists have drawn on studies in which participants’ manipulated emotions seem to cause changes in moral judgment (see, e.g., Prinz 2007; Sinhababu 2017). Famously, for example, participants inhaling a foul smell apparently think incest is morally worse than do participants in a control group (Schnall et al. 2008). There are many reasons why such studies, although numerous, fail to support sentimentalism (May 2014). The main problem is that meta-analyses suggest the effects are tiny, perhaps even non-existent (Landy & Goodwin 2015). Both the control and manipulation groups, for example, tend to rate the morality of the target actions the same. The mean differences, when found, are miniscule. Statistically significant does not mean significantly different (in the ordinary sense of the word). Now, there is a burden on the rationalist to explain why incidental

emotions could ever have an effect on moral judgment, even if rare and ever so slight (Prinz 2016). But I provide an explanation (in Ch. 2) in terms of our well-known susceptibility toward mis-attributing the causes of our feelings (see, e.g., Schwarz & Clore 1983).

Finally, emotions may seem essential to moral judgment because dysfunction in “emotion areas” of the brain seem to lead to moral incompetence (see, e.g., Nichols 2004; Prinz 2007). Psychopathy is the prime example (although I also discuss so-called “acquired sociopathy” and frontotemporal dementia). Psychopaths are characteristically callous, manipulative, and deficient in guilt and compassion (Glenn & Raine 2014). Some studies suggest that people with psychopathic tendencies *somewhat* struggle to draw the moral/conventional distinction (see, e.g., Aharoni et al. 2012), but it is doubtful that this is enough to attribute significant deficits in moral judgment to them. Moreover, it is often underappreciated that psychopaths exhibit not only “emotional” deficits, but also clearly rational or inferential ones. Psychopaths are notoriously irrational, particularly imprudent, as a result of their poor attention spans, impulsivity, difficulties learning from punishment, trouble detecting emotions in others, and so on (see, e.g., Maibom 2005; Marsh & Blair 2008). I conclude that, although psychopaths likely exhibit some deficits in moral judgment, these should not be overstated (compared to their deficits in moral motivation) and that the moral ineptitude they do exhibit can be explained in terms of their deficits in reasoning.

Another problem with the appeal to psychopathology arises from a broader concern about the supposed reason/emotion dichotomy. Talk of “emotion areas” of the brain has become rather dubious in light of evidence that functionally diverse brain networks, extended over clusters of brain areas, give rise to emotions and other similarly complex psychological phenomena. Partly for this reason, emotional processing appears to involve a great deal of unconscious inference, involving the application of concepts, categories, and prior knowledge. So, for example, psychopaths suffer from dysfunction at least in the amygdala and ventromedial prefrontal cortex, but these areas are part of networks that facilitate not only emotion, but also unconscious learning and inference more generally (see, e.g., Woodward 2016).

The reason/emotion dichotomy begins to look rather spurious, as many philosophers and scientists are starting to recognize (see, e.g., Huebner 2015). But this does not mean the rationalism/sentimentalism debate is confused or pointless. What we are learning is that emotions involve a great deal of inference or, to put it the other way around, that inference is infused with affect (cf. Railton 2017). This realization roundly supports the rationalist view that feelings are not required for distinctively *moral* cognition. Rather, moral cognition is like other forms of cognition: it requires unconscious inference that is facilitated by feelings or affect. This does not sit well with the sentimentalist tradition, which maintains that moral judgment, with its need for emotions, is importantly different from other domains of cognition.

Moreover, the affect that underwrites inference is a mere twinge of feeling, not traditional moral emotions, such as guilt, indignation, and compassion. Although such emotions are undoubtedly a prominent character in the drama of moral life, it is often because they are the normal *consequences* of our moral beliefs. For example, people who are vegetarians for moral reasons are more likely to become disgusted by meat (cf. Rozin et al. 1997). Compassion is likewise modulated by prior moral judgments. For instance, people feel little compassion for

JOSHUA MAY is an Associate Professor of Philosophy at the University of Alabama at Birmingham. His work is primarily in ethics and epistemology with an emphasis on how empirical work informs philosophical debates. He has published articles in a number of venues, including the *Australasian Journal of Philosophy*, *Cognition*, *Journal of Medical Ethics*, *Neuroethics*, *Philosophical Studies*, and *Synthese*. In 2010, he received the Emerging Scholar Prize essay at the Spindel conference on Empathy and Ethics. Before taking his current position in Birmingham, he spent 2 years teaching at Monash University in Melbourne, Australia.

a student who missed classes because she left town with friends, but they readily sympathize with a student who missed classes because she was involved in a car accident (Betancourt 1990). Similarly, those of us who react so passionately to racism, misogyny, and mass shootings do so *because* we believe they are terribly wrong. And we believe these acts are terribly wrong because we reason – we recognize, we learn, we infer – that they involve egregious violations of norms, which prohibit intentionally or recklessly causing unwarranted harm, disrespect, and so forth.

2.2. Moral inference

Let us now turn briefly to the positive case for moral reasoning. The crucial move here (Ch. 3) is to recognize that reasoning can be, and often is, *unconscious*. We can stipulate that the term “reasoning” only refers to conscious reasoning, but that is overly restrictive and unhelpful (Arpaly 2003; Mallon & Nichols 2010). Indeed, one of the counter-intuitive lessons from decades of convergent results in experimental psychology is that much of one’s mental life is unconscious. That includes reasoning or *inference*, in which we form new beliefs on the basis of previous ones. For example, think about when you watch the opening scenes of a film – even a kids’ movie – which typically leaves important information implicit, such as the relationships among characters. Viewers are often left to infer what is going on, but it is not as though we consciously go through all the steps – “Ah, they look to be living in the same dwelling, yet they are in separate rooms and they both look sad, exhausted, and angry. Ergo, they must be in a romantic relationship and just had a fight!” Even if you could reconstruct something like this reasoning, it need not have been conscious at the time.

Moral cognition is no different. There is now a rather extensive scientific literature which reveals that intuitive moral judgments are driven by largely automatic and unconscious inferences, particularly about the consequences of the agent’s action and how involved the agent was in bringing them about. “Agential involvement” turns on well-known distinctions in moral philosophy between acts versus omissions, intentional versus accidental actions, and harming as a means versus as a side-effect. Much of this literature employs the infamous trolley cases, but many of the studies ask participants to make moral judgments about more realistic scenarios. Besides, these hypotheticals have been useful for probing automatic moral intuitions across the globe and revealing that they are shaped by a variety of unconscious inferences about how much harm the action caused, whether it was intentional, whether it was an action versus an omission, and so on (see, e.g., Barrett et al. 2016; Cushman et al. 2006; Young & Tsoi 2013).

Perhaps the most contentious corner of this literature involves the distinction between bringing about an outcome as a mere byproduct of one’s action as opposed to a means to one’s end goal. Some studies have failed to replicate early demonstrations of this means/byproduct effect, which is a core element of the old Doctrine of Double Effect. However, drawing on a recent meta-analysis of more than 100 studies involving more than 24,000 participants (Feltz & May 2017), I conclude that the means/byproduct effect is a real, even if small, aspect of agential involvement.

Moral inference is not always unconscious of course. I distance my account from extreme versions of the “linguistic analogy” or moral grammar hypothesis (Mikhail 2011), which posit an innate moral faculty that is highly modular and impervious to conscious

reasoning. I adopt an extremely minimalist dual process account (cf. Campbell & Kumar 2012), on which moral cognition can be generated by both slow, conscious thought *and* automatic, unconscious processes. But there is no sound empirical reason to cast either mode of moral thought as uniquely unreliable, driven by emotion, or even “utilitarian.”

Throughout the book, I attempt what might be an impossible task: remaining neutral on what emotions are exactly. An ecumenical approach is enough, however, to generate a problem for sentimentalists. Suppose I come to realize that my country ought to take in Syrian refugees, but only after watching a video of the crisis. The video generates intense compassion that focuses my attention on their suffering which previously I had not fully recognized. Such emotions are relevant only insofar as they contain or cause changes in patterns of inference, attention, recognition, and the like. So, if emotions contain cognitive elements, then they can *directly* shape moral cognition by, say, directing one’s attention and vividly highlighting morally relevant features of a situation. If emotions are mere feelings, lacking any cognitive elements, then they can only hope to shape moral cognition *indirectly* by changing patterns of inference. Either way, emotions can influence moral judgment in the way that they can influence any kind of judgment – by shaping inference through directing attention and so on. An unexpected mathematical claim, for example, can generate a feeling of surprise that directs my attention to new information and thus changes my inferences. Whatever emotions are exactly, they get a grip on moral cognition *via* reason and in a way that is not particular to distinctively *moral* cognition.

3. Moral cognition: Status update

How well is moral cognition doing, given what influences it? Recent debunkers contend that our moral beliefs are commonly driven by problematic emotions like disgust (e.g., Kelly 2011; Nussbaum 2004), framing effects (e.g., Schwitzgebel & Cushman 2012; Sunstein 2005), evolutionary pressures (e.g., Joyce 2006), and automatic emotional heuristics (e.g., Greene 2014; Singer 2005). All of these challenges are too ambitious for their own good, although more selective debunking arguments may succeed.

3.1. Defusing debunking arguments

Chapter 4 shows that wide-ranging skeptical arguments succumb to a Debunker’s Dilemma (Kumar & May 2018). Debunking arguments in ethics rely on an empirical and a normative premise (Kahane 2011; Nichols 2014):

1. Some of one’s moral beliefs are mainly based on a certain factor.
2. That factor is morally irrelevant.
3. So: The beliefs are unjustified.

But the two premises are difficult to jointly satisfy when one’s target is large, because moral cognition is influenced by a variety of factors and these factors are only problematic in some contexts.

Take disgust. Although *incidental* feelings of this emotion are surely morally irrelevant (good normative premise), we have seen they hardly affect moral beliefs, if at all (bad empirical premise). Now, *integral* feelings of repugnance can influence moral cognition. Disgust toward the actions of sexists and corrupt politicians,

for example, is typically tracking morally relevant information (cf. Kumar 2017). But a sound empirical premise is now joined with an awful normative premise: attuned emotions are not debunking.

Framing effects suffer the same fate. For example, the mere order in which information is presented is morally irrelevant (good normative premise), but meta-analyses (Demaree-Cotton 2016) suggest that the vast majority of moral beliefs are unaffected by mere differences in order (bad empirical premise). Some moral beliefs might be substantially changed by mere framing (e.g., Tversky & Kahneman 1981), but meta-analyses suggest these are outliers (Kühberger 1998), and wide-ranging critiques need trends.

What about Darwinian forces? Our moral beliefs are undoubtedly influenced by our evolutionary past. However, although mere evolutionary fitness is morally irrelevant (good normative premise), that is not a main basis for our moral views (bad empirical premise). The proximate causes of our particular moral judgments are values such as altruism, reciprocity, justice (which induces a desire for the punishment of norm violators), and so on. The ultimate explanation of these values may involve the fact that they were fitness-enhancing in the Pleistocene, but as proximate causes these values are morally relevant considerations. Evolutionary debunkers might deny that we can rely on any moral values to assess the normative premise, but that is self-defeating (Vavova 2015). If we cannot help ourselves to an independent evaluation of the normative premise in the debunking argument, then neither can the debunkers in defending it. Too often debunkers mistakenly think their task is merely to raise the possibility of moral error rather than demonstrate it empirically (see May 2013b).

(Note: Many evolutionary debunking arguments target moral realism, specifically the objectivity of morality, which is *not* my topic. My concern is moral epistemology. I remain neutral on whether moral beliefs, when true, are objectively true or whether ordinary moral judgments presuppose as much.)

Finally, let us briefly examine automatic emotional heuristics. Are our non-utilitarian commitments unwarranted because they are “sensitive to morally irrelevant things, such as the distinction between pushing with one’s hands and hitting a switch” (Greene 2013, p. 328)? That is a fine normative premise, but the corresponding empirical premise is untenable. As Greene acknowledges, experiments demonstrate that our moral intuitions are not particularly sensitive to pushing alone, but rather pushing that is done with intent or as a means to an end (Feltz & May 2017; Greene et al. 2009). Indeed, our non-utilitarian intuitions are generally sensitive to how involved the agent was in bringing about a bad outcome. Of course, utilitarians believe this is morally irrelevant, but that begs the question at issue in their debate with non-utilitarians. Greene (2014) also says our non-utilitarian intuitions are driven by rigid heuristics that are applied to moral problems with which the heuristics have “inadequate evolutionary, cultural, or personal experience” (p. 714). Again, a fine normative premise, but our best evidence reveals that moral intuitions are much more flexible, particularly during childhood, as they change over time in light of new information and recent cultural developments (see, e.g., Henrich 2015; Railton 2017).

3.2. Selective debunking and moral disagreement

Although wide-ranging empirical critiques of moral cognition are flawed, more selective debunking arguments can succeed (Ch. 5). For example, one might point to empirical research on disgust and cognitive biases to debunk certain attitudes toward homosexuality,

human cloning, and factory farming – particularly among a certain group of believers. There is not enough evidence at the moment, but the Debunker’s Dilemma is unlikely to be a barrier.

Another form of selective debunking appeals to consistency reasoning (Kumar & Campbell 2012). Empirical evidence can reveal that we maintain different verdicts about two similar moral issues for morally irrelevant reasons. It could turn out, for example, that most people believe that harming pets is morally objectionable, whereas factory farming is not, primarily because pets are cute. Similarly, although it is too wide-ranging to critique all non-utilitarian intuitions, we can all agree that it is morally irrelevant whether someone you can easily help is simply near or far away. Yet we could acquire rigorous empirical evidence that people tend to believe they lack an obligation to aid refugees in other countries primarily for this reason. Now, I am unsure that any of these particular debunking arguments would eventually succeed, at least for a sizeable group of believers. But the point is that empirical debunking can be done – if done properly, which will typically be selectively.

I take much more seriously a different form of empirical critique, which comes from moral disagreement. Philosophers have been extensively examining whether we really know something when it is disputed by “epistemic peers” – people one should regard as just as likely to be right about the topic (e.g., McGrath 2008). But there has been little examination of the relevant empirical premise of the corresponding skeptical argument (cf. Vavova 2014, p. 304):

1. In the face of peer disagreement about a claim, one does not know that claim.
2. There is a lot of peer disagreement about foundational moral claims.
3. So: We lack much moral knowledge.

Yet there is a wealth of empirical data on moral disagreements. To locate foundational disagreements, we might be tempted to go straight for cross-cultural research. However, it is more powerful to identify epistemic peers lurking within one’s own culture.

Here I draw on Haidt’s (2012) famous moral foundations theory. Within a culture, liberals and conservatives apparently disagree about the relative importance of five (or so) fundamental values:

Care/Harm
 Fairness/Cheating
 Loyalty/Betrayal
 Authority/Subversion
 Sanctity/Degradation

Does this provide support for the second premise in the skeptical argument from disagreement? Perhaps, but the critique will be – no surprise – limited. First, not everyone is an epistemic peer. But that is true only so far as it goes, and the empirical evidence does suggest that we should all be humbler about our cognitive abilities, especially on controversial topics in ethics. Second, and more importantly, disagreements about the foundations should not be overstated. Most people are not extreme liberals or conservatives, and as a result most people tend to recognize all five values. We just apply those values more to different topics (e.g., purity of the body vs. purity of the environment), depending on our other beliefs. Liberals and conservatives do apply different weightings to the five foundations, but among moderate liberals and conservatives the differences are a fairly small matter of degree (see Graham et al. 2013).

Ultimately, many people do probably lack moral knowledge as a result of peer disagreement. But this is restricted to particularly *controversial* moral issues. Many people do or should recognize that their most controversial moral beliefs are disputed by people who are just as likely to be right (or wrong for that matter). Here we may just have sufficient empirical evidence to challenge a selective set of moral beliefs, at least among the masses. Still, the overall picture of moral cognition is not pessimistic.

4. Moral motivation: Sources

Let us turn now from thought to action. Even when we know right from wrong, does empirical evidence show that we generally act for the wrong reasons?

4.1. Egoism versus altruism

One reason for action that often conflicts with morality is self-interest. You should return the lost bracelet or harbor the refugee, not because it comes with a financial reward or will enhance your reputation, but because it is kind, fair, or just the right thing to do. But chapter 6 asks: Can we ever ultimately act on anything other than self-interest?

Most philosophers think so, but scientists often treat such an egoistic theory as a live empirical possibility. Fortunately, there are decades of rigorous experiments that back up the philosophers. C. Daniel Batson (2011), in particular, has shown that empathizing with another in distress, and hence feeling compassion, tends to increase helping rates, and not because such helpers want to gain rewards or avoid punishment. Moreover, experiments reveal that infants and toddlers help others they perceive to be in need, even when helping is not expected and requires the children to cease engaging in a fun activity (see, e.g., Warneken 2013). We can, of course, always cook up an egoistic explanation of the data, but it begins to look strained and rather implausible.

One might argue that none of this amounts to ordinary altruism, because empathy causes one to blur the distinction between oneself and the other. One is either in a sense helping oneself (egoism) or not quite helping a distinct other (non-altruism). Some theorists have proposed exactly this sort of account and it has some affinity with traditions that actively encourage such self-other merging, as in the concepts of *no-self* in Buddhism and *agapeic love* in Christianity (see, e.g., Cialdini et al. 1997; Flanagan 2017; Johnston 2010).

The problem with these proposals is that they cannot make sense of the data. The empirical support for a self-other merging account is flawed, but more importantly there is a conceptual problem (May 2011). When one helps another, there is a first-personal mode of presentation required to navigate the distinct bodies (cf. Perry 1979). I cannot, for example, actively help another person while conceiving of the two of us as merely *them* (third-personal). I need to know which arms and legs *I* must move to save her. Even *us* smuggles in a first-personal reference to a self. So we ought to treat empathy as inducing a concern for others represented as distinct from oneself. Ordinary altruism is thus possible and even prevalent, given that empathy, and the compassion it engenders, are not uncommon.

4.2. Rationalization and moral integrity

Pessimists might accept the existence of genuine altruism but argue on empirical grounds that it is limited, restricted primarily to our kith and kin. When we interact with acquaintances or

strangers, we might be primarily motivated by self-interest or otherwise the wrong reasons. However, chapter 7 covers ample evidence that people are quite frequently motivated by their moral beliefs. Oddly enough, the evidence comes from studies of bad behavior, particularly when we succumb to temptation.

Consider, for instance, the phenomenon of *moral licensing*. After doing something virtuous or affirming one's own good deeds, one sometimes justifies bending the rules a bit. For example, one study found that participants were a bit less honest and generous after conspicuously supporting environmentally friendly products, compared to a control group (Mazar & Zhong 2010). There are many studies of such moral licensing (Blanken et al. 2015), and they are just one form of the familiar phenomenon of motivated reasoning or, more generally, rationalization (Kunda 1990).

We can also look to studies in which one will fudge the results of a fair coin flip in order to steer a benefit toward oneself (Batson et al. 1997). Importantly, participants in such studies tend to rate their actions as morally acceptable, just because there is a sense in which they did use a fair procedure (flipping the coin), despite fudging the results in their favor. Flipping the coin provides just enough wiggle room for many people to rationalize disobeying the results. Such bad behavior is motivated not merely by self-interest but by a concern to act in ways one can justify to oneself – that is, by one's moral beliefs (or, more broadly, normative beliefs).

Notice that this is not just rationalization of one's bad choice after it happens (*post hoc*), but rationalization before the action in order to justify performing it (what I call "*ante hoc* rationalization"). This should be recognizable in one's own life. People do not just behave badly because it is in their interest. When they could just think (probably unconsciously) "I am going to keep this lost \$20, because I want the money," they instead think something like: "I probably need the money more than the owner," or "I have done more than my fair share of good deeds this week," or even "I bet it is that sleazy banker's money, and he has got plenty." These are thoughts that could potentially justify one's behavior, even if the reasoning is addled. (Indeed, even atrocities are rationalized, unfortunately.) After the rationalizing is done, one does not necessarily see oneself as doing anything morally objectionable (cf. Holton 2009). One's actions are in line with one's moral beliefs – at least temporarily, since later on one may be cool, calm, and collected or otherwise see matters aright, at which point guilt sets in.

What these various studies reveal is the motivational power of moral beliefs. Our focus has been on bad behavior because the relevant studies concern temptation. But there is no reason to think that moral beliefs play any less of a role in motivating *good* behavior. We are normative creatures, most of whom care deeply about acting reasonably and justifiably, whether we end up doing what is right or wrong. We care ultimately about acting in particular ways, such as being fair, which we regard as right, but we also ultimately care about doing what is right as such.

When we do what is right, then, we are not ultimately motivated by self-interest alone but by considerations we deem to be morally relevant or genuine reasons, such as considerations of fairness, justice, benevolence, loyalty, honor, and even abstractly "the good" and "the right." Like Hurka (2014), I adopt a pluralistic approach on which all these sorts of considerations are the right kinds of reasons or concerns (whether they are construed, to use some philosophical jargon, "*de dicto*" or "*de re*"). I adopt some terminology from Batson (2016) and call any such concerns to do what is

right *moral integrity*. This is a third intrinsic concern that we should add to human psychology – in addition to ultimately caring about one’s own self-interest (egoism) and the well-being of others (altruism). Indeed, moral integrity is plausibly related to the trait of “moral identity,” which varies in the population, and can be enhanced or suppressed (Aquino & Reed 2002).

4.3. The autonomy of reason

At this point, a theorist inspired by David Hume might argue that our moral beliefs, even if products of reason, are ultimately under the direction of desire (e.g., Arpaly & Schroeder 2014). Suppose, for example, that while on the bus you offer your seat to an elderly man standing in the aisle. A Humean might argue that you are only ultimately motivated to act because you happen to care about being respectful or about doing what is right. Do we have empirical reasons to *always* posit such antecedent desires which our moral beliefs serve? Does the scientific evidence show that reason is always a “slave to the passions”? These questions are taken up in chapter 8.

We certainly sometimes do what we believe is right because we are antecedently motivated to do what is right as such (“de dicto”) or to promote particular moral values, such as kindness, respect, and fairness (“de re”). But this need not always be the case. On a sophisticated anti-Humean view (May 2013a), one is capable of being motivated to do something simply because one believes it is the right thing to do, even if one has a weak or non-existent desire to do it or to be moral. For example, someone who engages in discriminatory behavior can be motivated to stop simply by coming to believe it is the right thing to do, even with no changes to his antecedent goals or motives. This anti-Humean picture is, despite appearances, entirely compatible with the science.

Take neurological disorders, which some Humeans have used to support their view. Here I will just mention the example of damage to the ventromedial prefrontal cortex. Patients with such damage who develop so-called “acquired sociopathy” tend to have difficulty making appropriate social, moral, or prudential choices. These patients seem to retain knowledge of how to act but struggle to translate their general normative judgments into a decision and action in the moment (Damasio 1994/2005). Some philosophers believe these patients support the Humean thesis that moral (or otherwise normative) beliefs cannot motivate by themselves (cf. Roskies 2003; Schroeder et al. 2010).

But that is based on a misunderstanding of the opposition. Of course one’s moral beliefs do not *always* generate the corresponding desire, but when they do they need not rely on an *antecedent* desire. Instead, the necessary element could be, say, a lack of full understanding – for example, a patient believes she ought to thank the host, but she does not entirely appreciate that she is in the relevant circumstances (cf. Kennett & Fine 2008). Or it could be that the relevant brain dysfunction disrupts her virtuous dispositions to be motivated to do what she knows she ought to do. Indeed, far from being incompatible with anti-Humeanism, acquired sociopathy reveals that normally our moral beliefs do motivate but that this can break down in cases of pathology.

Other empirically minded Humeans contend that desires are ultimately necessary for all motivation because they provide the simplest explanation of action (e.g., Sinhababu 2017). In particular, desires are goal-directed states that are inherently motivational, direct one’s attention to their objects, and cause pleasure when one anticipates satisfying them. These are characteristic features of desire that arise out of our intuitive folk psychology but also neuroscience, particularly our understanding of the brain’s

reward system (Schroeder 2004). Thus, it may seem that moral beliefs cannot play this same role without either being desires or being an unnecessary additional posit in psychology.

However, I show that desires do not have a monopoly on these psychological properties. Indeed, the reward system provides a framework for understanding any mental state that treats an event as positive or “rewarding.” In this way, moral beliefs (e.g., “Smoking near children is wrong”) share much in common with desires (e.g., wanting to smoke away from children), compared to merely descriptive beliefs (e.g., “Secondhand smoke causes cancer”). Both moral beliefs and desires treat a state of affairs as valenced – as good/bad or desirable/undesirable. The two states are importantly different, however, in that only beliefs are assessable for truth and thus suited to playing an integral role in reasoning – the forming of new beliefs on the basis of previous ones. However, when a belief does contain normative content, it represents its object in a positive light and thus typically generates some desire for it.

Consider how anti-Humeanism nicely explains a particular example. Suppose your friend goes on a meditation retreat and comes to realize that he is kind of a jerk. In conversations with others, he tends to boast, redirect the conversation toward himself, and rarely ask about his interlocutor’s problems or concerns. (Or imagine another moral failing you or a friend struggle to correct.) On the anti-Humean view, we can explain this kind of scenario in terms of two independent sources of intrinsic motivation. The jerk has an egoistic desire to feel good about himself and discuss his own problems. But he also believes it is important to be a good person, and recently has become thoroughly convinced that he has some relevant character flaws here. This moral conviction or belief – indeed, knowledge – generates a new desire in him to correct his behavior. Now, the Humean would insist on positing an antecedent desire to be moral, which this new belief serves, but I argue that we do not have any empirical reason to *always* do so. Reason is not destined to be a slave to the passions.

5. Moral motivation: Status update

So far, in the book, I argue that our moral beliefs are not hopelessly off-track and that these beliefs frequently drive behavior, through processes like rationalization. These are largely empirical questions, but in chapter 9 we ask again about normative status: Are we motivated by the right reasons? Much like attempts to debunk moral beliefs, one might try to debunk or “defeat” moral motivation using arguments of the following sort, which combine an empirical premise with a normative one to generate a normative conclusion:

1. Some of one’s morally relevant behaviors are mainly based on a certain factor.
2. That factor is morally irrelevant.
3. So: The behaviors are not appropriately motivated.

Here I speak of attempts to “defeat” moral motivation in order to connect my discussion with others (particularly, Doris 2015). The proposed morally irrelevant factors might be fleeting features of the situation (see, e.g., Doris 2015; Nelkin 2005; Vargas 2013b) or stable forms of self-interest (see, e.g., Batson 2016). However, as with debunking arguments, there is a formidable dilemma – the Defeater Dilemma – that afflicts any wide-ranging attempts to undermine virtuous motivation. Skeptics can often find support for one premise in their argument, but at the cost of failing to support the other premise. There is again a kind of trade-off or tension between the two.

5.1. Self-interest returns

Despite the existence of genuine altruism, we may too often rationalize serving self-interest, perhaps in a self-deceived manner. Even when we do what is right, we may often do so because unconsciously we ultimately want to curry favor or avoid being socially ostracized. Although the science does demonstrate that we can be ultimately motivated by more than self-interest, there is some evidence that this is less common than we would like to admit (Batson 2016). Virtuous motivation is threatened given that acting from self-interest is often the wrong kind of reason to do what is right.

Consider again studies of fairness. In some experiments, many participants only *appear* to be fair, by flipping a coin to determine who gets a reward, yet around 90% of the time the flip magically favors the participant. Clearly, there is some fiddling of the flip going on. Follow-up studies suggest the fiddlers do not just misremember whether they chose heads or tails. Instead, they are primarily motivated to avoid seeing themselves as immoral (Batson et al. 2002). In fact, fiddlers rate their behavior as moral, unlike those who do not flip at all.

Batson interprets this as “moral hypocrisy,” which he regards as a kind of egoistic motivation to look good to oneself. However, although seeking to appear moral to *others* is clearly egoistic, being ultimately motivated to look moral to *oneself* is just a concern to be moral. This is moral integrity even though, as with other forms of motivated reasoning, one’s conception of morally good behavior is corrupted at the time of temptation. Moreover, only some participants fiddled the flip, and only when there was enough “wiggle room” that they could justify flipping the coin but then ignore the results. We hardly have evidence for the cynical conclusion that our moral choices are dominated by self-interest alone without a concern to be moral.

Similar issues arise with studies of dishonesty. When participants can get away with it, many will lie about how many arithmetic puzzles they solved, in order to earn more money from the experimenters (see, e.g., Mazar et al. 2008). Interestingly, dishonesty is mitigated significantly, often nearly eliminated, when participants are reminded of moral standards (see Ariely 2012). In one case, for example, participants were first asked to write down as many of the Ten Commandments as they could recall. In another study, participants had to sign an honor code before they took a crack at the puzzles. Both interventions significantly reduced cheating.

Once again, we might be led to think that, when moral choices are available, egoism is rampant. However, as Ariely makes clear, the vast majority of people only cheat a little by claiming to have solved about 10% more of the puzzles than they did, and this dishonesty can be mitigated with moral reminders. Indeed, whether or not people cheat, the mechanism appears to be rationalization. One rationalizes cheating a little, for that is all one can justify to oneself. Some even rationalize not cheating at all by having one’s attention drawn to one’s considered moral beliefs. Either way, the proper motivation appears to be in play: People are primarily motivated by a concern to act in ways they can justify to themselves as morally acceptable, not merely by self-interest. If their ultimate concern were self-interest alone, they would not have worried themselves about the morality of their choices.

The Defeater Dilemma is evident here. It is a plausible normative premise that acting from self-interest is often the wrong kind of reason to act. But a careful look at the evidence suggests instead that people are quite motivated to be moral, and the corresponding normative premise is thereby implausible. We are not motivated by the wrong reasons if we are motivated to do what is

right. There is no doubt that we are motivated by egoism as well, but we should not overstate its power and prevalence, and likewise we should not ignore the power and prevalence of moral integrity (even when it is a result of motivated reasoning). The same dilemma arises for the challenge from situationism.

5.2. Situational forces

Countless studies support the situationist thesis that we are often unconsciously motivated by surprising features of our circumstances, at least more often than we intuitively expect. In one study, for instance, about twice as many participants at a mall helped someone make change for a dollar when in front of a bakery or a coffee roasting company, compared to participants who had the opportunity to help in front of a store that was not emitting such pleasing aromas (Baron 1997). Similarly, participants are much less likely to help someone apparently in need of serious help if there are other people nearby who are not helping (Latané & Nida 1981). And, infamously, people make decisions about who to hire and even who to shoot, based partly on implicit biases against the person’s race, gender, and other social categories (see, e.g., Bertrand & Mullainathan 2004; Payne 2001). These are just a few examples of the relevant sorts of studies. Some may not survive the replication crisis, but enough will likely remain to suggest that people can be influenced unconsciously by features of their circumstances.

Many philosophers and scientists have taken such results to threaten the existence of traditional character traits (cf. Alfano 2013) or of certain conceptions of free will and moral responsibility (e.g., Vargas 2013b). However, the more fundamental worry is that we are not motivated by the right reasons: Did I act primarily to help the person in need or because the pleasing smell of cookies put me in a good mood? As Dana Nelkin (2005) has put it when discussing the threat to free will: “the experiments challenge the idea that we can control our actions on the basis of good reasons” (p. 204). Thus, the situationist literature might seem to fund a wide-ranging critique of what motivates moral behavior.

The Defeater Dilemma remains an obstacle, however. Some situational forces do substantially influence morally relevant behavior, thus grounding a strong empirical premise in the skeptical argument. However, then the normative premise suffers. For example, meta-analyses suggest that circumstantial changes in mood do significantly impact helping behavior (e.g., Carlson et al. 1988), but the vast majority of studies concern acts that are morally optional or supererogatory. There, I argue, mood is a morally relevant consideration: Your mood is an appropriate consideration, among others, when deciding whether to help a stranger make change for a dollar, pick up some papers someone dropped in a mall, and so on. If helping is morally optional, then *whether you feel like helping* is a relevant consideration. Perhaps it is inappropriate to only help because you feel like it, but it is a relevant consideration that may tip the scales in favor of acting.

Other studies do have confederates who appear to be in serious need. It is not morally optional to help someone who, for example, appears to have fallen off a ladder. But here, too, the effects are driven by morally relevant considerations. In-depth studies of group effects suggest that most participants do not help in the presence of bystanders because participants firmly believe that no help is really needed (cf. Latané & Nida 1981; Miller 2013). Such a belief is unwarranted, but what it concerns is morally relevant.

The same cannot be said of other factors, such as implicit racial biases and genuine framing effects, which are clearly morally irrelevant. Here we have a plausible normative premise for the

skeptical argument, but its corresponding empirical premise becomes untenable. Our moral decisions are sometimes partly determined by the mere order in which information is presented. But again meta-analyses suggest that the vast majority of moral decisions remain the same in the face of genuine framing effects (Demaree-Cotton 2016). Although some framing effects produce dramatic results, these are outliers (Kühberger 1998).

Similarly, although implicit biases no doubt exist, recent meta-analyses suggest that their effects are quite small and do not predict much behavior (Forscher et al. 2017; Greenwald et al. 2009; Oswald et al. 2013;). Importantly, and I cannot stress this enough, that does not mean implicit biases cannot explain large-scale problems in society. Indeed, implicit biases may add up to explain the powerful discrimination any one individual experiences as a result of slights from many people, however well-meaning these people are. But the evidence to date does not suggest that *most* ordinary people base *many* of their morally relevant decisions *primarily* on their implicit biases. Some do, for sure, but we are looking for trends in the data that can fund wide-ranging critiques. When a police officer does the right thing and decides not to shoot an unarmed teenager who is brandishing a toy gun, it is probably not primarily because the suspect is white – although that may play a minor role. At other times, when the child is black, the minor role race can play might sadly be just enough to yield a pulled trigger and a lost life, particularly in a high-pressure situation when a split-second decision is made (which is precisely when most implicit biases show up in the lab). But, again, our inquiry concerns a *main* basis for *most* people's moral and immoral behaviors.

The foregoing is just a sampling of the situationist literature, but you get the idea. When targeting a wide range of morally relevant behaviors, it is difficult to identify a single influence that is morally inappropriate in all or nearly all contexts. Our moral decisions are based on many factors, only some of which are a main basis for any one choice. Moreover, a single influence can be inappropriate in some contexts but appropriate in others. Mood is sometimes an appropriate consideration when deciding when to help, but not when the situation is dire. Even race can be a morally relevant consideration in some contexts (e.g., when justifying certain affirmative action policies). Thus, rather than picking apart a few studies among many, I aim for the Defeater Dilemma to provide a principled and systematic way to resist challenges to virtuous motivation from situationism and related literatures.

6. Conclusion: Cautious optimism

If I am right, moral psychology is in an important sense continuous with other domains of human psychology. The heart of the rationalist view is that morality is not special; emotions are not essential to moral psychology in a way that is fundamentally different from how our minds grapple with prudence, social interactions, or even economics.

Now, the book in effect assumes that reason in general, as applied to any particular domain, is not deeply flawed. A full defense of optimistic rationalism would require responding to challenges to reason itself. But that is for another day. *Regard for Reason in the Moral Mind* already discusses a wide range of literature in just 10 chapters. It certainly has not settled these important issues in moral psychology and metaethics. I only hope to have carved out a reasonable alternative to the present orthodoxy. In light of the science, a rationalist view of moral psychology is defensible and, partly because of this, various skeptical challenges can be answered or defused.

The key is to examine the science critically and avoid caricatures of reason. Reasoning is often unconscious and flawed insofar

as it is influenced by motives unrelated to truth, which gives rise to rationalization (not just *post hoc* but *ante hoc*). Sometimes these bouts of rationalization are corrupted and bad behavior results. But just as often reasoning leads to virtuous action, typically through unconscious processes of inference, recognition, and learning. Of course, we care deeply about morality, so emotional reactions abound. But emotions are often the natural consequences, not causes, of our moral convictions. The distinction between reason and emotion is admittedly blurry, as gut feelings seem to underlie reasoning both in ethics and non-moral domains. However, although subtle affect may guide reasoning about moral matters, classic moral emotions such as compassion and shame are commonly a consequence of such reasoning.

The picture of moral psychology that has emerged has implications for how to enhance moral knowledge and virtue (Ch. 10). It is common now for scientists, philosophers, and even politicians to call for more emotional responses, such as compassion, disgust, and anger. But it should be clear that indiscriminately amplifying such emotions by themselves is not the best way to effect proper moral change. Our emotional reactions depend heavily on our prior moral beliefs, so it would be a disaster to get people to feel, say, more compassion without changing their patterns of inference and their conceptualization of situations. It is not just that empathy tends to be biased and parochial (Bloom 2016); people of different moral persuasions, such as liberals and conservatives, have different views about who deserves it.

For those with the right moral views, how do we get them to behave accordingly? This may require enhancing whatever motivation to be moral they already have (that is, moral integrity), but that will only go so far. The greatest barrier to good behavior is likely motivated reasoning and other cognitive biases. Perhaps we can nudge each other toward ethical conduct by structuring our environments with moral reminders and other technologies that help us avoid rationalizing bad behavior. Whatever the interventions, they will probably be most effective in childhood and focus on the full development of rational capacities, including understanding, learning, recognition, inference, focus, and humility.

Regard for Reason in the Moral Mind is meant to generate discussion among researchers working on different aspects of moral psychology. Despite there being rather distinct literatures on moral cognition and moral motivation, for example, the two are intimately connected and there is value in discussing them together. Indeed, skeptical challenges to both are structurally similar, as are the best available replies. A broad, systematic examination of our moral minds may be the best treatment for empirical pessimism.

Open Peer Commentary

Moral reasoning is the process of asking moral questions and answering them

Mark Alfano^{a,b} 

^aEthics and Philosophy of Technology, Delft University of Technology, 2628 BX Delft, The Netherlands; and ^bInstitute for Religion and Critical Inquiry, Australian Catholic University, Fitzroy, VIC 3002, Australia.

Mark.Alfano@gmail.com www.alfanophilosophy.com

doi:10.1017/S0140525X18002534, e147

Abstract

Reasoning is the iterative, path-dependent process of asking questions and answering them. Moral reasoning is a species of such reasoning, so it is a matter of asking and answering moral questions, which requires both creativity and curiosity. As such, interventions and practices that help people ask more and better moral questions promise to improve moral reasoning.

The new irrationalists would have you believe that “moral reasoning is really just a servant masquerading as the high priest” in the “temple of morality,” whereas “the emotions” in fact wield all the power (Haidt 2003, p. 852; see also Prinz 2016). Reasoning, they tell us, rarely plays a role in the formation of moral judgments, though it may sometimes be pressed into service by affects and emotions to provide “post hoc” justifications for judgments that the agent would have made anyway (Haidt 2001, p. 814). In *Regard for Reason in the Moral Mind*, May (2018) contends, to the contrary, that reasoning is causally efficacious and prevalent, though of course also prone to various errors and biases. Settling this debate requires us to establish a tentative characterization of reason or reasoning. In this commentary, I show that a promising new theory of reasoning – the erotetic theory – corroborates May’s position. Indeed, May concedes too much to the new irrationalists, who rely on an empirical base that does not replicate and is inconsistent with the erotetic theory.

According to May’s initial characterization (p. 8, sect. 1.2.2, para. 3), reasoning is “a kind of inference in which beliefs or similar propositional attitudes are formed on the basis of pre-existing ones.” This is a helpful start, in that it focuses on the process of reasoning rather than an alleged faculty of reason. Much mischief has been done by positing a reified and rarefied faculty of reason, and even more damage has been done (e.g., by Kant in the second *Critique*) by positing a domain-specific faculty of moral or practical reason. Moral reasoning is just a species of domain-general reasoning; it is reasoning that matters morally or is about moral matters (Cushman 2013).

Furthermore, the process of reasoning is so unmythical as to appear mundane. A reason is simply a consideration that counts in favor of adopting an attitude or a course of action. For example, a doxastic reason is a consideration that counts in favor of adopting a particular belief, a desiderative reason is a consideration that counts in favor of adopting a particular desire, and a practical reason is a consideration that counts in favor of performing or omitting a particular action. The process of reasoning is then a matter of being sensitive to such considerations when they are relevant and putting them together in sensible ways.

Two of the main arguments offered by the new irrationalists for their pessimistic conclusions rely on the premises that reasoning is necessarily conscious and that it is contrary to or at least in tension with emotion. They then aim to show that unconscious or emotional states and processes influence or guide many moral judgments. May opposes both of these propositions, and for good reason. First, a mental state or process is conscious when one is aware of oneself as embodying it (Rosenthal 2005). Someone of course *could* be aware of themselves as going through the process of reasoning (i.e., could be aware of themselves as asking and answering questions), but such self-awareness is not necessary or guaranteed. May (Ch. 3) is therefore right to argue that

reasoning need not be conscious, which disarms one of the central arguments offered by the new irrationalists.

Second, emotions implement and up-regulate people’s sensitivities to a range of associated considerations. For example, fear makes one more sensitive to threats and dangers (Brady 2013; Tappolet 2010), whereas disgust makes one more sensitive to impurities and corruption. Emotions thus play a central role in reasoning: alerting us to considerations that count in favor of adopting attitudes or courses of action. Emotions are one of the main ways in which reasoning – including moral reasoning – is implemented (Alfano 2016). May (Ch. 2) is therefore also right to argue that emotions should not be seen as irrational or arational disruptors of reasoning, which disarms another of the central arguments offered by the new irrationalists.

This is not to say that people’s emotional reactions are always appropriate or somehow infallible. Instead, emotions tend to make people especially sensitive to some reasons and less sensitive to others. As such, an emotional reaction can make someone oversensitive to some considerations and undersensitive to others. However, even in the case of disgust, one of the poster boys of the new irrationalists, such oversensitivity has at most a small impact on moral judgments (Landy & Goodwin 2015). Studies purporting to show large spillover effects from incidental affect or emotion to moral judgments do not replicate. But even if the influence of disgust or some other emotion on moral judgments makes us worry, that does not mean we should attempt to exorcise our sentiments. Just as inquiry is likely to go awry if one only asks a single question, so moral reasoning is likely to go awry if one only weighs reasons prompted by a single emotion. We are not forced to conclude, though, that inquirers should never ask questions, or that moral reasoners should not allow their emotions to provide inputs to their reasoning processes. Instead, we need to ask more and more diverse questions in our inquiries, and we need to experience more and more diverse emotions in our moral reasoning. Reasoning works best not when we wall ourselves off from our emotions but when we cycle through a range of them, letting each make its contribution before coming to an all-things-considered judgment or decision (Alfano 2017).

This analogy between inquiry and reasoning is instructive. Emotions prompt us to ask questions about the normative properties they are associated with. Fear leads us to ask, “Where is the danger”? Disgust leads us to ask, “Where is the corruption”? Anger leads us to ask, “Where is the insult or offense”? These are essential first steps in thinking through whether a moral wrong has been committed. According to the erotetic theory of reasoning, “reasoning proceeds by treating successive premises as questions and maximally strong answers to them,” and “systematically asking a certain type of question as we interpret each new premise allows us to reason in a classically valid way” (Koralus & Mascarenhas 2013, p. 318). To ask a question is to pose a set of mutually non-compossible options and attempt to settle on one of them. These options may exhaust the logical space, but in many cases they do not. If people ask enough and the right questions, their reasoning processes will be valid. However, if they ask the wrong questions or too few questions, they systematically fall into the errors and illusory inferences documented by cognitive scientists (e.g., Johnson-Laird 2008; Khemlani et al. 2012; Rips 1994; Walsh & Johnson-Laird 2004).

In the case of moral reasoning, asking enough and the right questions is typically prompted by cycling through a range of emotions. Poor reasoning – including poor moral reasoning prompted by a cramped emotional set – thus derives in many cases from a failure to express the intellectual virtues of creativity

(Koralus & Mascarenhas 2013, p. 324) and curiosity (Koralus & Alfano 2017, pp. 92–94). In my paper with Koralus, we showed that some of the same systematic patterns of error and bias crop up in untutored moral reasoning that have already been documented in untutored non-moral reasoning (Shafir 1993). This suggests that moral reasoning is of a piece with the rest of reasoning, and that the dispositions that foster good non-moral reasoning should also foster good moral reasoning. Furthermore, people exhibit the aptitude and skill associated with such reasoning to different degrees. Individual difference measures of both creativity (Silvia et al. 2012) and curiosity (Iurino et al. 2018) have recently been validated, and these may turn out to be useful covariates in the study of moral reasoning. In addition, we may reasonably hope that it is possible to acquire and cultivate these dispositions over time: Just as people can learn and be taught to ask more and better questions (Watson 2018), so they can learn and be taught to wield their emotional sensitivities in the service of creative and curious moral reasoning. If this is right, then May should take comfort not only in the fact that people engage in moral inference, but also in the facts that, spurred by their emotions, they engage in the corrigible activity of asking and answering of moral questions. There is much work to be done in establishing how people go about asking moral question and under what conditions they best answer them, but we no longer need to quaver before the new irrationalists.

Emotions in the development of moral norms within cooperative relationships

Jeremy I. M. Carpendale and Beau Wallbridge

Simon Fraser University, Burnaby, BC, Canada, V5A 1S6.
jcarpend@sfu.ca beau_wallbridge@sfu.ca
<http://www.psyc.sfu.ca>

doi:10.1017/S0140525X18002571, e148

Abstract

We support May's criticism of attempts to reduce morality to being primarily based on evolved emotional reactions. However, we question the clarity and consistence of his own position and suggest taking a developmental approach. We focus on providing a developmental approach to the role of emotions in the social origin of moral norms.

We applaud May's (2018) criticism of recent attempts to reduce morality to being primarily based on evolved emotional reactions, a view often inspired by Hume. We agree with May's skepticism regarding the emotions/reasoning dichotomy, and we support his move toward a more complex view of emotions and reasoning to avoid reifying emotions and reasoning as separate processes. May makes an important contribution by providing a careful review of a great deal of literature, but we encourage him to go further by taking a developmental approach, and we begin by examining possible contradictions embedded in his work. We argue that May's goal of increasing moral knowledge requires considering

the source of moral norms. We then offer an alternative approach to the typical nativist and relativist accounts of moral norms, based on the approaches of Jean Piaget and G.H. Mead.

Although May convincingly argues against sentimentalist approaches to moral psychology, his own approach is not clear. He takes for granted positions we consider problematic such as the computational theory of mind, and he also states that it is possible that humans might have evolved an innate moral faculty. On the other hand, he assumes that moral norms are relative to a culture (p. 17), and so are not objectively true. Thus, he seems to implicitly accept moral relativism because objectivism and relativism are generally considered the only two options, and he has not argued for a third alternative. However, he also seems to accept the possibility of some universal aspects of moral intuitions being "in some sense innate" (p. 102). This is too vague; in what sense innate? It has been pointed out that the multiple uses of the term "innate" are all problematic (Mameli & Bateson 2006). What is usually implied is that information is encoded in a genetic program, a highly problematic position because genes do not simply carry fixed information because they are always part of a process involving other factors and can have different effects depending on what other factors are present (e.g., Fisher 2006; Gottlieb 2007; Meaney 2010; Stiles et al. 2015). If what May is concerned with is regularity in outcome in typical human ways of life, then, from a developmental systems perspective, this can arise as an outcome from the whole human developmental system (e.g., Lickliter & Honeycutt 2015). From this perspective, it is essential to study this whole interactive and bi-directional matrix, and it is not possible to clearly separate social from biological factors because they mutually create each other (e.g., Carpendale et al. 2013). It is one thing to claim that something is common in human development, but using the word innate is not an explanation; instead it is what must be explained (Lickliter & Honeycutt 2009; 2015).

One way of framing debates within moral psychology is by considering what aspect of morality the various authors, all claiming to be studying morality, are trying to explain. For example, Haidt (2001), as a social psychologist, is concerned with explaining typical everyday behavior, such as the way people tend to justify their choices. Although this is an aspect of human social life (e.g., Carpendale & Krebs 1995), in doing so he focuses on a small part of the overall picture of morality. Haidt's theory, along with the other approaches that May reviews, tend to overlook moral norms as a problem to be dealt with because they are just explained away as either imposed by others through socialization or they are considered evolved innate ways of thinking. However, we have argued that explaining moral norms as arising solely from biology is problematic, and, although socialization has a role in moral development, it is not a complete explanation because it fails to account for the origin and change of moral ideas and it entails moral relativism (e.g., Carpendale et al. 2013).

Because May has not argued for another position, it would seem that he is left with moral relativism, but this seems to clash with his wish to enhance moral knowledge and virtue because his goal seems to assume progressivity in the sense of some positions being better than others. Thus, there is something missing in May's work, although presupposed implicitly. What is missing is a discussion of moral norms and their sources. Here we argue for a third option based on the view that moral norms do not pre-exist in either the individual or the previous generation but instead emerge through a social process in the context of interaction in particular types of relationships (Carpendale et al.

2013). To be clear, we are not arguing for objective moral truths, nor moral relativism. Our position, based on Mead and Piaget, is different and not on the usual dichotomous choice of possibilities. It involves stepping off the pendulum, and it is based on a constructivist view of knowledge.

We encourage May to take a developmental approach to emotions and to the source of moral norms and reasoning in particular forms of interpersonal interaction. As biological and cultural approaches to moral norms fail to fully capture what needs to be explained in the case of human morality, that is, the normative dimension of right and wrong (Carpendale et al. 2013), we argue for a third approach – that moral norms emerge within particular forms of interpersonal interaction that create the potential for mutual understanding and agreement. We suggest that to do so requires extending thinking about the role of emotions in morality to considering how they structure the relationships within which interpersonal understanding is achieved and moral norms can potentially emerge. We have drawn inspiration for this way of thinking from G. H. Mead (1934) and also Piaget (1932/1965) and Habermas (1983/1990) in rooting the emergence of moral norms in interpersonal agreement and communication (Carpendale 2009; 2018; Carpendale et al. 2010; 2013).

Setting this debate in a historical context is helpful in bringing out our points. At least part of the current pendulum swing away from reasoning and toward emotions was a reaction to Kohlberg's focus on moral reasoning, but, in fact, it fails to fully address the very problems he was concerned with to do with resolving conflicts between moral rules. It is true that one well-known aspect of Kohlberg's work was his focus on the development and use of forms of moral reasoning. This was based on Kohlberg's problematic interpretation of Piaget's idea of stages (Carpendale 2000). But a less well-known, and perhaps at least partially incompatible, aspect of Kohlberg's complex theory and research was his view of moral development as movement toward ideal role taking, a view converging with G. H. Mead (1934) and Habermas (1983/1990). This perspective is also consistent with Piaget's pioneering work, which has generally been overlooked, perhaps because it was considered merely the inspiration for Kohlberg and therefore grouped with Kohlberg's focus on reasoning. In fact, Piaget's work is different, and they could be said to approach the same point but from opposite directions (Carpendale 2009; Wright 1982).

Whereas Kohlberg (1981) began from reasoning, Piaget (1932/1965) started from activity. Piaget's work brings us back to a developmental approach to the link between emotions and reasoning. For Piaget, mutual affection between individuals structures the social relationships in which morality develops. Therefore, emotions are of central importance for Piaget, but this is a radically different role than that assumed by Haidt and others. Within cooperative relationships among equals children work out practical ways of getting along with each other and treating each other properly and with respect – that is, morally. Children like playing with their friends and to do so they must develop a lived morality, a way of treating each other with respect as embodied in their interactivity. They may only become able to articulate such values later as they come to be able to verbalize what is first implicit in their activity. Relationships of equality are best suited for reaching mutual understanding and arriving at a moral solution, through a moral process, what Mead (1934) referred to as a “moral method.” This is because individuals are obliged to listen to each other and explain their own positions. Such cooperative relationships contrast with relationships of constraint based on one sided respect and inequality. Piaget introduced these two contrasting types of relationships in terms of peer

relationships versus parent-child relationships, but he also acknowledged that any relationship is some mixture of the two types and that certainly not all peer relations are cooperative nor are all parent-child relationships completely constraining. From this perspective, moral norms do not pre-exist but can emerge given certain developmental conditions in the human developmental system. This approach is only recently being recognized as a source of moral norms within social interaction (Göckeritz et al. 2014).

It might seem that we are arguing for an overly optimistic view that clashes with the extent of injustice and equality clearly evident in our world. But Piaget's (1932/1965) point with this third option is that although there are many factors at play in subverting equality, such as cultural belief systems and power imbalances, oppression is inherently unstable, leading to a constant struggle toward more equality. There is a kernel or potential to move in the direction of more equality bound up in the conception of a person as embodied in interaction and communication (e.g., Carpendale 2018). There are many factors involved in explaining injustice and why people do not always do the right thing or do it for the wrong reasons. But what seems at least as important and more difficult is to explain how it is that such injustice can be recognized. It is also telling that the inhumane treatment of people and groups is typically accompanied by, and justified through, a dehumanization process. That is, the respect, and thus moral consideration, that comes from being treated as a person is denied to them.

The social character of moral reasoning

Nick Chater^a , Hossam Zeitoun^{a,b}
and Tigran Melkonyan^a

^aBehavioural Science Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom; and ^bStrategy and International Business Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom.

nick.chater@wbs.ac.uk hossam.zeitoun@wbs.ac.uk

tigran.melkonyan@wbs.ac.uk

<https://www.wbs.ac.uk/about/person/nick-chater>

<https://www.wbs.ac.uk/about/person/hossam-zeitoun>

<https://www.wbs.ac.uk/about/person/tigran-melkonyan>

doi:10.1017/S0140525X18002583, e149

Abstract

May provides a compelling case that reasoning is central to moral psychology. In practice, many morally significant decisions involve several moral agents whose actions are interdependent – and agents embedded in society. We suggest that social life and the rich patterns of reasoning that underpin it are ethical through and through.

May (2018) makes a compelling case for the importance of moral reasoning that inform our ethical judgments and actions. This conclusion is reinforced if we widen our scope to consider situations in which morality seems to depend on not only our own actions, but also the actions of others; and, more broadly, ethics

concerns rules and policies for the smooth operation of society, in which each person has specific roles and responsibilities. Moral agents are not lone and omnipotent decision makers, setting the course of a moral microcosm in which they have jurisdiction (e.g., whether to pull the lever in a trolley problem; whom to rescue in a shipwreck; and so on). They are instead active participants, alongside other active participants, in an endlessly complex social world of families, organizations, nations, professions, customs, conventions, norms and laws.

Consider, for example, the well-known transplant dilemma (Thomson 1985) that May discusses in chapter 3. The dilemma is whether a surgeon should forcibly remove the organs of one person to save the lives of five others, and hence apparently generate a net gain, from a utilitarian point of view (note that such actions are not allowed by the Pareto criterion in welfare economics). The extreme concern that most of us feel about this action might, of course, be set aside as emotionally driven squeamishness. But, on reflection, our distaste surely has a credible basis in moral reasoning. A world in which such practices were sanctioned would be one in which patients would refuse to go to hospital, staff would flee for their lives, doctors would be feared rather than welcomed, and surgeons would resign en masse. To sanction such behavior would be to risk pulling apart the entire fabric of the healthcare system, and to rip up fundamental tenets of law and policing. Indeed, an enthusiastic advocate of the utilitarian approach might attempt to prosecute doctors for refusing to make such transplants (leading, by assumption, to a net “loss” of four lives); and to prosecute police, prison officers and judiciary who refuse to comply. Such considerations seem to provide ample reason to explain our revulsion. Indeed, these considerations would surely be in the forefront of the minds of physicians, medical ethicists, and government policymakers, were the possibility of allowing such transplants a politically live issue. (May rightly makes a related point in terms of the reasons people give – regarding guilt, long-term psychological harm, shame or, potentially, undermining of religious beliefs and practices – when justifying “harmless” taboo violations; see Royzman et al. 2015a).

Some moral philosophers and moral psychologists might wish to wave aside such concerns, insisting that we focus only on the microcosm of the “thought experiment,” and nothing beyond it (as if, for the purposes of the example, the world consisted of six patients, a surgeon, and nothing more; or of an isolated careening trolley car, some people it may strike, some levers, and one or two hapless bystanders). But this *asocial* idealization, in which the ethical dilemma is disconnected from wider society, will be fundamentally misguided if, as we suggest, the fundamental rationale for our ethical principles and intuitions is the well-functioning of that society. Indeed, attempting to introspect, or collect data, on such putatively isolated moral problems may be akin to attempting to understand shoaling behavior by studying the movements of an isolated fish, out of water.

Indeed, such isolated examples are inevitably likely to yield limited insight into the rich web of moral reasoning which guides social life, because they are deliberately disconnected from that web. A parallel tack in epistemology would yield similar conclusions: Suppose people were asked what could be concluded *solely* from finding that the light passing through a prism forms a spectrum, or that feathers and cannon-balls fall at the same speed in a vacuum. If such questions must be answered without any connection to the rest of our knowledge of the physical world, then few conclusions will be forthcoming; and one might be tempted to

conclude that reason plays little role in science too. But, again, the disembodied example is stripped of useful reasoning – because the practically relevant reasoning concerns the relationship between specific experiments (or moral dilemmas) and the web of knowledge in which they are embedded.

Note, too, that the richness and complexity of moral reasons depends on our “location” in the social world – a matter ignored in many philosophical examples and psychological experiments. Consider, for example, the moral dilemma faced by a college-admissions tutor, who realizes that an applicant is the daughter of a close friend. The applicant’s test scores are just below the cutoff; but the tutor knows that the daughter has a phobia of tests and performs much worse than she could. For most of us, the case seems clear-cut: the tutor should apply the same rules to everyone or, and probably preferably, refer this student to a colleague. Why? Because there is an agreed process for impartially handling applications; and the admissions tutor’s role is to follow that process. These are the reasons that the tutor would presumably provide to explain making no exception. The consequences for the applicant (and for the applicant whom she might displace) are not relevant considerations (conversely, were the tutor to make an exception, a great deal of reasoning would be provided – the extremity of the case, the potential loss of a shining academic star, the personal devastation, and so on).

The moral psychologist or moral philosopher might be tempted to respond: but these reasons are all about why behaving in a particular way discharges a person’s job – here, what is *right for an admissions tutor*. But perhaps morality is about what is right *simpliciter*. We suggest that this type of response goes to the heart of the problem. If moral reasoning guides social behavior and the roles and responsibilities each of us has in society, then the very idea of “right” – independent of roles and responsibilities – verges on incoherence. The moral decision makers are not distant and omnipotent decision makers; they are real human beings, struggling with their conflicting roles of, here, being admissions tutor and helpful family friend.

As noted, much work in moral philosophy and moral psychology is not merely *asocial* and concerned with decision makers with no “location” in the social setting. Much such work appears, moreover, to be directed at a hypothetical omnipotent decision maker, rather than at participants with specific roles in an unfolding drama (see Sugden 2018, for a closely related argument in economics).

Often, the question at the heart of ethical debate – and implicit in many related psychological studies – is close to: What would you decide should be done here *if you ruled the world* (benevolently, of course)? But this is surely an unhelpful viewpoint! Each of us makes our ethical decisions locked within not just a specific role, with limited power, but at the mercy of many other decision makers, each making their own ethical decisions. And, worse, the results of our choices are interdependent, in potentially complex ways. Thus, we might expect that a good part of ethical reasoning will concern how we coordinate and negotiate our way through a mass of other people, each coordinating and negotiating as we are. And then the goal of ethics might properly be directed to helping individual citizens manage such challenges from their specific vantage point.

Consider, for example, a variation of the much-discussed trolley car example, originated with Foot (1967). Suppose that the trolley is hurtling toward 10 people whom it will kill instantly. A set of 5 people each has independent access to a switch that will divert the trolley to a parallel track. Unfortunately, this switch works on a toggle: each time the switch is pressed, the train flips

track again. So, if an odd number of switches are pressed, then disaster will be averted; if an even number is pressed, it is not.

Imagine, to start with, that it is common ground that all five people are well intentioned: they want to avoid calamity. But, still, what is the right thing to do?

Suppose, for example, that A knew that B, C, D, and E will do nothing. Then A should, of course, press the switch. But perhaps one of the others will press the switch; then A doing the same will cause, rather than prevent, disaster. Or perhaps two of the others will press the switch; in which case A must press, too. And the others, B, C, D, and E, face the same dilemma of course.

Note, though, that there is an intuitively elegant solution to this puzzle, which will doubtless already have occurred to the reader. Because there is an odd number of players, if all five people press the switch, then success is guaranteed.

Suppose, that each person notices this, each therefore presses the switch, and the good outcome is obtained. The reasoning involved here is rather subtle. One way to reconstruct this reasoning is for each player to ask themselves: If we could communicate, what policy would we agree? If it is “obvious” that the simplest and most general policy is that everyone chooses to switch, and that they would agree this policy were they able to communicate, then communication is unnecessary. A, B, C, D, and E simply imagine the outcome of the hypothetical process of reaching an agreement and implement the result. This is the type of reasoning we call virtual bargaining (Melkonyan et al. 2018; Misyak et al. 2014) – people imagine the outcome of a hypothetical bargaining process and directly implement the agreement.

Notice, crucially, from a virtual bargaining standpoint, ethical theory focuses on advising individuals about what they should do, given their collective challenge; it helps people align their behaviors to jointly achieve a successful outcome. The fundamental challenge for the moral philosopher is not: What should I command that these people do, if I ran the world, but rather, how might I help advise individuals in this situation to help them collectively bring about a good outcome?

Let us imagine, for a moment, that E chooses not to press the switch, and disaster occurs. What is the moral status of E’s action? The others may turn on E and blame her for the disaster: the moral emotions will be dialed up to maximum. But notice that reasoning is the source. Suppose E tried the following retort: “Well, if any one of us had done something different, all would have been well. I’m not especially to blame” (and indeed, many models of responsibility, e.g., Chokler & Halpern [2004], have difficulty with this type of case). This would be met with utmost scorn. But suppose E turned out to be misinformed – unlike the others, E had been told nothing about the functioning of the button; or perhaps E had been told there were six, not five, people with buttons. Then E is absolved of guilt; our collective rage might be directed at F, who deliberately, and with malice aforethought, misled E to bring out disaster.

Our moral emotions are directed at who seems to be to blame; and who seems to be to blame (no one, E, or F) depends on the outcome of subtle moral reasoning about hypothetical agreements.

A final possible objection. Can the proponent of an emotion-based account of moral psychology suggest that all this reasoning is not *moral*, but is simply reasoning about goal-directed social behavior (and that the goal in this case is saving lives, which is where morality enters)? We propose the very opposite: that morality suffuses every aspect of social behavior; that the prescriptions of what we should and should not do, which rules we should

live by, what is worthy of praise and blame, are moral through and through. Moral reasoning is the foundation for society in much the way that reasoning about the external world is the foundation for science. Laws, money, institutions, roles, rights, responsibilities and governments are all products of moral reasoning. May is right: moral reasoning is of primary importance. Indeed, the creation, critique and defense of moral reasons, large and small, is the essence of our emotional, social and political lives.

Optimism in unconscious, intuitive morality

Cory J. Clark^a and Bo M. Winegard^b

^aDepartment of Psychology, Durham University, Durham DH1 3LE, United Kingdom; and ^bDepartment of Psychology, Marietta College, Marietta, OH 45750. cory.j.clark@durham.ac.uk bmw002@marietta.edu
<https://www.dur.ac.uk/psychology/staff/?id=17418>
<https://www.marietta.edu/person/bo-winegard>

doi:10.1017/S0140525X18002558, e150

Abstract

Moral cognition, by its very nature, stems from intuitions about what is good and bad, and these intuitions influence moral assessments outside of conscious awareness. However, because humans evolved a shared set of moral intuitions, and are compelled to justify their moral assessments as good and rational (even erroneously) to others, moral virtue and moral progress are still possible.

We agree with May (2018) that many recent criticisms of moral cognition have been hyperbolic. Moral progress is an indisputable fact of human history (Pinker 2011b). It is therefore likely that moral reason, coupled with changing norms, institutions, and technologies, can change opinions and guide humans to new (and likely better) moral assessments (or at least better for that social and historical context). However, ultimately, moral reasoning stems from moral intuitions, which, like axioms in geometry, one must simply accept as givens. These unreasoned intuitions evolved in the same way desires to apprehend beauty, eat delicious food, and appear rationally consistent did.

The sources and nature of these intuitions might be knowable if we scrutinize them enough, but we likely can only know them the way we might know the structure of atoms or the substance of distant stars: By observing and carefully analyzing them. Furthermore, they often motivate our moral behavior and assessments in ways that remain inscrutable to most of us. When we judge that, say, stealing a marble rye from an elderly woman is wrong, we do not know why we have the intuition that it is wrong; we only know that we do have the intuition.

In the following, we will argue that the critical distinction about moral cognition is not between reason and emotion (two concepts arduous to define), but between conscious and unconscious processing. We believe that much of moral cognition is impelled by unconscious processes and that even conscious processes flow from moral intuitions whose causes remain obscure to introspection. However, because humans evolved desires for

good moral reputations, cognitive consistency, and to justify opinions and behaviors to other humans, moral virtue and moral progress are still possible.

May contends that reasoning can be unconscious. This raises a difficult definitional problem, which we will not try to resolve here. Still, however one defines “reason,” there are crucial differences between unconscious and conscious cognitive processes. For present purposes, chief among them is that we are often ignorant about the causes and contents of unconscious reasoning. For example, if one is strolling down a street and suddenly has the thought “breeding dogs is bad,” then one would be unaware of the causes of this cognition. It might be that one had carefully considered the consequences of dog breeding sometime before and that the fruits of such considerations finally burst forth as one was walking. However, it may also be that one recently saw an American Society for the Prevention of Cruelty to Animals (ASPCA) commercial and was feeling particularly emotional about the number of homeless animals, or that one was looking for an excuse to judge their snooty neighbor who just spent \$5000 on an exotic dog breed. The unconscious reasoner has no way of identifying what compelled their sudden conscious conclusion.

Furthermore, because humans are designed and motivated (often unconsciously) to persuade other people, they are often biased about the purported causes of their own cognitions. To persuade others often requires appealing to universal principles. Therefore, humans likely believe that their judgments are caused by such principles more often than they are (e.g., I dislike him because he is a jerk, not because the person I have a crush on likes him better than he or she likes me). Copious data suggest that humans are indeed often ignorant about the causes of their attitudes and behaviors (Nisbett & Wilson 1977) and are easily misled about the motives underlying their moral inferences (Haidt 2001). This suggests that scientists should be suspicious of the manifestations of unconscious reasoning. Because many of the actual causes of moral inferences are not accessible to consciousness, attempts to explain reasons for moral judgments are often mere speculation. For example, we might believe that killing babies is wrong but have absolutely no access to why we have this moral judgment. However, humans loath to admit that they have no introspective access to the causes of their judgments, so they often confidently assert something (e.g., “because that causes them pain and suffering”) that cannot be the actual reason for their judgment (e.g., most humans would still consider it highly immoral even if the murder could be made painless).

Despite this general lack of awareness, humans are compelled to explain and justify their judgments and behaviors to others (Mercier & Sperber 2011) – so as to maintain their reputations as good and reasonable actors. People care deeply about preserving their moral reputations (Vonasch et al. 2017) and so people will be compelled to produce explanations for their behaviors and assessments that serve this goal. And though the ultimate goal is to convince others that one is morally virtuous, one can be more persuasive if they also personally believe their own moral stories, and so self-deception would be useful. Moreover, people wish to be and to appear cognitively consistent (Festinger 1957) and want their moral judgments to appear rational and justifiable (Clark et al. 2017). These desires compel individuals to alter supposedly objective features of moral cases (such as how much control a moral actor had or whether a particular action caused harm) to appear morally coherent (Clark et al. 2015; Schein & Gray 2018). Thus, the reasons people produce

for their moral actions and assessments will be designed to signal virtue and justifiability rather than to describe the true underlying cognitive processes.

This inability to access one’s reasons for moral judgments coupled with the passions that moral assessments provoke does challenge moral rationality, and it makes moral discourse, debate, and reasoning supercharged – and often full of deception. In many cases, moral conversation is a façade that disguises underlying processes of which the interlocutors are utterly unaware. This might cause pessimism and cynicism about moral discourse. Even the explanations that people forward for their moral judgments that do appear reasonable and rational are often post hoc justifications. However, because there are pressures to justify one’s judgments and behaviors both morally and rationally – these judgments and behaviors often will be constrained by what humans can explain as moral and rational. For example, one might refrain from attacking a romantic rival because such a behavior would be difficult to justify, morally or rationally. That is, one might have a strong desire to denigrate or fight a rival, but then think, “could I justify this to another person”? If not, then one might not follow through on one’s desire. And so even if social norms about what is moral and rational are not the ultimate (or even proximate) causes of moral judgments and behaviors, these norms will influence and constrain moral judgments and behaviors.

Before we elaborate further on why we should not throw the moral baby out with the reason bathwater, we would like to clarify why intuition (or passion) is an inseparable part of moral judgment, as Hume contended (and many others misunderstood). It is not that emotions should drive our moral assessments (this is not the meaning of Hume’s famous “reason is and ought to be the slave of the passions”), but rather that it is not possible to reason one’s way to a moral conclusion without an intuition about valence (e.g., “pain is bad,” “it is good to maximize goodness for sentient creatures”). Moral judgments, by their very nature, must be grounded in intuitions about what is good and what is bad. If a cognitively sophisticated sadomasochistic robot shared the seemingly universal human intuition that it is generally good to maximize goodness for sentient creatures but also had an intuition that pain is good, a morally good and rational BDSM robot might then conclude that they ought to cause as much pain as possible. Without these pre-rational preferences, humans (and robots) would not have moral judgments because they would not have preferences at all.

Fortunately, humans evolved a shared set of moral intuitions from which moral reasoning can build. Humans generally agree that pain and suffering are negative experiences and that we should minimize negative experiences for ourselves and others we care about, and so we can make a variety of claims about what types of behaviors (those that cause pain and suffering in others) are morally wrong. This is a minor point, but a pervasive mistake in moral psychology. Yes, transient emotions or passions often influence moral judgments. That addresses more proximate causes of certain moral judgments and behaviors. But all moral judgments are based on unreasoned intuitions about what is good and bad in the same way that all aesthetic judgments are ultimately based on unreasoned intuitions about what is beautiful and ugly. The same way we evolved to find bodies of water, clear skin, and bright red strawberries appealing, we evolved to find generosity, honesty, and selflessness appealing.

Though presumably, May would disagree with this characterization of moral judgment, an implicit understanding of this reality permeates his writing. For example, he points to motivations

such as avoiding punishment, feeling better about ourselves, and being more likable to others as wrong reasons for moral behavior. But how do we know these are “wrong” reasons? May argues “something seems morally lacking in such actions.” We agree something seems lacking. But we did not reason that these are the “wrong reasons” for moral behavior. Rather, we share an intuition that desiring to benefit the self is not a morally good reason (and we suspect many or most humans share this intuition). One could argue, however, that given that promoting one’s moral reputation is an action that benefits the self, and also motivates behaviors that help others, self-interested moral motivations might often be virtuous.

None of this means that unreasoned intuitions are fully formed moral judgments or that we never use reason to make moral assessments. It merely means that we must build our moral judgments and arguments from the raw materials of our moral intuitions.

May argues that humans would not attempt to rationalize or justify their moral judgments and behaviors if they did not have regard for reason. Though we have explained that these rationalizations and justifications are often deceptive post hoc explanations, we agree that people do care about appearing reasonable. And though May might find something morally lacking in this type of “reasoning,” we remain optimistic about moral progress. The same way humans evolved desires to appear morally good and shared intuitions that harming others is bad, humans evolved desires to appear reasonable and shared intuitions about what is reasonable. These shared motivations and intuitions have shaped and will continue to shape moral judgment and behavior, compelling them to be more consistent with rational and universal rules and less nakedly selfish and parochial. And just because these virtues evolved for purposes other than enlightened prosociality, this does not mean we cannot admire them in the same way we admire ambition or beauty. And in fact, we should admire them because such admiration incentivizes virtuous behavior.

This means that moral discourse and argumentation will have significant, predictable effects on human behavior. If we want people to cease eating factory farmed meat, then we should appeal to their desire to seem reasonable. Point out that they would not torture a chicken to save a dollar on their next order of wings, but that factory farms do just that: they create torturous living conditions to provide lower prices to consumers. Although the effects of such arguments might be small at first, as more people come to agree with them, the social pressure makes it harder for people to justify actions that are incongruous with explicit moral pronouncements. One of the better angels of our nature, it turns out, is persnickety people who demand that we explain our actions.

Analyzing debunking arguments in moral psychology: Beyond the counterfactual analysis of influence by irrelevant factors

Joanna Demaree-Cotton 

Department of Philosophy, Yale University, New Haven, CT 06520-8306.
joanna.demaree-cotton@yale.edu

doi:10.1017/S0140525X18002716, e151

Abstract

May assumes that if moral beliefs are counterfactually dependent on irrelevant factors, then those moral beliefs are based on defective belief-forming processes. This assumption is false. Whether influence by irrelevant factors is debunking depends on the mechanisms through which this influence occurs. This raises the empirical bar for debunkers and helps May avoid an objection to his Debunker’s Dilemma.

In chapter 4 of *Regard for Reason in the Moral Mind (RRMM)*, May (2018) tackles sweeping debunking arguments that aim to show that ordinary moral beliefs are not justified because they are inappropriately influenced by morally irrelevant factors, such as incidental disgust, how a moral scenario is worded or “framed,” or whether or not an agent uses “personal force” to bring about a harm. May argues that, for the debunker’s argument to succeed, they need to identify an influence on belief that is both substantial (empirical premise) and defective (normative premise). A defective influence on belief is an influence on belief that, if substantial, renders that belief unjustified (such as wishful thinking). However, May argues that debunkers are faced with the following dilemma: Either the irrelevant factor in question is not a “substantial” influence on moral belief, thus undercutting the empirical premise, or the debunker has identified a substantial influence on belief that is not defective, thus undercutting the normative premise.

Throughout his book, May slides between two ways of presenting the would-be debunker’s empirical premise, sometimes writing in terms of:

1. Whether or not moral beliefs are substantially influenced by epistemically defective *factors* (that is, morally irrelevant factors)

And sometimes writing in terms of:

2. Whether or not moral beliefs are substantially influenced by epistemically defective processes

It is fairly clear what it is for a moral belief to be the product of an epistemically defective process. Wishful thinking, motivated reasoning, and paranoid inferences are all examples of belief-forming processes that result in unjustified beliefs (e.g., *RRMM*, p. 85).

But what is it, exactly, for a moral belief to be substantially influenced by an irrelevant factor? May seems to rely on an analysis in terms of *counterfactual dependency*. On this view, moral beliefs are substantially influenced by an irrelevant factor if the agent would have formed a different moral belief – a belief with a different polarity or valence – had that factor been absent (*RRMM*, p. 213). For example, my moral belief that an agent is doing something morally wrong is substantially influenced by incidental disgust in this sense if, *had* I not experienced incidental disgust, I *would not* have believed that the agent is doing something morally wrong. Similarly for framing: If one accepts this counterfactual analysis of the debunker’s empirical premise, then “what a debunking argument requires” [...] is “that people regularly tend to lose their belief or change its content” if the framing of a moral problem is altered (*RRMM*, p. 90).

It seems that the reason that May is happy to slide between talk of epistemically defective processes and talk of dependence on

irrelevant factors is that he assumes that if a moral belief is counterfactually dependent on a morally irrelevant factor, then it is the product of a defective process – one that, like wishful thinking, guesswork, or motivated reasoning, is unreliable, insensitive to evidence, or otherwise yields unjustified beliefs (*RRMM*, p. 85). This assumption is tempting and is widely held by debunkers (e.g., Sinnott-Armstrong 2008) and by anti-skeptical defenders of moral judgment alike. Indeed, I have explicitly outlined and relied on such a counterfactual analysis of the threat of irrelevant factors elsewhere (Demaree-Cotton 2016).

The trouble is that this assumption – the assumption that if a moral belief is counterfactually dependent on an irrelevant factor, then it is the product of a defective process – is false. Moreover, it concedes too much to the debunker.

The assumption is false because counterfactual dependence *per se* tells us exceedingly little about what the actual psychological process was that led the agent to form the moral belief that they did, let alone whether or not that process is a defective one. There are a number of reasons for this. One is that an irrelevant factor can influence moral belief precisely because it influences what kind of belief-forming process an agent engages in. Imagine an experimenter induces incidental feelings of anger, and then presents me with a moral vignette and asks whether the main character is doing something morally wrong. Perhaps because of the incidental anger, I feel engaged by the task – the topic of morality feels, right now, like something interesting and important, worthy of deep reflection – and consequently I carefully reflect on details of the scenario, weighing up whether or not there is a suffering victim, whether or not this suffering constitutes an injustice, and contemplating the main character's role in the production of that injustice. On the basis of these considerations I conclude the main character did something morally wrong. This is, I take it, a canonical example of a justification-conferring psychological process. The resulting belief is justified. Moreover, it would remain justified even if, had some irrelevant factor been different – for example, had I not just gone through the anger induction – I would have engaged in a different type of belief-forming process and formed a different belief as a result. This fact cannot in itself bear on whether or not my actual belief is justified, because it does not bear on what the actual psychological process was that led to that belief.

This is so even if, counterfactually, I would have engaged in a defective belief-forming process. Imagine that instead of anger the experimenter used a mood manipulation to enhance feelings of cheerfulness. Moreover, if she had, I would have taken little interest in the moral dilemma presented to me; while daydreaming about my weekend plans, I would have absent-mindedly formed the belief that what the main character is doing is probably perfectly morally permissible. This is not a reliable, rational way of forming moral beliefs, and the resulting belief would not be justified. Still, this counterfactual fact about me in no way impugns the careful reasoning I actually engaged in after the anger induction and the justificatory status of the belief I formed on the basis of that reasoning.

Another possibility is that a morally irrelevant factor affects, not what *type* of belief-forming process you engage in (e.g., reasoning based on evidence versus motivated guesswork), but merely what subset of all of the relevant evidence you (1) notice, and (2) pay attention to and weigh when arriving at your belief (see Avnur & Scott-Kakures 2015, on “positional” influences of irrelevant factors on belief). To return to the example above, perhaps my feelings of incidental anger remind

me of cases of injustice I experienced in the past, and consequently I am especially sensitive to aspects of the moral situation described in the vignette suggesting that the victim is suffering an unjust harm – aspects of the moral situation that influence my judgment and that I might not have noticed had I been in a different mood.

Of course, this is a purely hypothetical example, and not an empirically grounded one. This may well not be the mechanism by which incidental anger can be expected to influence moral belief. But this is exactly the point. For the debunker's argument to work, it is *not sufficient* for them to cite studies showing morally irrelevant manipulations have a significant effect on moral beliefs, even if the effect sizes in question are large. They must bring to bear theoretical interpretations of those effects on moral beliefs that give us reason to think that the proximate psychological mechanisms driving those moral beliefs are indicative of defective, non-justification-conferring psychological processes.

This also shows why May's slide between talk of a moral belief being “substantially influenced” by irrelevant factors and talk of a moral belief being “mainly based on” irrelevant factors is liable to mislead in a way that does his argument a disservice. We normally use the term “basis” to pick out, not just *any* substantial causal influence on belief, but a cause that is a crucial part of the psychological process leading to the belief (typically, a piece of the agent's evidence). For example, in the hypothetical anger induction example, my moral belief was counterfactually dependent on whether or not I had undergone an incidental mood induction; but my moral belief was *based* (in a psychological, evidential sense) observations I made about injustice. To conflate causal influence with psychological basis risks seeing defective belief-forming processes where there are none.

The final difficulty I want to outline regarding the counterfactual analysis is not just that it is false, nor just that it obscures the need for appeals to details of psychological process and mechanism, but that it concedes too much to the debunker – potentially in ways that threaten to weaken May's anti-skeptical arguments. A really important part of May's anti-skeptical argument is his motivation of the Debunker's Dilemma and the idea that *any* empirically based debunking argument will most likely face a tension between establishing that moral beliefs are substantially affected by some source and establishing that that source renders the belief unjustified. Why think the dilemma will generalize? According to May, because there are *many different* influences on belief – some of which are defective, and some of which are non-defective – it is unlikely that any *particular* defective influence will have a substantial effect on a large class of our moral beliefs (*RRMM*, pp. 103–104).

Furthermore, he argues that it is unlikely that lots of individually insubstantial defective influences (such as morally irrelevant factors) will add up to have a substantial effect on a large class of moral beliefs in a way that is debunking, because individually insubstantial *appropriate* influences on beliefs (such as morally relevant factors) can add up in exactly the same way (*RRMM*, p. 229).

But it is unclear why this would be true if we thought that mere counterfactual dependence on irrelevant factors counted as a defective influence on belief, where that influence is specified in a way that abstracts away from psychological process. Any moral belief we form has been affected by countless irrelevant factors in this purely counterfactual sense, insofar as the development our evidence (including our background beliefs and our

moral convictions) and what exact belief-forming process we engage in (including what evidence we consider and what type of thinking or reasoning we use to form the belief) is counterfactually dependent on an innumerable number of irrelevant factors, from our evolutionary history, to where we were born, to our socioeconomic status, to our health, to our mood, to whether or not we happened to have faced a similar moral problem previously, and so on. The number of factors that are morally relevant to the problem at hand that could influence us, by contrast, is necessarily restricted.

This problem goes away if we recognize that only a specific way of being influenced by irrelevant factors is pertinent to assessing the justificatory status of moral belief – namely, being influenced in such a way that leads you to engage in a particular kind of defective belief-forming process (e.g., because the factor in question tends to produce motivated reasoning, or because we tend to form moral beliefs that use that irrelevant factor as a heuristic, although the heuristic in question is unreliable). If we focus psychological dependence on irrelevant factors, rather than a more general sense of counterfactual dependence, then it is much more plausible that small appropriate influences are going to stack up against small inappropriate influences on belief.

In summary, I have presented an argument that is both critical of an assumption that May makes in *Regard for Reason in the Moral Mind*, but one that he should welcome if he wishes to defend ordinary moral belief. Counterfactual dependence on irrelevant factors does not matter. What this shows is that, to succeed, would-be debunkers have to meet a much more stringent empirical premise than May has allowed. May is absolutely right that statistical details matter (*RRMM*, p. 229). It matters how big of an effect irrelevant factors have on our moral beliefs. But it also matters *what kind* of an effect irrelevant factors have on our moral beliefs – the details of the psychological processes through which irrelevant factors come to affect our moral beliefs are of crucial importance when assessing the merits of debunking arguments in moral psychology.

The faces of pessimism

John M. Doris 

Philosophy-Neuroscience-Psychology Program, Washington University in St. Louis, St. Louis, MO 63130.

jdoris@wustl.edu

<http://www.moralpsychology.net/jdoris/>

doi:10.1017/S0140525X1800273X, e152

Abstract

In this commentary on May's *Regard for Reason in the Moral Mind*, I argue that many of the interdisciplinary moral psychologists whom May terms "pessimists" are often considerably more optimistic about the prospects for progress in moral inquiry than he contends.

1. May's (2018) *Regard for Reason in the Moral Mind* addresses the interdisciplinary, empirically informed, moral psychology that has proliferated in the philosophy and psychology of the

past 20 years. The overall tone of this work, May (preface, p. xi) contends, is *pessimistic* "about ordinary moral thought and action" Against such pessimism, May (p. xi) argues that "our best science helps to defend moral knowledge and virtue against prominent empirical attacks," and thereby casts himself as an *optimist*, defending traditional notions of reason and virtue. Yet May (p. 19) departs from the *a priori*, ascientific methodology that until recently dominated moral psychology in anglophone philosophy, remarking that "few optimists have taken the empirical challenges seriously, let alone answered them successfully."

May's assessment of his comrades in optimism may be a little unkind: Although there remain blissfully obdurate *a priorists* in moral psychology, increasing numbers of traditionally minded philosophers are engaging the empirical literature, most notably those working the burgeoning field of character studies (e.g., Miller et al. 2015). But May's book displays considerably more facility with the empirical literature than does the work of many optimists, making for an innovative and important contribution to moral psychology, which ought to be read by everyone in the field.

But (straightaway to the "but" endemic in these exercises) while I'm impressed by May's acumen, I have reservations about his management of the rhetorical space. In particular, I question his development of the optimist/pessimist dialectic. Although I will, for convenience, adopt May's nomenclature, I'll argue, as one of his pessimist foils, that this taxonomy is not generally apt, and I'll therefore, with no disrespect intended, henceforth flag our disagreement with "scare quotes" around *optimist*, *pessimist*, and variants where dialectical clarity requires it. Although there are certainly moments in the literature that are pessimistic in tenor, the sensibility driving the new interdisciplinary moral psychology is probably as often optimistic as not. Sharpening the taxonomy has a serious purpose, because misattributions of morally nihilistic pessimism help fuel the sometimes vitriolic repudiations of interdisciplinary moral psychology found in philosophical commentary.

2. An initial complication is that there are two, imperfectly overlapping, beneficiaries of May's optimism. The first is sometimes called folk morality; May (p. 7) declares "there are no empirical grounds for debunking core elements of ordinary moral judgment." The second is what we might call philosophical orthodoxy – the family of traditional philosophical understandings of moral psychology targeted by the "pessimists." May's (p. 7; cf. pp. xi, 3, 4, 6, 7, 19) frequent use of locutions like "our moral beliefs" and "our moral minds" notwithstanding, folk morality is far from a unity, and neither is there a monolithic philosophical orthodoxy, even within the comparatively narrow anglophone "analytic tradition" where this discussion lies. Nevertheless, certain commitments are often attributed to much of both folk morality and philosophical orthodoxy, at least in their anglophone guises: for example, that reflection has a central place in moral experience; that moral judgments are supported by tolerably undistorted reasoning; that character traits powerfully influence moral judgment and behavior.

May is right that those he dubs "pessimists" have frequently criticized such claims, in both folk and philosophical variants. Yet just as the optimist orthodoxy manifests considerable diversity, so does the pessimist insurgency. To the extent that the pessimism at issue is supposed to be pessimism about the possibility of progress in moral inquiry (metaethical difficulty surrounding "moral progress" hereby noted and skirted), many of May's "pessimists" are not pessimists at all. On the contrary, they understand

their work as *contributions* to progress in moral inquiry. Most often, his “pessimists,” at least those identified as philosophers, target *particular aspects* of philosophical orthodoxy, rather than moral inquiry *in general* (e.g., even Machery’s [2010] gloomily titled “The Bleak Implications of Moral Psychology,” is not generally pessimistic, but focused on difficulties with character and intuitions in ethics).

3. May identifies two main forms of pessimism, one about cognition and the other about motivation. Pessimists about *cognition*, May thinks, are dubious about the role of reason in ethics. In this, he’s not alone: according to D’Arms and Jacobson (2014, p. 253), “the champions of empirical ethics are united in holding that the emotional basis of morality systematically undermines its pretensions to rational justification.” Certainly, among the most central preoccupation of we “pessimists” – a main take home message for our students – is the influence of emotion on moral cognition and behavior, especially the disquieting influence of rationally arbitrary, “incidental,” emotions. But this needn’t entail the *derogation* of reason; a concern about rationally arbitrary influences may embody a *regard* for reason. Indeed, one thing scientific moral psychology can do is help show how people might reason *better*: For example, Cameron et al. (2013) used a simple intervention – a rather rationalist instruction to observe differences among one’s emotional experiences – to ameliorate the influence of incidental disgust on moral judgment.

Furthermore, as May himself notes, two of the authors most concerned about the influence of emotion on morality, Greene (2013; 2014) and Singer (2005; 2015), actually advocate highly aspirational utilitarianisms, rather than moral despair. And they are certainly not anti-rationalists; as D’Arms and Jacobson (2014, p. 255) read these two, they favor a “hyper-rationalist” approach. Finally, Greene and Singer are not even uniformly critics of commonsense morality: Their utilitarianism certainly has roots in everyday intuitions about the moral importance of harm and aggregate harm, and neither are above deploying thought experimental appeals to intuition (e.g., Singer 1999).

May (p. 6) also targets “a brand of *sentimentalism* which contends that moral cognition is fundamentally driven by emotion, passion, or sentiment that is distinct from reason (e.g., Nichols 2004; Prinz 2007).” But although Prinz (2007) may count as an anti-rationalist, other sentimentalists take different views. D’Arms and Jacobson (2014; *forthcoming*) defend “rational sentimentalism” and Nichols’ work has always had something of a rationalist feel, emphasizing the importance of rule-based inference, as well as emotion, in moral judgment (e.g., Nichols 2004, Ch. 1; Nichols et al. 2016). I suspect Nichols is more the sort of empirically inclined “pessimist” May takes in his sights, but D’Arms and Jacobson (2014, p. 254), though at pains to deplore “the scientism implicit in much empirical ethics,” are themselves awed in the business of crafting scientifically credible ethical theory.

In fact, many “pessimist” projects may be seen as animated by a quite orthodox concern with how to harmonize the deliverances of emotion and cognition in optimally reasonable judgments of ourselves and our worlds – a project, it seems to me, quite in the spirit of May’s own. So understood, they join May in extending a time honored philosophical enterprise.

It is true that “pessimists” are more likely than “optimists” to take seriously the science identifying the shortcomings of human rationality. Whether traditional *a priorists* or empirically concerned, “optimists” are more likely to adopt debunking perspectives on the science, apparently in hopes the orthodoxy can persist more or less unchanged. But the “pessimist” must despair

of progress in moral inquiry only if the *orthodox* way to think about morality is the *only* way to think about morality, and the antecedent is manifestly untrue. There is more than one way to think about morality, and these ways may depart orthodoxy to varying degrees.

4. This important point is further illustrated when we turn to May’s treatment of pessimism about *motivation*, particularly as he finds it in discussion of situationism and virtue ethics. May (p. 209) characterizes situationism as “the idea that human behavior is influenced by features of one’s circumstances far more heavily and more often than we tend to think,” though his concern “isn’t necessarily situationism in particular, but a view closely associated with it, to the effect that much of our behavior is motivated by factors we would recognize as arbitrary, alien, or non-reasons.” I’m guessing many drawn to views in the vicinity of situationism hold something like these positions; at least, I’m guessing I do. But May’s (p. 15) real concern is with something else, the thought that if “we are motivated by ethically arbitrary factors” it may be that “we’re chronically incapable of acting for the right reasons.” In May’s (pp. 5, 16, 173, 199–200) view, this is a kind of skepticism about what he calls “virtuous motivation.”

In this context May (pp. 15, 199, 210) mentions Nelkin (2005), Nahmias (2007), Vargas (2013b), and Doris (2015), apparently as pessimist exemplars. But none of us deny that people can act on the right reasons (whatever these turn out to be); indeed, Nelkin and Vargas are best known for their anti-skeptical “reasons responsiveness” accounts of morally responsible agency. May (p. 210) is sometimes more qualified, allowing that “some of these theorists wouldn’t consider themselves to be arguing for pessimism about moral motivation.” But, he (p. 210) thinks, “such frameworks can easily lead to it.” If I am right, a better reading of most theorists in question is that they go to lengths to evade the pessimism initially seeming to follow from taking the troubling empirical findings seriously.

Curiously, May does not cite the main work, *Lack of Character* (Doris 2002), in which I, perhaps with an excess of youthful ebullience, advocated situationism; in the later work he does cite, I (Doris 2015, pp. 14–16) explicitly decline to enter the “character controversy.” (May [213–22] contends that the arguments I make in Doris [2015] are subject to a fatal dilemma. I have [Doris 2018] contested this elsewhere.) When I *was* espousing situationist character skepticism, my target was a *particular conception* of character traits, understood as issuing in cross-situationally consistent behavior. This is only skepticism about “virtuous motivation” if virtuous motivation must flow from a robust “firm and unchangeable” character, as Aristotle may have supposed (see Doris 2002, pp. 16–18). But there are multiple ways way to think about traits and multiple ways to think about virtuous motivation. I was at pains, in developing character skepticism, to eschew moral skepticism; indeed, a central concern was to argue that moral thinking could get on, and indeed get on better, without reliance on empirically suspect notions of character.

I belabor this “inside baseball” issue not – at least not only! – out of the narcissistic pique common to scholars who imagine themselves misunderstood, but to underscore the difficulty with May’s taxonomy. Very often, the “pessimist’s” pessimism is tightly focused – in this case on a particular conception of character traits – whereas May’s objections often address more sweeping arguments that many “pessimists” eschew. Some moral psychologists may tend toward sweeping pessimisms about the prospects of moral inquiry, and I share May’s suspicion of these views.

But I don't think May has identified a more or less homogeneous cadre of interdisciplinary moral psychologists, say as exemplified by myself and my colleagues (e.g., in the Moral Psychology Research Group, www.moralpsychology.net), that is appropriately set up as pessimistic foil to his "optimistic rationalism." As May (p. 18) acknowledges, "pessimism comes in many forms" – and many of those, I'd insist, aren't all that pessimistic.

5. All this said, May is not wrong about the gestalts diverging. For May (pp. xi, 4) is right that many of those he dubs "pessimists" believe that commonsense morality, in many of its many forms, is in need of "serious repair." Here, they often appeal to systematic empirical research, but I suspect that many of them, like me (Doris 2002, Ch. 3; Doris & Murphy 2007; Murphy & Doris, *forthcoming*), are equally motivated by the horrors of human history – as well as an appalling present and terrifying future. Call this the *pessimistic abduction*: Part of the best explanation of why the story of humanity is at so many points a story of moral horror is that our moral thinking is in serious disrepair.

In this respect, the "pessimist's" glass is half empty. And May's, perhaps, is half full. In discussing data suggesting that "a politician's followers are inclined to rationalize continued support even in the face of rather egregious scandals," May (p. 207) concludes, hopefully, that "the love isn't unconditional and supporters will eventually jump ship." On January 23, 2016, a U.S. presidential candidate boasted, "I could stand in the middle of Fifth Avenue and shoot somebody and I wouldn't lose voters." That candidate is now president, and events have done distressingly little to suggest that he is wrong and May is right. (To take one of uncounted examples: if vicious middle school mockery – on camera – of a disabled person does not cost you the love of your diehard supporters, what will?) Here, me and many of my empirically minded colleagues in moral psychology may well be pessimists: We think the impediments to thinking clearly and humanely are many, and the obstacles to behaving accordingly are still more. But there's also a sense in which we are cock-eyed optimists: We are animated by the conviction that a scientifically credible understanding of why we so often go wrong is a necessary part of finding ways to do better. And that, many of us empirically minded moral psychologists would say, is why we do what we do.

Acknowledgments. Many thanks to Justin D'Arms, Edouard Machery, Shaun Nichols, and Stephen Stich for very helpful comments on earlier versions.

Rationalism, optimism, and the moral mind

Quinn Hiroshi Gibson 

Global Perspectives on Society Program, New York University Shanghai, Pudong New District, Shanghai, Peoples Republic of China, 200122.

qhigibson@berkeley.edu
<https://www.quinnhiroshigibson.com>

doi:10.1017/S0140525X1800256X, e153

Abstract

I welcome many of the conclusions of May's book, but I offer a suggestion – and with it what I take to be a complementary

strategy – concerning the core commitments of rationalism across the domains of moral psychology in the hopes of better illuminating why a rationalist picture of the mind can deliver us from pessimism.

I welcome many of the conclusions of May's (2018) *Regard for Reason in the Moral Mind*. He does for the domains of moral cognition and moral motivation what many other philosophers, including myself, have tried to do for agency and moral responsibility. I think "optimistic" and "rationalist" philosophers across these domains share a number of core concerns. Consider the following example (adapted from Gibson 2017, p. 34):

Colin is considering moving to either Delaware or Colorado for work. Colin has the following quirk that he employs to help him decide what to do. He feels that he can focus more on the facts that are relevant to the decision if he writes "Colorado" and "Delaware" down on a piece of paper, rapping his pen against the page while he ruminates. He typically writes them in that order, figuring that when other things are equal (which, attempting to be unbiased, he strives to make them as much as he can) alphabetic order is as good as any.

Colin may be subject to *implicit egotism* (Pelham et al. 2002) with respect to the name of the state he is considering moving to. He may also be subject to an ordering effect. Some philosophers (such as Doris 2015) have used cases like this to argue against rationalist theories of agency and responsibility. One way for a rationalist to respond is simply to question whether the purportedly agency-undermining effects are real (Simonsohn 2011). Another way of responding is to say even *if* they are real, it still remains whether they operate by *bypassing* whatever the supposedly necessary mental processes are for moral responsibility, or whether they operate by running through them. The relevant effects may operate on Colin simply causing him to *attend* to all of the lovely features of Colorado. Then it is far from obvious that his agency is undermined.

At bottom, the debate between rationalists and their empirically motivated opponents over agency and responsibility is about whether having the kind of contact with the normative domain that is thought by rationalists to be required for agency or responsibility is ruled out by an up-to-date conception of the mind. Crucially that involves disputing two different things: (i) what the rationalist picture of the mind really involves and (ii) what the commitments of a distinctively rationalist outlook really are. There is room for rationalist pushback on either score.

Much the same can be said about the debates May wants to intervene in. But May's book is less about (ii) than one might expect. It is very much about mounting an effective rationalist response by thoroughly investigating (i). But it is also about delivering us from pessimism. One could perhaps schematize the arguments that May is attributing to his opponents as follows:

1. Empirical premise describing moral cognition (descriptive sentimentalist premise) or moral motivation (descriptive egoist, Humean, or situationist premise)
2. If (1) then pessimism
3. Pessimism

This argument is about the stakes. But I often found myself wondering if some of his opponents' arguments are not better schematized as:

1. Empirical premise describing moral cognition (descriptive sentimentalist premise) or moral motivation (descriptive egoist, Humean, or situationist premise)
2. If (1) then not-rationalism
3. Not-rationalism

This argument is about mental mechanics. Now, I think it is fair to say that May wants to argue that neither pessimism nor the denial of rationalism should be thought to follow from any of his opponents' arguments. But these are actually quite different positions, and the connection between them is not always clear. (There is no doubt that May's opponents bear much of the blame for this confusion – Haidt (2012) calling rationalism a “delusion” (as cited by May 2018, p. 6) being a prime example of why.)

Perhaps the main virtue of May's book is that it provides ways to resist either argument schema without succumbing to the temptation to simply run the corresponding *modus tollens* against them. However, the debate, as May understands it, between rationalists and their opponents is then forced to turn on the respective roles played by some distinctively rational set of states or faculties, on the one hand, and theoretically competing states such as emotion, affect, or desire, on the other. Seen in this way, the operant questions are then: Which comes first? Which is primary? Which is *essential*? Focusing on these questions requires diving into the studies that purport to answer these questions and giving them a sober look over. May does this admirably, and the rationalist position comes out looking better for it. But framing the disagreement in this way threatens to obscure how defending a rationalist picture can help us avoid pessimism in the first place.

This is because we can always ask *why* it is important that we have proper regard for reason. I take it that at least part of the answer goes beyond simply making empirical room for us to have justified moral beliefs and for those beliefs to at least sometimes move us to action. There is a more general kind of pessimism at issue, and I suspect that those who are inclined toward various forms of rationalism in the domains of moral cognition, motivation, agency, and responsibility are united in resisting it. The concern, it seems to me, is to provide a picture according to which thought and action can be meaningfully connected to the normative. That is, it is important to have proper regard for reason because “reason” picks out a distinctive set of capacities in virtue of which we are able to make contact with considerations that weigh for or against courses of action and states of mind. Those considerations are simply *reasons*.

The capacity to respond to reasons is what makes us normatively sensitive creatures. It seems that those like Colin can still be sensitive to reasons even if they are subject to the operation of any number of “biased” or “unconscious” cognitive mechanisms. Still, one could say that what the twenty-first century view of the mind implies is that what you might call *intellectualism* about how we come into contact with those reasons would inevitably lead to pessimism. We simply do not consider the reasons for our thought and action deliberately, explicitly, or consciously, nearly enough of the time to generally count as responsible agents or justified moral believers if that is what such things require. It obviously helps a lot to acknowledge, as May does (pp. 8–10), that inference and judgment are largely unconscious. But I wonder what the distinctive value of having *cognitive* (rather than some other kind of) contact with reasons is.

May says he sympathizes with the characterization of rationalism (construed in this context more narrowly as a view about

moral judgment) “as the thesis that moral judgment is ultimately ‘the culmination of a process of reasoning’ (Maibom 2010, p. 999)” (May 2018, p. 12). I do not deny that there is a (perhaps largely empirical) debate to be had about whether moral judgment originates in reasoning or in emotion. But from the perspective of rebutting a pessimist about justified moral belief, I might have thought the issue was less whether our moral beliefs are (or rest on) judgments, and more whether they are appropriately connected to moral truths. One source of pessimism is that our moral beliefs are not under our rational control. But to resist this we do not need the etiology of a particular belief to run through judgment (though in cases like Colin's it probably does). It is enough for the state itself to be what Scanlon calls a “judgment sensitive attitude” (Scanlon 1998, p. 20). The etiology of the belief notwithstanding, it is still a state for which reasons can be asked and offered.

The celebrated Huck Finn case (Arpaly 2003; Arpaly & Schroeder 1998) is usually read as one where Huck is *praiseworthy* for helping Jim escape slavery despite his explicit judgment that it would be wrong to do so. On Arpaly's reading of the case, Huck has come to see Jim as a human being after undergoing a “perceptual shift” (Arpaly 2003, p. 77) that resulted from spending time with Jim. After this shift Huck has, on some level, the moral belief that Jim is deserving of certain forms of treatment. One could say he came to this belief as the result of a bunch of unconscious inferences. But my intuition that he is *praiseworthy* does not change if we simply stipulate that Huck has come to this belief purely as the result of non-inferential processes. Still, it seems, Huck has made a kind of contact with the moral domain that leads him to action – through justified belief, no less. Similarly, one can imagine Aristotle's *phronimos* coming to moral belief in much the same way. Being confronted with the particularities of *this situation here and now* the *phronimos* forms the belief that such-and-such is to be done. This belief might be the result of well-conditioned unconscious reasoning serving up that morally correct belief. Indeed it is plausible that in many cases this is how it will be. But I see no reason why in some cases it might not be. To draw the parallel with agency, it is not obvious that if there were some process that made Colin sensitive to the lovely features of Colorado in a way that bypassed judgment his conduct would be any less agential.

This need not be a capitulation to sentimentalism because, on this view, at least from the perspective of the dispute between pessimists and optimists, there is no morally relevant reason/emotion dichotomy. And this is not just because, as May says, we can “place great weight on the cognitive aspects of emotion that can facilitate inference and related belief-forming processes” (p. 228). It is also because both reason and emotion, or even mere feeling, can be ways of getting onto the reasons that are there and we can be accountable for getting on to them well or poorly. There are lots of different *kinds* of reasons and the difference between being sensitive and being insensitive to them need not track a simple – nor at any rate a “fuzzy ... at best” (228) – distinction between kinds of mental states or processes. Some reasons require reasoning to apprehend. But there is no reason to think that all reasons do. This would appear to provide a response to the pessimistic sentimentalist irrespective of whether they think of emotion as brute or as partly cognitive (pp. 52–53).

Humeanism has a pessimistic character because we also want to have rational control over the states that are capable of moving us to action. But again, I think it is enough for the states that are capable of moving us to be judgment sensitive attitudes. One move that May makes against the pessimistic Humean is to

impute to belief many of the functional properties of desire (pp. 193–95). Why this is a move that leads us to optimism is easier to see with the broad aims of rationalist picture and the idea of a judgment sensitive attitude held out front: desiderative states connect us to reasons and enter into relations of justification with other judgment sensitive states.

I found that working through my own reasons for gravitating toward a more optimistic picture across the domains of moral psychology helped me focus more clearly on the stakes of May's project. Some of the routes to optimism that I have suggested are shorter than the ones that May takes, but I do not mean to cast doubt on the value of taking the longer, thornier route that he does. I consider much of what I have said here to be complementary and congenial to May's overall goals, but I am genuinely curious whether he agrees.

Moral foundations are not moral propositions

Daniel Haas 

Department of Philosophy, School of Arts and Sciences, Red Deer College, Red Deer, AB, Canada, T4N 5H5

Daniel.haas@rdc.ab.ca

<https://rdc.ab.ca/programs/academic-departments/school-arts-and-sciences/bachelor-arts/ba-philosophy/faculty/daniel-haas-phd>

doi:10.1017/S0140525X18002728, e154

Abstract

Joshua May responds to skepticism about moral knowledge via appeal to empirical work on moral foundations. I demonstrate that the moral foundations literature is not able to do the work May needs. It demonstrates shared moral cognition, not shared moral judgment, and therefore, May's attempt to defeat general skepticism fails.

Part of Joshua May's (2018) project in *Regard for Reason in the Moral Mind* is to address the threat to genuine moral knowledge that is raised by peer disagreement about morality (pp. 116–28). Moral knowledge skeptics argue that the fact that there is widespread disagreement among epistemic peers about moral claims, including foundational ones, gives us reason to suspect that we typically lack moral knowledge (p. 108). May proposes that few foundational moral disputes are among epistemic peers (pp. 123–28) and more importantly, the widespread moral disagreement that skeptics envision has not been backed up by empirical data (pp. 16–123). In fact, several empirically informed projects, including the moral foundations literature suggest that there is actually a lot more agreement about foundational moral claims than one might think (pp. 120–23).

I do not really have much disagreement with the way May is approaching this argument, or even with where he ends up. I think he is right that the kind of moral disagreements we see, between peers, are not sufficient to warrant widespread general moral skepticism as proposed by the moral skeptics, but that there is sufficient disagreement that we should adopt a limited moral skepticism (p. 130). And I agree with May that there is

reason to be optimistic that empirical threats have not uncovered widespread fundamental flaws in ordinary moral deliberation and judgment that do not also apply more generally to cognition itself. That said, I want to press two issues. First, moral knowledge skeptics make the claim that “there is a lot of peer disagreement about foundational moral claims” (p. 117) and whether this premise is consistent with the available evidence or not depends upon on what is meant by a moral “claim” and what exactly a moral “foundation” is. I doubt that these are the same things. May is certainly treating them as if they are but we should be more reluctant to make that claim. Second, I agree with May that what is warranted is limited skepticism. May takes his limited skepticism to recommend optimism about moral knowledge, in general, but I think we should be more cautious. If general, albeit limited skepticism is justified by the available data, then we know significantly less about morality than we thought we did.

May addresses the question of whether there actually is disagreement among epistemic peers about foundational moral propositions by looking in a very reasonable place: at the difference between conservatives and liberals and by looking at what, at first glance, appears to be the relevant empirical data, the moral foundations literature in psychology. He takes the moral foundations literature to demonstrate that within a society there is little disagreement about fundamental moral propositions between epistemic peers. This is because all five of the moral foundations (care, fairness, loyalty, authority, and sanctity) appear to be used by both conservatives and liberals. But, it seems like the moral foundations literature could be interpreted as supporting the claim that there is more fundamental disagreement here than May suggests. Jonathan Haidt, for example, claims that the way these foundations are weighted matters. Although May notes this, it seems reasonable to arrive at a different conclusion than May does, given these different weightings.

It is true that Haidt goes to pains to demonstrate that liberals do use all five moral foundations, differing from conservatives in that they merely target different issues (Haidt 2012, p. 179). This could support May's interpretation. But Haidt also argues that even though liberals and conservatives do seem to use all five of these foundations, liberals tend to consciously disown the use of the sanctity, authority, and the loyalty foundation (Haidt 2012, pp. 186–87). Liberals only acknowledge care and fairness as legitimate foundations for morality. They may use loyalty, authority, and sanctity when forming actual moral beliefs, but they do not consciously acknowledge this and according to Haidt, they typically will disown using these foundations (Haidt 2012, p. 179; Haidt 2016, p. 208). By contrast, conservatives are comfortable acknowledging that they do use all five foundations explicitly.

So, here is the worry. Perhaps conservatives and liberals do have a deeper disagreement about foundations than May's reading suggests. Maybe the liberal tends to see three of these foundations as more akin to cognitive biases than as legitimate foundations for morality whereas the conservative accepts all five foundations as legitimate. Assuming some liberals and conservatives are epistemic peers, we should then worry that there is fundamental disagreement within a society between epistemic peers about moral foundations. Some people think that only care and fairness are legitimate foundations for morality and others think sanctity, loyalty, and authority are equally compelling moral foundations.

Moving on, May is trying to assess whether or not there is widespread disagreement about foundational moral judgments (i.e., beliefs or propositions) within North America, particularly

between liberals and conservatives. But the moral foundations work he focuses on seems to be addressing different kinds of questions. Specifically, the moral foundations research seems mostly focused on explaining how we process information and how we arrive at judgments when it comes to morality. These are questions about moral cognition, not the frequency of various foundational moral beliefs within a population. I am not sure we can infer shared foundational moral propositions from shared moral cognition.

Jonathan Haidt argued that most moral judgments are not the result of careful, rational reflection. On his social intuitionist model, our moral deliberation starts with intuitions (Haidt 2012, p. 5). We have an automatic intuitive, response toward an event or scenario, this leads us to judge that said event is wrong or right/ good or bad, and then we provide a post hoc justification for why we have the judgment we have (Haidt 2012, pp. 55–60). These intuitive responses, not reason, are the foundation for morality, according to Haidt (pp. 103–108).

Haidt goes on to propose that these intuitions are innate and universal moral foundations (p. 130), which he takes to be specialized cognitive modules that evolved to address humanity's shared adaptive challenges (we need to care for children which led to a caring cognitive module, we need to form partnerships with non-kin which led to the fairness module, avoiding disease resulted in the sanctity module, etc.) (p. 146).

The important thing to note is that moral foundations are not moral propositions, at least not for Haidt. To claim that we use care as a moral foundation is not to claim that most people agree that we should care for our young or that we should care for each other. It is to simply say that we are built such that we do care for our young and what enables us to do that is we have certain kinds of emotive responses to the suffering of others particularly those we feel close to. This need not imply a commitment to any specific statements or positions or views on morality. It is not that many people value caring or place a premium on loyalty or sanctity. It is rather that we have evolved specific mental modules that are implicated when we form moral judgments and responses to the world.

Suppose this is an accurate model of what's going on in moral deliberation. This does not address the issue of whether or not people typically share the same foundational beliefs about morality. What it tells us is something about moral cognition, about how our brains work when we consider a moral issue. What we would need to be able to demonstrate as a result of the moral foundations project is that there is something about the way this processing occurs that leads us to be optimistic that, in general, there is significant agreement about foundational moral propositions.

Although this is interesting research, it does not seem to be the right kind of data to address whether or not there are commonly shared foundational moral propositions among epistemic peers. What we have is an account of moral cognition, but what we need to know to address the second premise of the skeptical argument is whether or not there is common agreement about foundational moral propositions among epistemic peers.

What we are looking for, or should be looking for, I would think, are whether or not there is sufficiently widespread agreement about foundational metaethical and normative ethical principles, to allow for agreed upon foundational propositions among epistemic peers. The moral foundations literature may give us reason to be optimistic that there could be, as it suggests, that similar mental processes are implicated in moral judgment. But it does

not demonstrate agreement about foundational moral propositions. We might be able to use this literature in the way that May proposes, but what we would need to do is provide an argument demonstrating that this shared brain machinery implies that most of us do share similar foundational moral propositions.

What we really need is more clarifying work on what a moral foundation is, how moral foundations operate, and how much convergence there is in regards to the general foundational moral judgments or claims that individuals arrive at within a society.

To close, then, May is engaged in a valuable project but we need to go further. What we need is data addressing whether or not there is sufficient agreement about foundational moral propositions within a culture, not whether or not most humans are working with the same mental mechanisms when we engage in moral cognition. Until data of this sort is generated, we should adopt moderate skepticism toward moral agreement within a society. But moderate skepticism is grounds for withholding judgment as to whether or not we have widespread agreement about moral foundations. I agree with May that it is not grounds for pessimism, but it is no more grounds optimism.

Valuation mechanisms in moral cognition

Julia Haas

Department of Philosophy, Rhodes College, Memphis, TN 38112.
Haasj2@rhodes.edu www.julishaas.com

doi:10.1017/S0140525X18002686, e155

Abstract

May cites a body of evidence suggesting that participants take consequences, personal harm, and other factors into consideration when making moral judgments. This evidence is used to support the conclusion that moral cognition relies on rule-based inference. This commentary defends an alternative interpretation of this evidence, namely, that it can be explained in terms of *domain general valuation mechanisms*.

In *Regard for Reason in the Moral Mind*, Joshua May (2018) argues that our “moral judgments are often governed by rule-based inference” (p. 55). Here, inference is understood as a mode of reasoning in which “beliefs or similar propositional attitudes are formed on the basis of pre-existing ones” (p. 8). To support this view, May cites a body of evidence suggesting that participants take consequences, personal harm, and other factors into consideration when making moral judgments. On the basis of this evidence, May concludes that, in making moral decisions, “we often rapidly infer the moral status of an action in part by relying on general principles” (p. 70).

The evidence May cites admits of at least two interpretations, however. On May's interpretation, the transition from consequences, intentions, and other factors to moral judgments is underwritten by an inference from general principles. On an alternative interpretation, moral cognition is instead underwritten by non-inferential, subpersonal, *domain general valuation mechanisms*.

This commentary defends the latter interpretation. Moreover, it suggests that, insofar as May fails to consider this latter possibility, the book's argumentative strategy is weakened. May's argument proceeds by elimination: He aims to show that the evidence regarding the emotions is not as compelling as generally thought, and hence that "what's left is inference" (p. 71). But if emotion and reason are not the only explanatory options to choose from – if valuation mechanisms are a viable, interpretive possibility – then we should not be so quick to conclude that "moral judgments are generated by fundamentally cognitive and rational processes" (p. 18).

Roughly, domain general valuation mechanisms refer to a body of computational, neural, and behavioral mechanisms thought to underwrite decision making and economic choice (for an excellent overview, see Glimcher et al. 2009). On one way of describing these mechanisms more precisely (in what is called "reinforcement learning"), decision making is underwritten by multiple distinct decision systems, each relying on a distinct computational approach to predicting future reward and value (Sutton and Barto 1998). Here, reward refers to the intrinsic desirability of a given state, whereas value refers to the total, expected, future reward of a given state. The distinction allows for a state that is not intrinsically rewarding to nonetheless be assigned value, contingent on its causal relations to future rewarding states.

Human decision making is thought to depend on at least three distinct, domain general valuation mechanisms (for reviews, see Dayan & Abbott 2001; Dayan & Niv 2008; Rangel et al. 2008). A *Pavlovian* valuation mechanism governs automatic approach and withdrawal responses to appetitive and aversive stimuli, respectively (Mackintosh 1983). A second, *model-free* valuation mechanism caches positive and negative state-action pairs, and assigns values to actions based on their previous outcomes. A third, *model-based* mechanism explicitly represents possible choices and determines the sequence of actions that maximizes value. This procedure is typically represented by a decision tree: Each node in the tree representing a possible choice, where total value is the sum of the rewards minus the sum of the punishments along a given branch. Notably, however, none of these mechanisms involve rule-based inferences or an appeal to general rules. Rather, as their name suggests, domain-general valuation mechanisms rely on a basic notion of value and the calculation of value to underwrite our everyday decision making.

These (and possibly other) domain general valuation mechanisms are increasingly thought to play a role in moral cognition. Taking a computational approach, for example, Crockett (2013; 2016a; 2016b) argues that these mechanisms – and their interactions – can account for how the features of a moral situation can be transformed into a moral decision and, by extension, a moral action. Analogously, but adopting a more explicitly neuroscientific approach, Shenhav and Greene (2010 p. 671, Table 1) show that participants' ratings of moral acceptability are correlated with degrees of activation in their posterior cingulate cortex and ventromedial prefrontal and medial orbitofrontal cortices, that is, with brain activations relatively similar to those seen in instances of valuing physical goods and actions. These mechanisms have further been implicated in the processing of normative statements (Berns et al. 2012), experiences of trust (Fehr et al. 2005), the feeling of compassion (Montgomery et al. 2017), and the phenomenon of implicit bias (Huebner 2016).

So how might these domain general valuation mechanisms explain the specific findings May cites in favor of reasoning in moral cognition? I focus on the components of *consequences*, *personal harm*, and the *means versus by-products distinction*, though

analogous arguments can be made for May's discussion of intentions and actions versus omissions (see especially Crockett 2013; Cushman 2013; though, see also Ayars 2016)

First, May (p. 58) argues that ordinary moral judgments are "sensitive to the quantity of harmful consequences that follow from an agent's options," including relatively complex considerations, and takes these to be evidence of the role of rule-based inference as characterized previously. However, domain general valuation mechanisms can equally account for such assessments of consequences, and notably can do so without appealing to general principles or rules. Specifically, both the model-free and model-based mechanisms characterized above predict instrumental state-action pairs; that is, they predict the consequences of various action alternatives as selected in various future states. Similarly, though again using a more neuroscientific emphasis, in the study cited above, Shenhav and Greene (2010) demonstrate that participants make surprisingly fine-grained assessments of consequences – even taking uncertainties and risks in account – using precisely those domain general valuation systems that underwrite other types of decisions, including economic decisions.

Second, May argues that principles relating to harm, notably principles relating to force, contact, and battery, play an important role in moral cognition. As above, the suggestion seems to be that such assessment requires rule-based inference. May suggests that although there is not precision in this literature, a quip from Paul Bloom can characterize the problem, suggesting, "Here is a good candidate for a moral rule that transcends space and time: If you punch someone in the face, you'd better have a damn good reason for it" (Bloom 2013, p. 10). However, as in the case of consequences, assessments of personal harm, including such factors as force and contact, can be accounted for by using domain general valuation mechanisms.

Cushman (2015) makes just such a case, suggesting that the model-free system can explain why participants provide inconsistent responses to the trolley problem. On Cushman's view, the model-free system elicits aversion toward direct, physical harm, where this aversion "can be understood as the consequence of negative value assigned intrinsically to an action," namely, the physical harm (Cushman 2015, p. 59). This learned aversion in turn prevents participants from endorsing the pushing of the bystander in the "Footbridge" case, but still allows them to endorse the pulling of the switch in the "Switch" case. Notably, again, however, such an assessment depends on the cached attribution of value rather than on an appeal to general principles.

Third, May argues that moral cognition involves making a sophisticated distinction between an outcome that is a direct means to one's end and one that is merely a by-product of one's endeavors. Although the empirical results are, by May's own account, rather mixed on this front, the basic implication of May's argument remains the same: moral cognition depends on inference and, in particular, on some set of morally relevant general principles. Again, however, domain general valuation mechanisms may do the trick without appealing to principles at all.

Crockett (2013) cites a special kind of interaction between the Pavlovian and model-based systems to defend this latter possibility. Specifically, Crockett appeals to the phenomenon of "pruning," in which the Pavlovian system "cuts" the branches of a model-based decision tree in the face of aversive alternatives represented early on in the decision tree. Crockett argues that in means cases, as when an individual is directly used to stop a trolley, the harm – killing the individual – is sufficiently high up on the tree that this alternative is "pruned," causing participants to

deem it an unacceptable alternative. By contrast, in by-product cases, as when a trolley avoids killing five individuals but *then* kills one individual, the benefits of saving the five occurs before any pruning can take place. Consequently, “the side-effect death is incidental to saving the five individuals, so it can be safely pruned away while preserving the contribution of saving five toward the overall action value” (Crockett 2013, p. 365). Domain general mechanisms thus provide a plausible explanation of the means versus by-product distinction in moral cognition.

To summarize: Based on his assessment of these and the other factors, May concludes that “a clear picture is emerging from the science of moral judgment. We often rapidly infer the moral status of an action in part by relying on general principles that identify as morally relevant various features of agents, actions, and outcomes” (p. 70). However, moral cognition may instead involve the assessment of consequences, personal harm, and means versus by-product effects (as well as the other factors) *without* a cognitive appeal to general principles. Moreover, as noted above, insofar as May aims to provide an argument by elimination, May’s case for moral reasoning is not as strong as it might first appear.

Moral judgment as reasoning by constraint satisfaction

Keith J. Holyoak^a  and Derek Powell^b

^aDepartment of Psychology, University of California, Los Angeles, CA 90095-1563;

and ^bDepartment of Psychology, Stanford University, Stanford, CA 94305-2130.

holyoak@lifesci.ucla.edu derekpowell@stanford.edu

http://reasoninglab.psych.ucla.edu http://www.dereempowell.com

doi:10.1017/S0140525X18002546, e156

Abstract

May’s careful examination of empirical evidence makes a compelling case against the primacy of emotion in driving moral judgments. At the same time, emotion certainly is involved in moral judgments. We argue that emotion interacts with beliefs, values, and moral principles through a process of coherence-based reasoning (operating at least partially below the level of conscious awareness) in generating moral judgments and decisions.

May (2018) makes a compelling empirical case that reason, not emotion, is the primary causal factor driving human moral judgments. Of course, many philosophers (some predating Kant by a couple of millennia) have similarly considered the essence of moral judgment to be a matter of correct *understanding*. In the *Analects*, Confucius issued a critique that might well be applied to modern capitalism when he observed, “The superior man understands what is right; the inferior man understands what will sell.” But the sentimentalism against which May argues has attracted its own strong proponents (not all of them philosophers or moral psychologists). Ernest Hemingway laid out a simple test: “what is moral is what you feel good after and what is immoral is what you feel bad after.” Based on that subjective criterion (from *Death in the Afternoon*), Hemingway was able to attest to the moral rightness of bullfighting – for after the fight ends in

the usual way, “I feel very sad but also very fine.” According to the great novelist’s moral emotions, the artistry and allegory more than compensate for the dead bull and the dying horses.

The sentimentalist’s account of moral judgment has an attractive simplicity – “Theirs not to reason why, theirs but to laugh or cry.” Moreover, few doubt that emotions play *some* role in moral decision making – even Kant (1785/2002) recognized that “sympathies” and “sentiments” are integral to proper moral functioning. The difficult question, which May tackles head on, is to determine how emotion and reason relate to one another in the context of moral judgment, and to assess their relative importance. We have argued previously (Holyoak & Powell 2016) that theories of moral psychology have often been premised on outmoded conceptions of both emotion and cognition. Rather than being strictly separable processes, modern work in both psychology and neuroscience has emphasized their intricate interactions. In social psychology, appraisal theories postulate that emotions are caused by processes in which stimuli are evaluated on such cognitive dimensions as goal relevance, coping potential, and agency (Moors et al. 2013). At the neural level, compelling evidence indicates that emotion and cognition interact in the prefrontal cortex (Pessoa & Pereira 2013), where cognitive and emotional signals appear to be combined in complex ways.

In the case of cognition, an outmoded conception is that thinking consists solely of the conscious application of deterministic rules. Current cognitive theories differ widely, but the dominant overarching view is that both inductive and deductive reasoning are largely based on forms of probabilistic inference (e.g., Cheng 1997; Griffiths et al. 2008; Oaksford & Chater 2013). Probabilistic inference supports structured and systematic reasoning even when grappling with highly uncertain beliefs, premises, or observations. Probabilistic cognitive models also suggest how even simple intuitive judgments that do not draw on explicit or conscious deliberation (e.g., will a block tower fall or be stable?) might be supported by complex and highly structured knowledge, such as an intuitive theory of physics encompassing Newtonian mechanics (Battaglia et al. 2013; see Kubricht et al. 2017). Moreover, high-level human abilities such as creative problem solving (Holyoak 2019; Kounios & Beeman 2015) depend on complex interactions between some processes that depend on conscious attention and working memory, and others that depend on unconscious activation of neural networks distributed throughout the cortex (Knowlton et al. 2012).

As May recognizes, a dual-process conception of moral reasoning that posits a strict separation between an unconscious emotional system (identified rather oddly with the philosophical position of deontology) and a conscious system for rational reasoning (supposedly dedicated to the computation of utilitarian outcomes) ignores the evidence for emotion/cognition interactions, as well as for unconscious aspects of reasoning. May suggests that dual-process theorists might be better off drawing a distinction between fast and intuitive versus slower and more deliberative *cognitive* processes (neither necessarily dependent on emotion). We would press the point further, and suggest that the popular notion of dual-process models is itself simplistic. Even dual-process theorists are uncertain about the nature of the two processes, or indeed their number (Evans 2009). The fast/intuitive versus slow/deliberative distinction provides a useful shorthand to mark the extremes of a continuum, but most complex cognitive abilities – including moral judgment – are likely to be based on multiple, integrated mechanisms that quickly blur any binary division.

May briefly considers the possible role of *consistency* or *coherence* in resolving moral issues that involve conflict or ambiguity. Coherence-based reasoning is a domain-general mechanism that applies to moral reasoning as a special case. Its operation has been observed in a variety of complex decisions in which moral issues arise, such as legal cases (Holyoak & Simon 1999; Simon 2012), attitudes to war (Spellman et al. 1993), and attributions of blame and responsibility (Clark et al. 2015). A key property of coherence-based reasoning is that values, beliefs, and emotions may change to increase their coherence with the emerging decision (contrary to the usual assumption that these core elements are typically fixed over the course of a reasoning episode). The outcome of decision making is not simply the choice of an option, but rather a restructuring of the entire package of values, attitudes, beliefs and emotions that relate to the selected option (for reviews see Simon & Holyoak 2002; Simon et al. 2015).

Considerable evidence indicates that moral judgments are often based on a process of constraint satisfaction that is directed at achieving local (and perhaps transient) coherence (Holyoak & Powell 2016). Coherence-based reasoning could be applied to adjudicate among competing moral principles. To take two examples from those suggested by May, a person might value the Consequentialist Principle, “All else being equal, an action is morally worse if it leads to more harm than other available alternatives” (p. 57); but also the Principle of Agential Involvement, “All else being equal, it is morally worse for an agent to be more involved in bringing about a harmful outcome” (p. 69). The latter has a decidedly deontological flavor, and both will typically be based in part on a causal analysis of the situation (Lagnado & Gerstenberg 2017; Waldmann & Dieterich 2007). The “all else being equal” implies that neither principle is absolute (and indeed, they are simply two among many). Depending on the relative strengths of these and potentially many other competing principles, coherence-based reasoning may lead to different judgments about what is the right course of action (see also Zamir & Medina 2010).

At the same time, whenever a set of factors leads someone to render a judgment in conflict with a given principle, coherence-based reasoning implies that the strength of that principle may be reduced (Horne et al. 2015). Fluidly shifting one’s moral principles in this way might seem decidedly unprincipled, but the drive to achieve coherence in one’s moral beliefs is of a piece with Rawls’ (1971) notion of “reflective equilibrium.” Through coherence-based reasoning, such equilibria are sought dynamically and potentially unconsciously during the course of moral decision making.

Coherence-based reasoning is consistent with the thrust of May’s empirical debunking of the various lines of argument themselves intended to debunk the role of reason in moral judgment. Many philosophers have sought to debunk commonsense moral beliefs (i.e., to argue that those beliefs are unjustified) by arguing that the grounds or processes by which those beliefs are formed are unsound. May argues that those seeking to undermine the reliability of human moral judgments on psychological grounds invariably find themselves on one of the horns of the “debunker’s dilemma”: either the purportedly corruptive process backing those moral beliefs actually proves reliably informative in some circumstances, or it turns out that the impact of the corruptive process on moral judgments and beliefs is weak enough to be quite inconsequential. For example, May argues that far from leading us astray, emotions are often highly informative in moral situations. On the other hand, where emotions are incidental, their influence is generally exceedingly small (also see Horne & Powell 2016).

Coherence-based reasoning may explain why would-be debunkers are left facing May’s dilemma. Generating judgments by constraint satisfaction allows reasoners to incorporate a diverse set of factors into their decision-making process while constraining the influence of any one of those factors. For instance, coherence-based reasoning can enable emotions to influence moral judgments yet not entirely override other relevant factors; this reasoning mechanism can also alter emotional responses to a situation based on the emerging judgment (Simon et al. 2015).

The picture of moral reasoning that emerges from May’s arguments, and in particular from his critique of sentimentalism, is quite the opposite of the kind of encapsulated, special-purpose “module” some evolutionary psychologists have envisioned (for a discussion, see Bolhuis et al. 2011). Rather, all the mechanisms that impact judgment and decision making in non-moral domains – including those characterized as heuristics and biases – guide moral judgments as well (e.g., Rai & Holyoak 2010). Human reason, with or without inputs from emotion, is certainly fallible. But May aptly quotes Kahneman (2011, p. 4), who observed that, “the focus on error does not denigrate human intelligence, any more than the attention to diseases in medical texts denies good health.”

May’s renewed focus on the centrality of reason in moral judgment suggests that morality should be included, along with language and high-level thinking, on a short list of domains that lie at the core of what it means to be human (see Penn et al. 2008, for discussion of thinking, and Wynne & Bolhuis 2008, for a discussion of morality). The claim that a sense of morality is distinctively human is of course controversial. Contemporary comparative psychologists (often relying on anthropomorphism) routinely report finding evidence of moral motives in non-human animals, such as chimpanzees’ apparent concern for the equitable distribution of rewards (e.g., Brosnan et al. 2005). But when put to critical tests, simpler explanations have been found for many of these behaviors (e.g., Engelmann et al. 2017).

If morality is indeed a type of specifically human cognition, aligned with language and abstract thought, the common thread linking them may well be the requirement to be able to explicitly represent and think about higher-order relations. May describes empathy as involving a kind of “relational desire” – for example, the wish to ease a pain one feels by easing that of another. More generally, morality begins when one understands the values of others to whom we are related in some specific way – as relatives, fellow citizens, humans, or perhaps sentient beings – and makes concern about the values of these others a part of one’s own values. This is the crucial step that renders the life and well-being of another one’s own concern. As Aristotle observed in *Nicomachean Ethics*, it is also a crucial step toward friendship: “The best friend is the man who in wishing me well wishes it for my sake.” Perhaps reason is the bedrock of the most distinctively human emotions.

What is sentimentalism? What is rationalism?

Antti Kauppinen 

Department of Practical Philosophy, University of Helsinki, 00014 Helsinki, Finland.

antti.kauppinen@helsinki.fi <http://anttikauppinen.weebly.com>

doi:10.1017/S0140525X18002649, e157

Abstract

May argues successfully that many claims about the causal influence of affect on moral judgment are overblown. But the findings he cites are compatible with many of the key arguments of philosophical sentimentalists. His account of rationalism, in turn, relies on an overly broad notion of inference, and leaves open crucial questions about how we reason to moral conclusions.

In the first part of *Regard for Reason in the Moral Mind*, Joshua May (2018) mounts a bold defense of a form of moral rationalism against sentimentalism. But what exactly is his target, and does he offer a credible alternative?

As I have observed previously (Kauppinen 2013b), sentimentalism comes in many logically independent forms, in which emotions or more broadly pro- and con-attitudes play different roles. *Explanatory sentimentalists* hold that sentimental reactions fundamentally explain our moral verdicts; *judgment sentimentalists* hold that moral judgments consist in sentiments or otherwise make essential reference to sentiment; *metaphysical sentimentalists* hold that moral properties are grounded in actual or possible sentimental responses; and *epistemic sentimentalists* hold that we come to know moral truths ultimately by way of sentimental responses. Sentimentalists offer different sorts of *a priori* arguments for these claims, appealing, for example, to the apparent importance of attitudes that have a world-to-mind fit in explaining the action-guiding character of moral thought. Recently, some sentimentalists, most notably Jesse Prinz (2007) and Shaun Nichols (2004), have also offered *a posteriori* arguments for these views, drawing on scientific findings.

It is the *a posteriori* arguments that are May's main target, although he merely points to arguments of others when it comes to *a priori* sentimentalism. This is worth emphasizing for two reasons. First, though May mounts a very promising case against the *a posteriori* arguments, we may nevertheless have sufficient reason to subscribe to a sentimentalist view on *a priori* grounds. Second, I think it is fair to say that what defines the various sentimentalist views are the conclusions of the *a priori* arguments. Only explanatory sentimentalists, for example, are committed to *causal* claims about the role of emotion in generating moral judgments, and these claims are sometimes significantly weaker than May's targets. Adam Smith, for example, holds that "the greater part of our moral judgments [...] is regulated by maxims and ideas derived from an induction of reason," while arguing that it is "absurd and unintelligible to suppose that the first perceptions of right and wrong can be derived from reason" (Smith 2002, p. 377). On this kind of view, emotions do not play a causal role in every moral judgment, but rather explain why we find certain *act-types* right or wrong in the first place. According to even more modest social transmission views, emotions play a causal role in explaining why certain patterns of moral judgment prevail and get transmitted (Kauppinen 2014; Nichols 2004). Assuming that people pick up their moralizing tendencies from others, this view entails that emotions ultimately (but indirectly) explain even the judgments of those who never respond emotionally.

The evidence May adduces in chapter 2 against exaggerated claims about the causal influence of emotion on moral judgment is compatible with a view like Smith's being true. And of course it does not bear on other varieties of sentimentalism, which

make no causal claims in the first place. The best kind of evidence against a Smithian sort of explanatory sentimentalism would show that there are individuals who lack the postulated kind of sentiments altogether, but nevertheless make genuine moral judgments. The closest results in this respect come from studies on psychopaths – but alas, it is far more ambiguous, because psychopaths do have emotions (even if abnormal), and there is active debate on whether their moral judgments are genuine (see, e.g., Smith 1994). And the social transmission view is of course not committed to the claim that emotions directly explain the judgments of particular individuals, so it is not necessarily threatened even if psychopaths know perfectly well what is right or wrong.

Why hold on to even modest explanatory sentimentalism, however, if the observed effects of emotional manipulation are as weak as May argues? Perhaps the most convincing argument is based on the close parallel between independently evolved emotional tendencies and widely accepted moral principles. There is an extremely plausible adaptive rationale for the tendency of social animals like us to have negative emotional responses to actions like cheating, failing to reciprocate, insulting, and grabbing a share of resources that is disproportionate to one's contribution (e.g., Sober and Wilson 1998). Other primates have analogous responses, which lends additional credence to the claim that they are independent of moral judgment. Yet there is a striking parallel – even if not an exact correspondence – between these adaptive emotional tendencies and widespread patterns of moral judgment (e.g., Boehm 2012). Some use such facts as a premise in a debunking argument of moral beliefs (Street 2006), but that is not the sentimentalist claim. The explanatory sentimentalist contention is that the parallel is best explained by the fact that moral judgment is deep down driven by emotion, though competing accounts differ on the details of just how this happens. David Hume (2006, p. 260), for example, emphasizes the need to correct for bias in our untutored responses for morality to perform its social function. (This would explain why there is only a parallel, not an exact correspondence.) For the rationalist, in contrast, the parallel between adaptive emotion and moral judgment is a coincidence: Reason just happens to tell us to disapprove of the very things we in any case tend to feel negatively about, at least when we are ourselves at the receiving end. This comparison does not flatter the rationalist.

So far, I have focused on what sentimentalism is and what it is not. Let us now turn to rationalism, as May understands it. His claim is that "moral judgment is fundamentally an inferential enterprise that is not ultimately dependent on non-rational emotions, sentiments, or passions" (p. 7). May relies here on an extremely broad conception of inference, which includes "unconscious, unreflective, or implicit processes that nonetheless amount to reasoning" (p. 55). But he acknowledges that not *every* transition among beliefs (or other contentful states) amounts to reasoning (p. 9). Otherwise rationalism would be devoid of distinctive content.

What is reasoning, then? Here we must bear in mind that bad reasoning, too, is a kind of reasoning, so we cannot appeal to what are in fact genuine requirements of rationality (Broome 2013). It is common to hold that at least the following elements are necessary: doxastic states whose contents serve as premises, doxastic or conative states whose contents express the conclusion, and some form of endorsement of the move from the premises to the conclusion, such as tacit acceptance of a pertinent rule of inference or taking the conclusion to follow from the premises (Boghossian 2014).

Although this minimalist account is compatible with non-conscious reasoning, many of the computational mental processes that May argue play a role in moral judgment do not qualify as inference by its lights, because any kind of inference requires both premise-beliefs and somehow basing the conclusion on their content. For example, May holds that *categorization* of an ordinary object as a piece of furniture involves *inference* from a belief or belief-like state like “This objects resembles sofas, chairs, and tables” (p. 70). However, this is a non-starter as an account of categorization, as the very same (non-inferential) recognitional capacity that allows us to categorize something as furniture is required to make the judgment that it *resembles* items in the furniture category. If we can perform the latter without inference (and surely there are *some* such judgments on anyone’s view), there is no reason to think unconscious inference must be involved in the former. Similar considerations hold for high-level perception (e.g., Audi 2013), like the perception that someone is on drugs – we can be sensitive to complex information without any kind of inference from premises to a conclusion.

The same goes for moral categorization: there is no evidence for a necessary inferential step. Curiously, May half-acknowledges that the evidence fits the alternative hypothesis that our principles like the Doctrine of Doing and Allowing merely *describe* the pattern of our moral judgments rather than *guide* them (p. 70). What makes this only a half-acknowledgment is that he describes this in terms of “reasoning in accordance with” the principle. But that our judgments accord with a principle is no evidence at all that they result from *reasoning* – indeed, if it is acknowledged that the principle does not guide our reasoning, it would be a miracle of sorts if reasoning guided by some *other* rule yielded the *same* output in every case.

Second, even if we were to accept May’s broad notion, the evidence he cites only shows that inferences about *non-normative* facts, such as the extent to which the agent was involved in bringing about the outcome, play a role in moral judging. This is something that sentimentalists *accept*. Already Hume emphasized that although sentiment renders the final verdict, “in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn” (Hume 2006, p. 189). So even on the arch-sentimentalist Hume’s view, it is not only true that moral sentiments are “sensitive to information” (p. 74), but also that they sometimes require conscious reasoning about non-normative facts. What he and other sentimentalists deny is simply that this *suffices* to explain or justify moral judgment, because there is a gap between non-normative and normative conclusions. On their view, emotions do not just “facilitate” inference by directing attention, but either fundamentally explain or justify crossing the gap. Unless May shows that the process that takes us from non-moral premises regarding, say, intentions and consequences, to moral verdicts is distinctively rational, his view is importantly *incomplete*.

Finally, and related to the previous point, any process of inference must begin from premises, which on pain of regress cannot always be justified by further inference. Take the following simple piece of (good) reasoning:

1. Clinton lied.
2. Lying is wrong.
3. So, Clinton did something wrong.

No one denies that it is possible to reason from premises 1 and 2 to conclusion 3, and thereby gain justification to believe 3, if one is

justified in believing the premises. But what justifies belief in premise 2? (Let us assume for simplicity that it is true.) On pain of a different regress, the answer cannot be “testimony.” So traditional intuitionists say, roughly, that it is *self-evident*: Anyone who understands the content thereby has justification to believe in it (Audi 2013). Many epistemic sentimentalists say, roughly, that it is a legitimate inductive generalization from the contents of emotional responses, such as resentment, that *present* particular acts of lying as wrong (e.g., Tappolet 2016). These are both the right kind of answers in that they do not appeal to further premises. May does not argue against such views. But more importantly, while he discusses evidence that we engage in reasoning *from* moral principles, I was unable to find any discussion of how we reason *to* moral principles, although he acknowledges the need in passing (p. 79).

To sum up, May tends to construe sentimentalism extremely thinly, as a claim that moral judgments are explained by or consist in purely non-cognitive *feelings*, and rationalism extremely broadly, as something like the claim that moral judgments are sensitive to information about their targets. On such construals, it is easy to declare rationalism as the better theory. But as I have tried to sketch here, at least when it comes to philosophy, both of these characterizations are ill-fitting. More work is needed to refute the arguments that sentimentalists actually make, and to develop a credible rationalist alternative.

It is thus fortunate that most philosophical sentimentalists from Hume and Smith onward are no less optimistic than May. They hold that as long as there is “some particle of the dove, kneaded into our frame, along with the elements of the wolf and serpent” (Hume 2006, p.259), we will approve of just and benevolent actions, constrain our egoism in virtue of internalizing the reactive attitudes of actual or imagined others, and make moral progress by reasoning about non-moral facts before rendering our judgment and by extending our natural empathy beyond our immediate circle. Doesn’t it warm your heart just to think about it?

What sentimentalists should say about emotion

Charlie Kurth 

Department of Philosophy, Western Michigan University, Kalamazoo, MI 49008-5328.

charles.kurth@wmich.edu www.charliekurth.com

doi:10.1017/S0140525X18002601, e158

Abstract

Recent work by emotion researchers indicates that emotions have a multilevel structure. Sophisticated sentimentalists should take note of this work – for it better enables them to defend a substantive role for emotion in moral cognition. Contra May’s rationalist criticisms, emotions are not only able to carry morally relevant information, but can also substantially influence moral judgment and reasoning.

What sentimentalists should say about emotion

Not every form of sentimentalism is plausible, and Josh May’s (2018) book shows that there is reason to doubt some recent,

prominent formulations. But it does not follow from this that we should be rationalists. Rather, I believe that May's criticisms help us see what a better sentimentalist metaethic should look like. More specifically, investigating what a sentimentalist should say about the nature of emotion reveals that emotions play a more significant role in moral cognition than May presumes. A sophisticated sentimentalism thus remains an important rival to rationalism.

Emotions for sentimentalists

As May sees it, sentimentalists face a dilemma. If emotions are just non-cognitive feelings, then they play no substantive role in moral cognition. By contrast, if emotions are partly cognitive (i.e., belief-like states), then the substantive work that they do in moral thought is best explained by their cognitive – not sentimentalist – features (pp. 51–52).

In response, sentimentalists should reject the picture of emotion that May's dilemma presupposes. At a gloss, emotions are intentional mental states with evaluative content. To be angry about a comment is to see that comment as an affront – as something that calls for a response; to feel compassion toward another is to see her as suffering – as someone to be helped. Pushing deeper, sentimentalists should follow emotion researchers in seeing emotions as states that involve multilevel content and processing (e.g., Griffiths 2004; Izard 2007; Kurth 2018, Ch. 2; Levenson et al. 2007).

At a low-level, emotions have course-grained, non-conceptual evaluative content that is intimately tied to feeling and action. So, for example, to feel angry is to experience the actions of another as *challenge-to-standing-bad*; feelings of shame convey something like *social-rank-asymmetry-bad*; compassion presents its target as *another-suffering-bad*. Here the hyphenated strings are gestures toward the distinctive, motivationally laden, evaluative dimensions of these emotions' low-level, non-conceptual content.

At the high-level, an emotion's distinctive evaluative content is both fine-grained and conceptual in a manner that facilitates their use in reasoning. So, for example, anger toward a comment presents that comment as, roughly, an affront to one's (moral) standing. With shame, one sees oneself as having failed to live up to an ego-ideal. In both cases, the high-level conceptual content facilitates inferences about (respectively) being wronged and one's social-moral inferiority.

Importantly, a single emotional experience (e.g., a token of anger) will typically engage *both* types of content and *both* levels of processing (Griffiths 2004; Kurth 2018; Wringer 2015). Moreover, although the two channels of emotion content/processing generally perform complementary – though distinct – functions, they can come apart in ways that lend support to the above picture.

Consider, for example, experimental work on “repressors.” When these individuals are presented with a threatening stimuli, they display the attentional and physiological changes associated with fear – but they *deny* being afraid. What we appear to have, then, is a dissociation of low- and high-level emotion processing: while the low-level processing of repressors generates the action-oriented attentional shifts and physiological responses characteristic of fear, their high-level processing fails to categorize the situation under the relevant concept (FEARSOME OR DANGER). Hence, they deny feeling the fear that they otherwise seem to be experiencing (Derakshan et al. 2007; Kurth 2018, pp. 58–59).

Notice as well that emotions are not unique in being mental states with multilevel content/processing of this sort. Work in vision science, for instance, indicates that the content of visual

perception is the upshot of two distinct channels: one (the ventral) that is involved in the perception of action and another (the dorsal) that is tied to memory and speech-processing. As with emotions, although visual perception typically combines these two sources of content as part of a unified visual experience, the two channels can be forced apart (Aglioti et al. 1995; Wringer 2015).

In the present context, recognizing the multilevel structure of emotion is important because it opens up space for a distinctly sentimentalist thesis about the content and function of emotions. More specifically, with the above account in hand, sentimentalists can maintain that the low-level, motivationally laden, evaluative content of an emotion *grounds* the evaluative concept(s) distinctive of that emotion's high-level content. So, for example, shame's low-level content (i.e., *social-rank-asymmetry-bad*) fundamentally shapes and constrains both one's concept SHAMEFUL and shame's associated high-level content (roughly, the evaluation that I have failed to live up to an ego-ideal). Similarly, compassion's low-level content (namely, *another-suffering-bad*) fundamentally shapes and constrains one's concept of COMPASSION-WORTHY and compassion's associated high-level content (roughly, the evaluation that the target of one's compassion is enduring a serious and underserved misfortune that merits one's attention).

Crucially, the dependencies here are fundamental in a distinctly sentimentalist sense: Our understanding of the high-level evaluative concepts that are associated with emotions like shame and compassion comes by way of the motivationally laden, non-conceptual content carried by these emotions' low-level evaluations (D'Arms 2005; Izard 2007; Kauppinen 2013a). We find empirical support for this sentimentalist thesis in work on the evolutionary origins and development of emotion. For example, research in anthropology, psychology, and cognitive science provides evidence of high-level emotion content being shaped and constrained by low-level content for a range of emotions including shame (Fessler 2007), fear and anxiety (Kurth 2016; 2018; Öhman 2008), and disgust (Tyber et al. 2013).

Moreover, the idea that low-level, non-conceptual content can ground high-level content is not unique to emotion. Consider color. The “unity relations” (that is, the phenomena of, e.g., reds looking more similar to oranges than greens) are thought to be non-conceptual features of color experience that shape and constrain both our color concepts and high-level, color content (e.g., RED and GREEN pick out “opposites” but RED and ORANGE do not) (Cohen 2003; Johnston 1992).

The payoff: A sophisticated sentimentalism

If emotions are states of the sort sketched above, then – contra May – sentimentalism can explain how emotions are able to both “carry morally relevant information” (p. 52) and “substantially influence moral judgment” (p. 28).

Taking these in turn, first notice that emotions are concerned with fundamental human values: compassion concerns the suffering of others, shame concerns the loss of social status, anger concerns challenges to one's standing. But notice, as well that the protection and promotion of these values is at the core of what we take morality to be. If that is right, then the above sentimentalist account of the content of emotions entails that they carry morally relevant information.

May might object that this connection between emotion and morality is too indirect – although emotions might highlight morally relevant information, they are not essential for making moral

judgments (pp. 13–14). However, if the sentimentalist is right that emotions are essential to our understanding of evaluative content – grounding, for example, the distinctive badness of SHAMEFUL, the special neediness of COMPASSION-WORTHY – then May’s objection is misplaced. Acquiring evaluative concepts is not something a “sophisticated robot” could do (p. 14). At best, a robot could *approximate* emotion’s distinctive evaluative content by drawing on information provided by actual emoters (cf. Kauppinen 2013a).

Turn then to the question of emotions’ influence on moral judgment. The above account of emotions and their connection to moral/evaluative content, entails that emotions contribute to moral inferences insofar as they are essential sources of morally relevant content. Here too May is likely to protest that an influence of this sort is too thin to vindicate sentimentalism – though emotions “facilitate information processing,” they are not essential to moral inference in a deeper way (pp. 13, 71). But again notice that, on the above sentimentalist account, the low-level content of emotions is *foundational* for our understanding of the associated, high-level evaluative concepts that we use when making moral inferences. So, contra rationalists like May, moral inferences are “ultimately dependent on non-rational emotions” (p. 7).

Yet one might still worry that even if emotions are fundamental in this sense, the role that they play is still too paltry – after all, their distinctly sentimentalist-friendly low-level content only plays an indirect role in moral inference. In light of this, it is important to recognize that emotions’ low-level content also has a direct impact on moral decision making and inference.

For example, the low-level content of emotion can block the inferences and conclusions that one is brought to via explicit reasoning. Huck Finn’s deliberations told him he ought to turn Jim over to the slave hunters. But the compassion he felt for his friend interfered, preventing him from endorsing the conclusion of his reasoning (Tappolet 2016, p. 180). Additionally, emotion’s low-level content can also lead us to question the moral judgments we have made: Martin Luther King, Jr., for example, spoke of the anxiety he felt about his conclusion that it would be wrong to protest the Vietnam War – in particular, he saw his anxiety as central to his realization that his decision not to protest was mistaken (Kurth 2018, Ch. 6).

In both of these cases, the low-level content of emotion not only provides morally relevant information that was *not* captured via deliberation, but also *directly influences* these individuals’ subsequent decisions and actions.

Most significantly, emotions can be immediate, non-inferential drivers of basic moral beliefs and judgments. To draw this out, first notice that May allows that we can come to have beliefs without engaging in any (explicit or implicit) reasoning. He thinks this happens when, for instance, you immediately (i.e., non-inferentially) come to the conclusion that the door opening before you retains its rectangular shape: such a judgment is not the result of reasoning, but rather the upshot of you “simply taking your visual experience at face value” (p. 9).

But now notice that moral judgments can be formed via emotions through the same kind of immediate, non-inferential process: I immediately come to believe that I have been insulted from the anger that I feel at your comment; your judgment that the invalid needs help springs immediately from the compassion you feel on seeing her crumpled on the sidewalk. Basic moral beliefs like these need not be the upshot of (implicit) reasoning. Rather – just like May’s door example – they can result from simply taking your emotional experience at face value. Moreover,

although this point has been made by sentimentalists who take emotions to be perceptions (e.g., Tappolet 2016), the above account of emotion indicates that it holds for sentimentalism more generally.

In short, we have a range of examples showing not only that emotions carry morally relevant information, but also that they can play a significant role in moral judgment and inference.

Emotions are not mere consequences

At this point, May might object that the sentimentalism sketched here fits poorly with empirical findings suggesting that emotions are merely a consequence of (non-emotion-based) moral inferences and beliefs, not the drivers of them (pp. 38–41). In particular, May could extend the conclusions that he draws from experiments investigating the temporal order of subjects’ judgments about the disgustingness and moral wrongness of certain actions (Yang et al. 2013). This work suggests that disgust judgments *follow* moral judgments – a conclusion that fits poorly with standard sentimentalist proposals.

However, the relevance of these experiments is questionable. First, it is unclear how much we can draw from experiments focused on just one emotion (disgust). Moreover, research on other emotions (fear and anxiety) suggests that the temporal ordering of emotion and higher cognition is more in line with the sentimentalist account sketched here (e.g., Hofmann et al. 2012, Kurth 2018, pp. 52–53). Most significantly, the task used in Yang et al.’s Go/No-Go experiments was complex: subjects were asked to make a decision about what button to push based on comparisons of their assessments of the disgustingness and moral wrongness of an action. But given that this was the task, the experiment does not appear to provide insight of the sort May needs (namely, evidence about the temporal order of *feelings* of disgust in comparison to moral *judgments*). Rather, it appears to focus on something else: how we make comparative assessments about (i) our *judgments* regarding the disgustingness of an action and (ii) our *judgments* of the moral wrongness of that action.

Stepping back, we can see how a richer understanding of what emotions are provides sentimentalists with new resources that help them vindicate a central role for emotion in moral cognition.

Cautiously optimistic rationalism may not be cautious enough

Justin F. Landy 

Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604.
jlandy@fandm.edu justinflandy.com

doi:10.1017/S0140525X18002613, e159

Abstract

May expresses optimism about the source, content, and consequences of moral judgments. However, even if we are optimistic about their source (i.e., reasoning), some pessimism is warranted about their content, and therefore their consequences. Good reasoners can attain moral knowledge, but evidence suggests that most people are not good reasoners, which implies that most people do not attain moral knowledge.

Regard for Reason in the Moral Mind (May 2018) is an impressive work. Drawing on the latest psychological research, May pushes back against prominent sentimental theories in normative ethics and moral psychology that view moral judgments as the products of unreasoned, emotional processes. Ultimately, he defends a “cautiously optimistic” form of rationalism (p. 227): Our moral judgments are the product of reasoning, so “virtue is within reach” (p. xi), because we are capable of acquiring moral knowledge and knowing right from wrong (p. 4).

May is therefore optimistic about the *source* of our moral judgments (i.e., they are the products of reasoning, which is capable of tracking moral facts), which leads to optimism about the *content* of our moral judgments (i.e., we can know right from wrong), which leads to optimism about the *consequences* of our moral judgments (i.e., we can act in accordance with them). I agree with many of the positions that May argues for: In my view, moral judgments are products of reasoning (Royzman et al. 2015b), moral cognition does not fundamentally differ from other kinds of cognition (Landy & Bartels 2018), and emotions are consequences, not causes, of moral judgments (Landy & Goodwin 2015; Royzman et al. 2014a).

However, I think that cautiously optimistic rationalism may not be cautious enough. Even if we accept May’s (2018) optimism about the *source* of our moral judgments, we ultimately care about this because it speaks to our ability to actually attain moral knowledge and act accordingly – that is, because it speaks to whether we should be optimistic about the content and consequences of our moral judgments. In other words, although virtue may be “within reach,” this is important because it is relevant to the question of whether we can be reasonably expected to successfully reach out and actually take hold of virtue. Two observations lead to the conclusion that this may not happen as often as we would hope, and that a tempered pessimism about the content (and, therefore, the consequences) of our moral judgments is warranted.

First, it seems plausible that those of us who are better at reasoning are more likely to successfully reach out and grasp virtue, and, conversely, that those of us who do not reason well are less likely to do so. Indeed, May (2018) seems to accept at least a weak form of this position: “sophisticated” reasoners are “likely to have more well founded moral beliefs than those ignorant of the key details or more prone to cognitive errors” (p. 236). Many sentimentalists will dispute the claim that reasoning has any relationship at all with the content of our moral judgments (see, e.g., Haidt 2001; Schnall et al. 2008). Nonetheless, research has shown that there is substantial variation in people’s domain-general ability and propensity to reason thoroughly, and that variation in this kind of domain-general reasoning performance does predict the content of people’s moral judgments (see, e.g., Landy 2016; Royzman et al. 2014b; Royzman et al. 2015b; for a recent review and synthesis, see Landy & Royzman 2018). So, I will accept the premise that better reasoners are more likely to arrive at well-founded moral beliefs than are worse reasoners.

The problem for cautiously optimistic rationalism is that most people seem to be unable or unwilling to think through reasoning problems when they are faced with them. For example, the modal number of correct answers on the much-studied Cognitive Reflection Test (CRT; Frederick 2005) – a three-item performance measure of reasoning – is usually found to be zero (e.g., Campitelli & Gerrans 2014; Frederick 2005; Pennycook et al. 2016; Royzman et al. 2014b), despite the fact that the three problems require only rudimentary cognitive work to solve correctly. Performance on the CRT is thought to depend on both reasoning

ability (similar to an IQ test) and the *propensity* or *motivation* to reason through problems (see Pennycook & Ross 2016). So, most people seem not to be very good reasoners, because they lack the necessary ability, motivation, or both.

Rationalism implies that “if all goes well, you form the correct [moral] judgment, it’s warranted or justified, and you thus know what to do” (May 2018, p. 19), but the problem is that we have little reason to assume that “all goes well,” most of the time. The research suggests, instead, that things often go rather poorly. The premise that bad reasoners are unlikely to form well-founded moral beliefs, combined with the empirical evidence that most people lack either the ability or the motivation to engage in good reasoning, leads to the conclusion that, for most people, much of the time, either virtue is “out of reach,” or they lack sufficient motivation to extend their arms and grab it. Either way, people may not find themselves with virtue in hand very often.

May (2018) engages with a version of this problem (Ch. 5, sect. 3.3), in which he discusses widespread cognitive biases that interfere with domain-general reasoning. He argues that this is not problematic, though, because “they don’t afflict *moral* judgment in particular but reasoning generally” (p. 125). May and I seem to agree that moral judgments are products of the same kind of domain-general reasoning mechanism that produces other kinds of judgments, given his argument that “moral judgment is just like other forms of cognition except that it involves moral matters” (p. 228). If this is the case, then widespread defects or biases in reasoning represent a potentially serious threat to the attainment of well-founded moral knowledge in most cases. Most people, as he notes, have “little claim to being a moral guru” (p. 126), but he does not acknowledge this as a serious problem for his optimism regarding the content of our moral judgments. Here, his comparison of moral reasoning with mathematical reasoning strikes me as apt. When it comes to both math and morals, what we presumably care about is arriving at the right answer via the right kind of process. Although the “basic capacity is not fundamentally flawed” (p. 129), anyone who has taught a statistics class can attest that many people never successfully reach out and grasp mathematical competence, and those that do often do so only with considerable effort. Rather than “sweeping pessimism about *only* the moral domain” (p. 230, emphasis added), a more tempered pessimism seems warranted about our reasoning in general. Of course, this entails pessimism about both our mathematical cognition and, more germane to the present discussion, our moral cognition. Even if we are optimistic that moral judgments result from the kinds of domain-general reasoning processes that also drive mathematical cognition, if those processes frequently go awry, we have little reason for optimism about the content of the moral judgments they produce.

May also notes that it is beyond the scope of his book to address “deep skepticism about the reliability of our general cognitive, learning, and reasoning capacities” (p. 106). Fair enough. I offer this commentary in the spirit of advancing the discussion beyond the already considerable amount that he has accomplished in the book. Importantly, though, I am not arguing that “all cognition, moral and non-moral, is bunk” (p. 230). My point is that the empirical literature suggests that good reasoning is not *impossible*, but it is relatively *rare*. This is a separate “empirical threat to the acquisition or maintenance of well-founded moral beliefs” (p. 20) from the two that are addressed in chapter 5, and it is one that May’s cautiously optimistic rationalism does not currently speak

to. The argument that I am making for a moderate amount of pessimism cannot be dismissed as merely radical skepticism.

A defender of cautiously optimistic rationalism might reply that part of May's argument is that inferential, cognitive processes do not need to be conscious and explicit to qualify as reasoning (see, e.g., pp. 8–9, 54–55). They might then argue that the CRT and similar psychological instruments primarily tap conscious, "System 2 reasoning," so it is possible that most people are rather good at more intuitive "System 1 reasoning," and therefore that we do reach out and successfully grasp virtue reasonably often. The first premise in this argument can be contested – reasoning is often associated with "System 2," but not "System 1," processes (e.g., Kahneman 2011; Kokis et al. 2002) – though I do personally find May's argument that at least some instances of effortless, automatic cognition can qualify as a kind of "reasoning" to be compelling (see also Landy & Royzman 2018, fn. 2).

However, whether or not we accept this first premise, the second premise and, therefore, the conclusion, are problematic. The CRT is usually thought to measure success at overriding a response that is prepotent, intuitive, and incorrect (Frederick 2005, though see Pennycook et al. 2016). That is, even if we accept the first premise in this reply, low scores on the CRT can be thought of as reflecting failures of "System 1 reasoning" to produce the correct response, as well as failures of "System 2 reasoning" to recognize this error and override it. This assertion is bolstered by the fact that CRT performance is negatively correlated with susceptibility to intuitive heuristics and biases (Toplak et al. 2011). This is not definitive evidence that people are, by and large, bad intuitive reasoners, but it does at least undermine the argument that we can safely assume that "System 1 reasoning" is generally reliable, and therefore that we should be optimistic about our chances of successfully taking hold of virtue.

Of course, to even be able to say whether we have virtue in hand on any given occasion requires an independent normative criterion by which moral judgments are right and which actions are virtuous and which are wrong. There are some defensible metaethical theories that would posit that no such criterion can reasonably be said to exist (e.g., moral error theory, see Mackie 1977), but even if one believes that moral properties are mind-independent and truth-apt, it is still the case that no theory of normative ethics has attained consensus after some 2,500 years of work in this area. How, then, are we to know when we have reached out and taken hold of virtue, and when we have not? We do not have a noncontroversial answer to this question, as of yet.

In sum, I agree with May that moral knowledge is "possible" (p. 5), but I doubt that it is all that *probable*, in most cases. Given what we know about the prevalence – or rather, the lack thereof – of good reasoning, a moderately pessimistic form of rationalism seems more appropriate than a cautiously optimistic one. Yes, our moral judgments are largely products of reasoning, but reasoning is not something that most of us are especially good at.

Moral principles in May's *Regard for Reason in the Moral Mind*

Colin Marshall 

Department of Philosophy, University of Washington, Seattle, WA, 98195.
crmarsh@uw.edu
<https://sites.google.com/site/colinmarshallphilosophy/>

doi:10.1017/S0140525X18002674, e160

Abstract

Joshua May offers four principles that might serve as the rational foundations of moral judgments. I argue that these principles, if they are independent of affect, are too weak to be the basis of any substantive moral judgment and do not fit with the idea that morality is categorical.

Kant famously held that reason plays a fundamental role in our grasp of morality. Because he believed that reason was outside the empirical realm (Kant 1996, p. 99), Kant denied that our moral judgments could be ultimately understood through empirical investigation (though he gave "moral anthropology" an important secondary role [Kant 1996, p. 372]). Kant even claimed that moral facts could come radically apart from any empirically detectable facts (Kant 1998, p. 544). However, the philosophical climate has shifted from Kant's time. Most philosophers today believe that any complete account of human moral judgment must be closely tied to empirical psychology. Some, such as Joshua Greene, think this bodes poorly for deontological rationalist views like Kant's (Greene 2008).

Joshua May's (2018) *Regard for Reason in the Moral Mind* argues that the best empirical evidence does not threaten either the rationalist claim that moral judgments are based in reason or our acceptance of broadly deontological moral principles. The book is clearly written, philosophically rich, and enjoyable to read. Its optimistic claims and tone was, for this reader at least, very welcome. More often than not, I found May's arguments persuasive. In case after case, May provides a "hold on – let's look at the details" check on empirical results from which many have drawn pessimistic, anti-rationalist conclusions. Whether or not May intends it to be, his general defensive approach is itself in the Kantian tradition, because Kant principally emphasized that the empirical facts *left room* for strong moral facts.

Although May's primary aim is to defend a broadly rationalist view, he also offers some pieces of the particular moral view he himself is drawn to (on both philosophical and empirical grounds). That view is more modest than many earlier rationalist moral views, but it is far from trivial. May claims that "a creature with unlimited time and resources needn't possess emotion [or, more specifically, affect] to make distinctively moral judgments" (p. 13 – though May later allows that emotional affect might be a requirement for all cognition [p. 80]). In chapter 3, May offers four principles that may underlie moral inference, only one of which is directly consequentialist:

Consequentialist Principle: All else being equal, an action is morally worse if it leads to more harm than other available alternatives. (p. 57)

Intentionality Principle: All else being equal, it's morally worse to cause harm intentionally as opposed to accidentally. (p. 61)

Action Principle: All else being equal, harm caused by action is morally worse than harm consequent upon omission. (p. 62)

Principle of Agential Involvement: All else being equal, it is morally worse for an agent to be more involved in bringing about a harmful outcome. (p. 69)

May does not commit to these principles being foundational, though he notes that he is inclined to think there are at least

some foundational moral principles which guide our thought in the roughly the same way that rules of grammar do (see pp. 70, 78–79). Though May generally characterizes reason as the capacity for inference (basing beliefs on other beliefs [p. 11]), he does not claim that these principles are the result of inferences. They would count as rational, presumably, just by being non-affective and by *supporting* inferences in some way.

My aim here is to assess the form of rationalism that would hold if May's four principles were the rational foundation of moral belief. The contrast I will consider is with a sentimentalist view that agrees with May about the psychology reality of the principles, but takes them to somehow rest on affect. May acknowledges that a sentimentalist could agree about these principles (pp. 55–56, 71), but argues at length in chapter 2 that we lack any strong empirical grounds for accepting sentimentalism. In this discussion, I am going to assume that May is right about the empirical factors he considers, and instead offer abductive considerations that favor sentimentalism about the principles over rationalism. If I am right about those considerations, then May's most obvious line of response will be to revise or expand the principles. My remarks are thus primarily intended to provide May an opportunity to further develop his positive view.

To begin, consider again the most famous rationalist view of moral judgment: Kant's. Kant held that there is a single fundamental moral principle, whose main formulation is: "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (Kant 1996, p. 73). Though there are hard questions about how exactly Kant's principle gets applied (see, e.g., Herman 1993), it is fairly clear that Kant believed that the principle of reason *deductively entails some substantive verdicts* about what to do. This is what we would expect given Kant's general view of reason, which, in his mature work, he introduced in terms of syllogistic inference (Kant 1998, pp. 387–91). Kant was not alone in this – other rationalists like Spinoza also believed that foundational moral principles have some substantive, deductive implications (see, e.g., Spinoza 1988, pp. 586–87). No doubt this was tied to their view of reason as the capacity that allows us to make mathematical inferences and grasp necessary truths.

By contrast, May's four principles have no substantive deductive implications by themselves, because they are *ceteris paribus* principles. They apply only when other things are equal, but give no indication of which other things are relevant. For all the principles say, for example, consequentialist considerations cease to be relevant when trolleys are involved. Hence, if these principles were the true foundations of moral judgment, then any substantive moral judgment based on them (as in the trolley case) would involve an irrational inferential jump. This matters both for May's account of moral judgment and for his defense, in chapter 8, of the idea that moral beliefs can generate the primary motivation for a particular action by generating a desire. It is hard to see how this later idea could work if the beliefs in question had no substantive implications for particular actions.

Contrast this with rules of grammar for English, such as the rule that every declarative sentence requires a verb. At face value, this principle has substantive deductive implications – it dictates that "I cat" is not a declarative sentence. That does not mean that speakers of English have an articulated belief in the rule, of course. But it has some psychological reality for them, and explains how they are able (without any irrational inferential jumps) to recognize some novel sentences as problematic. By contrast, consider the difficulty in learning spelling in English or

learning logograms in Chinese. Although certain rules of thumb apply, the rules are not generative, so brute memorization is needed for accurate recognition.

Presumably, though, May would take fundamental moral principles to be more than rules of thumb that require supplementation by memorization. Structurally, then, they seem to be closer to rules about seasoning foods such as "other things being equal, add some salt." This principle has no direct substantive entailments. It is not, however, merely a rule of thumb that is supplemented by memorization. Instead, it gestures at a *pattern of responses* we have to food: most humans respond positively to salt, but not always. These responses are not determined by reason in any sense, however, but are instead arational and affective. Hence, I suggest that if the most foundational principles for a domain are merely *ceteris paribus* principles, we have abductive reason to think that the principles are really just gesturing at some arational pattern of responses.

May's book offers what might be a potential response by analogy: our way of classifying objects as furniture or non-furniture (p. 70). There might be principles of furniture-identification, such as "if the function of an object is for sitting, then it is furniture." Such principles, though, "merely identify prototypical features that are statistically frequent in the category or exemplars with which I can compare the object in question" (p. 70), and are stronger than mere *ceteris paribus* principles. Recognizing furniture, however, is a broadly rational accomplishment, in May's broad sense of "rational," because it involves forming a belief on the basis of other beliefs – albeit in a non-deductive way.

Understanding May's principles in this way, however, suggests that they are not really foundational principles, but instead stem from some more foundational representations of moral prototypes. By contrast, principles of grammar and seasoning do not seem to hinge on prototypes, though we might use prototypes to identify particularly good examples of grammar and seasoning. This may be connected to why it seems harder to explain cross-historical and cross-cultural convergence by appeal to prototypes than to innate rules (this is part of why traditional rationalists modeled their moral principles on logical and mathematical principles). Because May is drawn toward the grammar analogy partly to explain intercultural convergence (p. 78), he therefore has reason to not rely too heavily on the furniture analogy.

I will make one more broadly Kantian point. Arguably, part of the reason that Kant held that the moral law was unconditionally binding ("categorical") was that its verdicts were clear (at least, when the right questions were posed). However, it seems that one way to not be bound by a principle is to be unable to see what it implies. For example, there are limits to how much we can hold someone accountable for being un-American (even given full knowledge of other relevant facts), because it is often not clear what being "American" requires. Likewise, there are many cases where we would not fault someone for being unsure whether something was furniture, even if she had a full knowledge of the physical properties of the item in question. If we do indeed regard morality (unlike convention) as unconditionally binding in situations where we know the relevant non-moral facts, then that suggests that we take moral requirements to be reliably clear. Assuming we are coherent in seeing morality as binding and clear, it would therefore seem that we must take morality to be guided by something with clearer substantive implications than *ceteris paribus* principles or exemplars. If absolutist principles like Kant's are off the table, then strong affects would seem like the best candidate, because even in novel cases, we are often very confident (albeit sometimes

wrongly) about how other people will respond to a given level of saltiness. Of course, May could say that we are simply wrong to assume that our rational principles have clear implications, and so wrong to see them as unconditionally binding, but attributing such an error to us seems like a cost of a view.

In sum, then, if May's four principles are the foundations of moral judgment, there is reason to think that, like principles of seasoning, they rest on sentiments in some way. The obvious response is for May to deny that, as stated, these principles exhaust the foundations of moral judgment, perhaps leaving the matter up to further empirical investigation. I expect that May is not willing to go as far as Kant and identify a single clear principle with substantive, deductive implications, because he later appeals to the fact that there is "nearly always ... wiggle room in the application of moral principles" (p. 167). Even so, there is plenty of middle ground between weak *ceteris paribus* principles and a Kantian view. I therefore hope that May will develop his view further, perhaps in a way that can explain why we are inclined to think that the right moral answer is often obvious, even in novel situations.

Moral reasoning performance determines epistemic peerdom

William H. B. McAuliffe^a  and Michael E. McCullough^a 

^aDepartment of Psychology, University of Miami, Coral Gables, FL 33146.
w.mcauliffe@umiami.edu <http://williamhbmcauliffe.com/> mikem@miami.edu
<http://local.psy.miami.edu/faculty/mccullough/>

doi:10.1017/S0140525X18002595, e161

Abstract

We offer a friendly criticism of May's fantastic book on moral reasoning: It is overly charitable to the argument that moral disagreement undermines moral knowledge. To highlight the role that reasoning quality plays in moral judgments, we review literature that he did not mention showing that individual differences in intelligence and cognitive reflection explain much of moral disagreement. The burden is on skeptics of moral knowledge to show that moral disagreement arises from non-rational origins.

In chapter 5 of Joshua May's (2018) *Regard for Reason in the Moral Mind*, he concedes that moral disagreement among "epistemic peers" – people who have equally good access to the truth of a matter – can undermine the claim to moral knowledge. However, he also rightly points out that moral disagreement often arises from poor thinking, such as motivated reasoning, and disagreement about the non-moral premises undergirding moral conclusions, such as whether same-sex marriage undermines social stability (p. 120). May concludes that it is difficult to identify whether a disagreeing peer is also an epistemic peer in practice, especially if he or she is from a different cultural milieu. We agree with this conclusion, but contend that researchers have been successful in identifying some factors that, in aggregate, show that much of moral disagreement does *not* occur among epistemic peers. Here, we review this literature because May did not, and because it points up the importance of rational factors – namely, differences in

cognitive ability, education, and tendency toward cognitive reflection – in explaining the quality of moral judgments.

Consider the simple case of competently judging whether a moral violation has taken place. The judge must assess (a) whether there was an actual or potential patient of harm, (b) whether there was an agent who intended that harm, and (c) whether the harm was a means to selfish ends (Sousa & Piazza 2014). Achieving these tasks requires the judge to experience empathy for the putative victim, to deploy theory of mind regarding the agent's intent, to apply accurate background beliefs about the act's typical consequences, and to impartially consider whether the act would be acceptable regardless of the identities of the agent and the patient (Gibbs 2013). The judge must then check whether her initial impression coheres with her other moral beliefs and whether there are relevant mitigating circumstances (Holyoak & Powell 2016). All the while, the judge must ensure that self-interest or a desire to pander to a certain audience does not corrupt any of these processes (Krebs & Denton 2005). Each of these tasks considerably increases in difficulty if the violation in question is not common in the judge's everyday life (Davidson et al. 1983) or concerns several stakeholders from diverse walks of life (Gibbs 2013). The judge must perform optimally in all of these tasks to be an epistemic peer of another person who performed optimally.

All components of moral judgment require the application of sophisticated cognitive and socioemotional capacities that differ in strength across people and do not fully develop until at least adolescence. It is no surprise, then, that differences in intellectual achievement are strong predictors of differences of moral opinion. For example, intelligence at age 10 predicts anti-traditional beliefs (e.g., endorsement of gender equality in the workplace, opposition to retributive justice, and rejection of racism) at age 30, even after controlling for educational achievement (Deary et al. 2008). Also, meta-analyses indicate that illiberal attitudes are positively associated with about a dozen different measures of cognitive rigidity (Jost 2017). And to round the bases on May's aforementioned example, intelligence is positively associated with support for same-sex marriage (Perales 2018).

Additionally, young people who are still at low levels of cognitive development tend to make category mistakes in moral reasoning. For example, still-developing minds tends to confuse morality with power dynamics, self-interest, peer approval, and the status quo (Gibbs 2013; Piaget 1932). Moreover, intelligence is strongly associated with successful distinctions between moral violations (i.e., actions that intrinsically have detrimental consequences for others) and convention violations (i.e., actions that disrupt social order within a given culture, but would not be harmful in other contexts; Aharoni et al. 2012; Royzman et al. 2014b). A failure to make this distinction is partly responsible for why less reflective people tend to treat violations of the "binding foundations" of morality – authority, tradition, and purity – as intrinsically wrong (Landy 2016). Contra May, then, one need not grant that people who moralize different sets of values are epistemic peers (p. 123). Someone who does not know which distinctions really count when making moral judgments is not an epistemic peer of someone who does know which distinctions really count.

Of course, intelligence is not everything: One must also be motivated to apply it when making a judgment. Seminarians, for example, typically recognize morally mature arguments, but some of them choose to relinquish reason in favor of obedience to God (Lawrence 1987). Failure to think through the details of a dilemma can lead people to dogmatically champion one moral consideration to the neglect of legitimate alternatives. For

example, reflective thinkers regard either the deontological or utilitarian resolutions to moral dilemmas as morally permissible (Royzman et al. 2015b). Less reflective people prefer a particular resolution, suggesting that they do not acknowledge that there really is a dilemma. Although taking a hard line on an issue sometimes reflects principled belief, in dilemmatic contexts strong opinions more often reflect an unthinking adherence to a rule, an inflexibility that most children eventually learn is inadequate for dealing with the complexities of life (Lourenço 2003).

Perhaps the most pernicious misapplication of intelligence to the moral domain is the motivated rationalization of views one wants to maintain (Stanovich et al. 2013). For example, intelligence is positively associated with prejudice against conservative targets such as corporations, Christians, and the military (Brandt & Crawford 2016). People rely on stereotypes less when individuating information is available (Jussim 2017), but they must be willing to seek out such information in the first place. Hence, the Big Five trait that is most negatively related to generalized prejudice is agreeableness, which reflects lenience in judging and a desire to get along with others, not openness to experience, the trait most linked to intelligence (Crawford & Brandt 2019). This example reinforces the point that because making good moral judgments depends on so many distinct capacities, suboptimal performance on any one task can compromise one's claim to epistemic peerdom.

Moral reasoning is also not a mere academic ability, as its importance is evident when examining moral heroes and moral transgressors. In their landmark comparison of rescuers of Jews during the Holocaust to bystanders, Oliner and Oliner (1988) found that rescuers were more likely to have been raised to adopt a universal care ethic and less likely to accept an ethic of obedience. Walker et al. (2010) found that moral reasoning ability was the distinguishing feature of a subset of people who had won lifetime achievement awards for their prosocial contributions to society. Among ordinary persons, scores on moral reasoning tests are positively associated with volunteering and going into a helping profession (Comunian & Gielen 1995; Rest et al. 1999). Similarly, cognitive ability is positively associated with charitable giving (Bekkers & Wiepking 2011). Conversely, moral reasoning scores relate negatively to selfish, manipulative tendencies (Marshall et al. 2017). Criminal offenders have lower scores on moral reasoning tests than do non-offenders (Stams et al. 2006), a group difference that is likely mediated by deficits in cognitive empathy and general intelligence (O'Kane et al. 1996; Van Langen et al. 2014). Among criminal offenders, lower moral reasoning scores predict increased recidivism (Van Vugt et al. 2011) and psychopathic traits predict deficits in detecting social contract violations (Ermer & Kiehl 2010). Both lines of evidence suggest that recalcitrant offenders have difficulty obtaining and applying moral knowledge.

The role of reason in promoting prosocial behavior and inhibiting antisocial behavior is even evident in the historical record: As societies became more cosmopolitan over time, the justifications governments gave for helping the needy became more distinctively moral (McCullough, forthcoming). The earliest justifications for regard for the poor were mostly self-serving inasmuch as they secured reputational benefits for rulers, enabling them to consolidate their power in the face of competing interests, establish peaceable kingdoms, and lubricate trade relations with other societies and ethnic groups. Later justifications were based on prudential arguments about the collateral effects of poverty on the prevalence of disease, crime, vice,

and social unrest. It was only during the enlightenment era that arguments about helping the poor and preventing poverty became distinctly moral in character, invoking distributive justice, the equal dignity of all persons, and the maximization of utility at the societal level. Thought experiments involving veils of ignorance, original positions, and children drowning in shallow ponds would not come until the latter half of the twentieth century. The spread of literacy, along with reductions in the prices of books and the speed with which information could travel, also encouraged distinctively moral reasoning by providing people with humanizing portraits of poor and distant victims. Similar advances in moral reasoning, education, and literacy also help to explain the decline of violence between states over the past 500 years (Pinker 2011a). The qualitative changes in moral justifications for helping others and against harming others across generations is remarkably similar to qualitative changes in moral reasoning within the lifetime of single persons (Gibbs 2013).

A by-product of moral progress is that the standards for becoming an epistemic peer in the moral domain have increased now that access to information is more available than ever (McCullough, forthcoming; Pinker 2018). Newspapers, radio, television, and the Internet make it easier to learn about other people's plights, which ideally enable people to come to better agreement about when societies have moral obligations to combat injustice and improve the lot of those in dire need. And now with considerable historical precedent for offering impartial reasons for one's point of view, it is harder to get away with a patently self-interested moral compass (Shermer 2015).

May concludes that people need only be skeptical toward people's ability to obtain moral knowledge about particularly controversial issues, where reasonable people disagree because the relevant empirical premises are uncertain and the temptation toward motivated reasoning is strong (p. 128). We agree that intellectual humility is an antidote to counterproductive polarization, but we counsel against relying on the proportion of people who believe a certain point of view to determine which moral conclusions are beyond our ken. For history reveals not only that the moral compass of the masses has improved over time, but also that there have always been individuals who were centuries ahead of their time in their moral outlook. For example, long before sizable abolitionist movements caught hold in the United States, there were those who cogently argued that there are no differences between blacks and whites that entitle whites to subordinate blacks (Lepore 2018). Otherwise reasonable people – including some founders of the U.S. constitution who were ahead of their time in other ways – disagreed, but their counterarguments were self-interested and based on false claims such as that blacks wanted to be ruled or that they did not possess rational capacities. Others, such as Benjamin Franklin, were resistant at first, but changed their minds after reflecting on abolitionist arguments and taking the time to observe black communities in a disinterested manner. What this example shows is that simply counting the number of learned people who hold a certain moral point of view is not an infallible means of detecting whether that view is reasonable, positive correlations between cognitive ability and moral positions notwithstanding. In all cases, one must examine the reasoning and evidence that each side has brought to bear to determine who is an epistemic peer of whom. Only those who are not committed to taking the time to think about and research an issue need withhold judgment.

Do framing effects debunk moral beliefs?

Kelsey McDonald , Siyuan Yin, Tara Weese
and Walter Sinnott-Armstrong

Philosophy Department, Duke University, Durham, NC 27708.
kelsey.mcdonald@duke.edu siyuan.yin@duke.edu
tara.weese@duke.edu ws66@duke.edu
<https://sites.duke.edu/wsa/>

doi:10.1017/S0140525X18002662, e162

Abstract

May argues that framing effects do not undermine moral beliefs, because they affect only a minority of moral judgments in small ways. We criticize his estimates of the extent of framing effects on moral judgments, and then we argue that framing effects would cause trouble for moral judgments even if his estimates were correct.

In his précis (sect. 3.1) and book (May 2018, p. 85), Joshua May schematizes moral debunking arguments like this:

1. Some of one's moral beliefs are mainly based on a certain factor.
2. That factor is morally irrelevant.
3. So: The beliefs are unjustified.

This logically invalid argument depends on a suppressed premise that moral beliefs are unjustified whenever they are mainly based on a morally irrelevant factor.

May accepts the normative premise (2) as “eminently plausible” (2018, p. 90) in cases of genuine “framing effects” that are “clearly morally irrelevant” (p. 89). May's examples of irrelevant framing include order of presentation, equivalent wording (Petrinovich & O'Neill 1996), and grammatical person (Nadelhoffer & Feltz 2008).

After granting the normative premise in these examples along with the suppressed premise, the only remaining way for May to avoid the conclusion (3) is to deny the empirical premise (1). Against premise (1), May replies “meta-analyses suggest that the vast majority of moral decisions remain the same in the face of genuine framing effects” (Demaree-Cotton 2016)” (sect. 5.2; see also May 2018, pp. 91, 218). Thus, May admits that a minority of moral beliefs are mainly based on a morally irrelevant factor, so premise (1) is true of those beliefs. His primary reply concerns the number and kind of moral beliefs that are changed by genuine framing effects.

We will argue that this reply is inadequate for three reasons. First, Demaree-Cotton's meta-analysis does not show as much as she and May claim. Second, even if it did show that framing does not affect “the vast majority” moral judgments, the number of judgments affected would still be enough to cause trouble for popular views about how moral beliefs are justified. Third, this trouble applies even to obvious moral beliefs that are immune from framing effects. We discuss these three points in turn.

First, May relies crucially on one meta-analysis by Demaree-Cotton (2016). (May also mentions Kühberger [1998], but that analyzes studies of framing effects on risky choices rather than morality.) Demaree-Cotton's meta-analysis is taken by May to

show, “Roughly 80 percent of people's moral intuitions subject to framing effects don't change, and that figure excludes studies that found no effect” (May 2018, p. 91; cf. p. 218). Unfortunately, Demaree-Cotton's meta-analysis suffers from several flaws.

One technical problem is that Demaree-Cotton arrives at her conclusion that “80 percent [...] *don't* change” simply by taking the difference between the proportion of moral judgments in distinct frames. She takes the difference between frame groups to show the proportion of people whose moral intuitions are changed by the frame. This statistical interpretation, however, obscures differences among subjects and among types of moral intuitions in susceptibility to framing effects. Moreover, simply taking the difference between two point estimates of group averages (i.e., means) ignores the uncertainty of the measurements; indeed, there are cases in which a large observed difference between two averages is still statistically not significant because of large measurement uncertainty. Furthermore, simply observing an effect in a sample does not provide a valid statistical basis for inferring an effect of this magnitude in the general population. For example, Demaree-Cotton assumes that a 70% difference between framing groups is statistically equivalent to a 70% chance that a randomly selected person's moral judgment is determined by the frame. This inference misconstrues how true effect sizes are estimated and generalized to new populations. To determine whether the effect of an independent variable generalizes to new subjects, one would need to run a random-effects, or mixed-effects, regression analysis that treats “subject” as a random factor.

Additional problems with Demaree-Cotton's meta-analysis concern not her statistical analysis but the studies she surveyed. Most framing effect studies included in Demaree-Cotton's meta-analysis are between-subjects rather than within-subjects experimental designs. This distinction is important, because any difference observed in a between-subjects design also carries differences between subjects, because each person sees only one experimental condition. Within-subjects designs, in contrast, expose each person to several treatment conditions, so individual differences among groups do not conflate the observed difference between conditions in within-subjects designs.

In addition, the studies that Demaree-Cotton reviewed cover only a very limited range of moral dilemmas. Six out of seven of them concern killing one person to save others, and they all specify that consequences will definitely occur without any mention of how likely these consequences are. There is little basis for generalizing from this small subset to other kinds of moral dilemmas, especially scenarios that involve risk and uncertainty. As May (2018) says, “we're just bad at reasoning with probabilities and risk generally” (p. 91), so it would not be surprising if framing effects were greater in risky moral dilemmas.

These problems with Demaree-Cotton's sample in addition to flaws in her statistical analysis undermine her and May's conclusion that roughly 80% of moral intuitions are reliable and not susceptible to framing effects. To estimate the extent of framing effects more accurately, we are preparing a larger meta-analysis of more than 80 studies of framing effects on a wider variety of moral judgments. Before completing our meta-analysis, we cannot be sure whether the 20% average rate of framing effects is too high or too low. In any case, it is premature for May (2018) to conclude that framing effects on moral judgments are “negligible” (p. 92) or that the “vast majority” (p. 90) of moral judgments are immune from framing effects.

Our second point concerns whether a 20% rate of each framing effect is enough to cause trouble. It seems so, especially

because of the multiplicity of framing effects, understood broadly as effects of morally irrelevant factors on moral judgments. May discusses effects of order, wording, and grammatical person. Other studies find effects of morally irrelevant factors, such as videos (Valdesolo & DeSteno 2006), disgust inducers (Schnall et al. 2008), cleaners (Helzer & Pizarro 2011), sleep deprivation (Killgore et al. 2007; Olsen et al. 2010), and social setting (Kelly et al. 2007). These factors lie in the environment of the person judging rather than in the situation of the judged act, so they make people in different environments ascribe contrary moral properties to the same act in the same situation. It would be inconsistent to say that both of these contrary judgments are correct, so the environment of the person making the judgment cannot affect the moral status of the judged act. That ensures that such factors are morally irrelevant.

In response, May repeatedly points out that each of these effects is small and occurs in only some people. Good point! Nonetheless, multiple small effects each on a minority sometimes add up to large effects on the majority. Imagine that we find five framing effects that each affects 20 people out of 100. If the same 20 people are vulnerable to all five framing effects, then the remaining 80 are not vulnerable to any of those effects. However, if each framing effect occurs in a separate 20 with no overlap, then every one of the 100 is vulnerable to one framing effect. Moreover, if one framing effect makes a person's moral judgment a little less confident or a little less extreme in content, then another framing effect moves it further in the same direction, and another framing effect moves it even further in that direction, then all three framing effects together can amount to a large effect. Thus, to determine how many people's moral judgments are misled by framing and how much, we need to look not at each framing effect in isolation (as May does) but at many framing effects together.

Furthermore, let's grant May's assumptions that frames affect about 20% of moral judgments and that the same people are subject to all of the various framing effects, so 80% are immune. Even on these generous assumptions, 20% of us still make moral judgments distorted by framing effects. May (2018) denies that 20% is "substantial" (p. 91). We disagree. A gambler does not know that a six-sided die will *not* come up six, although the chance of a six is less than 20%. Analogously, we do not know that a particular moral judgment is not distorted by framing effects when there is a 20% chance of distortion. Admittedly, it can be rational to bet a little on one die not coming up six. However, such bets become irrational when there is little to gain and mistakes are costly. In important and controversial moral debates, mistakes are costly, and not much is lost by suspending belief about moral claims that lack confirmation.

The problem arises because we as individuals often do not know whether we are in the unreliable 20% or the reliable 80%. If we do somehow know that our own moral judgments are not distorted by framing effects, even though other people's moral judgments are so distorted, then we might be justified in trusting our moral judgments. But how would we know that we are so lucky? We would need an independent test of whether our moral judgments are correct. And even if we had an independent test, that test rather than the mere fact that a certain moral judgment seems right to us would be what makes us justified. Framing effects thus show that we are not justified in relying on moral intuitions by themselves without any independent confirmation that our moral intuitions are reliable.

To all of this, May (2018) and others are likely to object that some moral judgments seem obvious. May's example is "condemning

someone for intentionally and successfully poisoning an innocent co-worker" (p. 92). We agree that this example and others are obvious. We also agree that such obvious cases are unlikely to be changed by framing effects (cf. Tanner & Medin 2004). However, such cases are not enough to show that moral judgments in controversial cases can be justified. Those controversial moral beliefs are both interesting and important, so framing effects can debunk many significant moral judgments, even if not all.

Moreover, framing effects on the other moral judgments can reveal *how* moral judgments are justified even in the obvious cases. When so many other moral judgments are based on framing effects, one cannot be justified in believing any particular moral judgment without some reason to believe that this particular judgment is not subject to framing effects. Just as one cannot know that one is a reliable moral judge, whereas others are not, unless one has some reason to believe that one is special in some relevant way; so one cannot know that any particular moral judgment is undistorted by framing effects, whereas others are distorted, unless one has some reason to believe that this judgment is special in some relevant way. We still might be justified in believing that judgment, but only if we have independent confirmation that it is somehow immune from framing effects. Then that confirming evidence is what makes the moral judgment justified, instead of the mere fact that it seems obvious to us.

In this way, scientific studies of framing effects shift the burden of proof and create a need for independent confirmation. That is enough to refute philosophical moral intuitionism, which claims that moral judgments do not need independent confirmation when they seem intuitively obvious to us. This refutation of moral intuitionism does not prove that moral judgments are not justified, but it does show that they are not justified by intuition alone. That is the primary challenge of framing effects in moral judgment. We would like to see May respond to it.

Baselines for human morality should include species typicality, inheritances, culture, practice, and ecological attachment

Darcia Narvaez 

Department of Psychology, University of Notre Dame, Notre Dame, IN 46556.
dnarvaez@nd.edu
<https://www3.nd.edu/~dnarvaez/>

doi:10.1017/S0140525X18002625, e163

Abstract

Empirical studies involve WEIRD (Western, European, industrialized, rich, democratic) but also un-nested (raised outside humanity's evolved nest) and underdeveloped participants. Assessing human moral potential needs to integrate a transdisciplinary approach to understanding species typicality and baselines, relevant evolutionary inheritances beyond genes, assessment of cultures and practices that foster (or not) virtue, and ecological morality. Human moral reason (*nous*) emerges from all of these.

May (2018) has waded impressively through a great deal of empirical research and philosophical argument to propose an account of “optimistic rationalism.” He has many ideas about how to deal with the inconsistencies found in experimental research. Much of what he proposes aligns with my view of moral complexity where moral functioning involves the conscious deliberative mind interacting with numerous subconscious processes – including, preferably, *well-educated* intuitions built from appropriate experience (Narvaez 2010). Still, I find his view of morality and human nature narrow and pessimistic because he does not address species typicality, baselines for morality, evolutionary inheritances beyond genes, cultures and practices of virtue, and ecological morality.

Species typicality

May implicitly adopts the common view that current psychological research assesses species-typical moral functioning, at least to a reliable degree. To his credit, May briefly mentions the WEIRDness (Western, European, industrialized, rich, democratic; Henrich et al. 2010) of the persons populating most data sets. These samples are also the source for most theories and conclusions drawn about human nature. He does not take these facts to their conclusions – that the nature of human nature cannot be established from such samples. A broader scope is needed to establish species typicality. The rest of my critique focuses on several aspects regarding the need for setting transdisciplinary-informed baselines when discussing human psychology and morality.

Baselines for morality

May provides no real empirical baseline for typical moral functioning of the human species apart from experiments in (mostly) social psychology. Although WEIRDness is important to realize, there are two additional features of most research participants that should influence the interpretations of these psychological studies. The first is a critique that others have raised – that participants in psychological experiments are mostly undergraduate sophomores (around age 19), which is especially important to note when studying morality. Undergraduates typically are not yet adults in their executive functions (e.g., foresight, empathy) or practical wisdom (Arain et al. 2013). So we should not expect to be able to assess typical adult moral functioning in this age group, for reasons of life experience and cognitive development (Rest et al. 1999).

But there is a less widely known reason to question empirical research results as representative of humanity. Most if not all participants have likely been raised outside of humanity’s species-typical developmental system – outside our evolved nest (what we can call “un-nested”). Why does this matter? Humans are more immature than any other ape at birth (presenting like fetuses until 18 months of age; Trevathan 2011) and have a several decade long maturational schedule. Early experience especially bears on neurobiological development, influencing health and well-being for life (Shonkoff & Phillips 2000) but also sociality and morality (Narvaez 2014). The human nest in early life – whose long term importance is corroborated by developmental and neuroscientific studies (e.g., for reviews see Narvaez et al. 2013b; Schore 2003a; 2003b) – includes soothing perinatal experiences (no separation of baby and mother or painful procedures), nearly constant (positive) touch, several years of infant-initiated breastfeeding, responsiveness that keeps the child optimally aroused, self-directed free social play in the natural world, positive social climate, and a community of responsive caregivers and

support (Hewlett & Lamb 2005). Humans who grow up in our ancestral environment (mobile small-band hunter gatherers, the type of society that represents 95–99% of human existence; Lee & Daly 2005) are raised within the evolved nest and demonstrate greater sociality (e.g., greater self-control and cooperation) (e.g., Ingold 2005; Narvaez 2013). A violation of a child’s “blueprint for normality” (Winnicott et al. 1989, p. 264) or lack of experience-expected care (Greenough & Black 1992) during sensitive periods (Knudsen 2004) –that is, a degraded nest – undermines neurobiological development, arresting or impeding social development (Schore 2003a), pushing a child’s trajectory toward relational disconnection and lifelong stress reactivity (Lupien et al. 2009). Thus, is it vital to take into account humanity’s species evolutionary history and not attend only to contemporary culture, practices, and behavior.

Evolutionary inheritances

Although May cites some evolutionary theory, there are critical aspects missed that bear on morality. Humanity’s evolutionary inheritances beyond genes are multiple (Jablonka & Lamb 2005; Oyama et al. 2001) and include not only basic needs and the evolved nest that fulfills them, but also self-organization around experience because of a highly dynamic, socially constructed human nature (Overton 2013). The plasticity and epigenetic malleability of human brains and body systems, especially early in life, is a characteristic not shared with chimpanzees (Gómez-Robles et al. 2015). Further, Charles Darwin (1871) in *The Descent of Man* noted humanity’s inheritance of the moral sense (social pleasure, empathy, concern for the opinion of others, habit control), which he found universal among preindustrialized societies. However, rather than being innate as Darwin implied, the moral sense appears to require postnatal experience that aligns with the evolved nest (Narvaez 2017; 2018b), as data from our lab is suggesting (e.g., Narvaez 2016a; 2018a; Narvaez et al. 2013a; 2016). Darwin found the moral sense less apparent in his British male compatriots (whose childhoods are far from supportive; Turnbull 1984), perhaps in part because boys have less built in resilience and take longer to mature, and thereby are more greatly affected by early life experience (Schore 2017). Postnatal early life shapes capacities critical for moral functioning such as self-control, empathy and cooperation (e.g., Kochanska 2002; Thompson 2012). The evolved nest extends beyond the mother and close caregivers to the community and culture.

Culture and practice of virtue

May makes no mention of a culture’s influence on moral development and does not evaluate contexts for development. The lack of attending to the cultural level is a common problem within psychology too where psychopathologies have been normalized and societal members are instead helped to adjust to societal impositions of individual isolation, impersonalism, and disconnection, among other dehumanizing things outside of our species-normal experience (Kidner 2001; Narvaez 2016b; Narvaez & Witherington 2018). In contrast, non-industrialized societies would consider U.S. culture to be quite harmful to the development of virtuous behavior because of a degraded evolved nest and missing practices (described later). Undermining human development and disrupting relationships in the ways described earlier foster relational and emotional

disconnection, sources of danger for everyone because they lead to harm of others (Lee 1979; Ross 2006).

Among native American communities, virtue development is a lifelong practice. Although a person's ability to get caught up in ego or misguided behavior is assumed in virtually all societies, among traditional native American communities, humans are raised to be on a path of continued self-improvement (Deloria 2006; Four Arrows 2016). In these societies, the evolved nest extends across the life span because human beings can become imbalanced outside of supportive relationships and require ongoing attention to relational harmony and balance. Community life in traditional societies entails rituals, practices, and stories that keep members focused on humility and proper relationships with others. Morality is about living in the right way and acting in the right manner in every circumstance. Moral slippage can occur when an individual's ego inflates or her responsibilities to the community are forgotten. As these societies are aware, a person's mindset can be shifted away from (or toward) relational trust and connection, making it critical to keep the self in an appropriate mindset (gratitude, humility) and aware of relational connection to all entities (WindEagle & RainbowHawk 2003). A harmful behavior reverberates across the social fabric undermining trust, and so healing circles for expressing and mending harm are part of a process of restorative justice common in native American communities (Ross 2006). Consequently, justice has to do with repairing relationships – restoring respectful and caring connection – toward self, others, community, landscape, and the unseen spiritual world. In contrast, societies like the United States assume disconnectedness as part of a capitalistic human life and set up institutions and practices that, perhaps mindlessly, undermine connection (e.g., person-to-person, person to community, person to natural landscape) (Kidner 2001).

Ecological morality

May makes no mention of ecological morality – moral mindedness toward other-than-humans. The morality he describes seems to have no grounding in living on and with the earth. In contrast, for most societies across history, treating the local landscape with humble respect was part of the moral life (Descola 2013; Merchant 2003; Nelson 2008). Indigenous or native science (Cajete 2000) is holistic, understanding that everything is connected (as, for example, Western physics and biology have noted at the quantum level and in terms of shared DNA), even into the future (seven generations), with a responsibility to promote flourishing of the landscape, not just of human individuals or communities. Indigenous societies are highly attached to and respectful of the landscape (Narvaez et al. 2019). In contrast, un-nestedness and disconnection are characteristic of industrialized, capitalistic societies, leaving their members detached from one another, as well as from the natural world (Polanyi 2001), driving the many ecological and social crises that threaten biodiversity and life on the planet (Intergovernmental Panel on Climate Change [IPCC] 2014; Kolbert 2014; Millennium Ecosystem Assessment 2005). Should not this be part of the conversation about morality, especially in a discussion of rationalism? It is highly irrational in any way you slice it to be destroying planetary ecological systems and futures for the sake of money and power. One would hope that researchers and philosophers would enlarge their purview to include how individuals and

communities live their lives within ecological systems on a day to day basis.

Finally, moral evaluation is not the same as moral decision making or virtuous behavior. Armchair, detached observance and judgment of the world do not have much to do with morality in the flesh. In this regard, May seems to miss a key notion from our historical past, the distinction between reason (Greek *nous/noos*) and rationality (Greek *logos/dianoia*). The former is shaped by experience and involves one's whole being (e.g., embodied cognition, including intuition) and was considered the prior and superior faculty to the latter. This form of reason is "evolutionary [...] makes use of forms of perceptual and motor inference present in "lower" animals...mostly unconscious...largely metaphorical and imaginative...not dispassionate, but emotionally engaged" (Lakoff & Johnson 1999, pp. 4–5). In contrast, *logos* refers to the more explicitly rule-focused, utilitarian, mechanical, detached, internally consistent type of knowledge, which unfortunately characterizes most moral psychological experiments. Perhaps the many human weaknesses or failings in moral functioning observed in experiments are tapping into a lack of appropriate cognitive, emotional and social experience to build embodied personal knowledge (*nous*), leaving individuals (including research experimenters) to rely on semantic knowledge (*logos*). The focus on *logos* instead of *nous* fits with the Western world's shift to a "left-brain" focus on static objects instead of on the shifting patterns of connection among living dynamic entities (McGilchrist 2009; Muller 2018), a focus characteristic of those with brain damage, not representative of human potential. David Kidner's (2001) question for psychology comes to mind as a question for philosophy too: "why should the social sciences be based on the one power which separates humans from [...] other organisms"? Why not, instead base it on the *many* powers that *relate* us to other organisms" (p. 59, italics in original)? Indeed, why should philosophy dig itself into the same hole, especially when it appears that detached thinking has been instrumental in creating the ecological crises faced today?

In conclusion, a wider examination of human behavior across time and societies is needed when discussing human moral potential. Few people are well educated or well developed in virtuous morality in industrialized societies like the United States for the reasons mentioned, making humanity appear to be innately morally flawed. Individuals and culture vary in their opportunities and support for virtue development, demonstrating that a focus only on the individual is inadequate. Cross-generational effects on development and cultural factors matter greatly. Perhaps in a follow-up book May can take up a broader scope and include more of the factors that contribute to humanity's moral potential, focusing then on "optimistic reasonableness."

Kantian indifference about moral reason

Adam J. Roberts 

Holywell Manor, University of Oxford, Oxford OX1 3UH, United Kingdom.
adam.roberts@oxon.org

doi:10.1017/S0140525X18002650, e164

Abstract

The pessimistic arguments May challenges depend on an anti-Kantian philosophical assumption. That assumption is that what I call *philosophical* optimists about moral reason are also committed to *empirical* optimism, or what May calls “optimistic rationalism.” I place May’s book in the literature by explaining how that assumption is resisted by Christine Korsgaard, one of May’s examples of a contemporary Kantian.

In the first chapter of *Regard for Reason in the Moral Mind*, May (2018, p. 5) claims that moral theories in the tradition of Kant “have taken a serious beating” from sceptics about the role of reason in moral cognition. “To be fair,” he says, “Kantians do claim that we can arrive at moral judgments by pure reason alone,” at least in the sense that they think we can arrive at basic moral principles without appealing to emotions. In that sense, they are what we might call *philosophical* optimists about moral reason. However, exactly they think of reason, they believe that it alone can rule some general moral judgments in or out.

May follows much of the experimental literature in taking Kantians to also be committed to an *empirical* optimism about moral reason, or what May (p. 4) calls “an *optimistic rationalism*.” There are at least two reasons why one might think that Kantians have that empirical commitment. The first is particular, to do with their premises, and the second is more general, to do with reason’s priority. First, one might think it plausible that the premises of Kantians’ moral theories could be undermined by research into the psychological grounds of our accepting them. Kantians would be empirical optimists in assuming that such research would not undermine their premises if there was a modest chance those premises could indeed be “empirically debunked,” (p. 80) in “the now common epistemological” sense (p. 83).

Second, one might think that if the science seems to show our moral judgments are driven by emotion, then surely emotion – rather than reason – ought to play the basic role in our moral philosophy. There are more and less “common sense” ways of trying to argue this. Among the more philosophical ways, one might try to appeal to some kind of motivational internalism or to a science-first methodology. To really press the point, one might attempt a version of an argument Kant (1996, pp. 4:448–53) himself has been read as making: If morality consists in rational requirements, it only binds us if we are rational – and we are not on questions of morality.

What any individual Kantian is committed to is neither here nor there. However, at least some Kantians are not obviously committed to anything deserving to be called empirical optimism. They are what I have called *philosophical* optimists, but as to whether our moral judgments are emotionally driven, rather than being optimists, pessimists, or something in between, they instead defend an active indifference. On the one hand, these Kantians take their theories to have premises which are not vulnerable to empirical debunking in the same way as intuitions about what to do in cases or what kinds of thing morally matter. On the other, they argue that reason must play the basic role in moral philosophy not because it drives us, but because we need its concepts to help us make the right moral judgments. To make the point I want to in this commentary – about where May’s book sits in the literature – I will have to explain one Kantian’s position in a little detail.

Perhaps the best-known defender of a Kantian kind of empirical indifference is Christine Korsgaard, one of May’s (2018, pp. 5–6, 176, 179, 183–85) examples of a contemporary Kantian moral theorist. The premises of Korsgaard’s (e.g., 2009b, pp. 18–26, 64–67) arguments are descriptive claims in the philosophy of action, concerning how we conceive of what we are doing when trying to act. Those conceptions cannot be false as such, because they are not meant to correspond to some fixed features of the world (cf. Kant 1996, pp. 5:54–57; Korsgaard 2008, pp. 322–24). As our own conceptions, there is also plausibly a limit to how wrong our claims about them can be.

For those claims to be *debunked*, we would have to discover that something like an irrelevant emotion affected how we thought we conceived of our agency but not how we actually conceived of it. As I understand her, Korsgaard (cf. 1996b, pp. 254–58, 125–26) does not mean to leave much space for that possibility. How we think we conceive of our agency is at least part of how we do conceive of it. What we can be mistaken about is what claims we are committed to about our agency whatever else about it we may also happen to think. Korsgaard’s (1996b, pp. 113–25; 2009b, pp. 20–25) best-known argument for our having Kantian commitments starts from our own particular conceptions of our agency, not from intuitions about those commitments. It attempts to show that we must take our common humanity to be a source of reasons to take ourselves to have reasons as teachers, lovers, citizens, or whatever else.

To be justified in believing her conclusion, Korsgaard might have to be justified in taking herself to be able to make and follow valid conceptual arguments. I think Korsgaard would argue the justification need not be empirical, but even if it had to be, supposing one was possible would not obviously make her an optimist. She could still think that swathes of intuitions about moral dilemmas are ripe for debunking, that moral knowledge based on such intuitions is impossible, and that virtue is unattainable. What she would be supposing in offering an empirical justification – and as I said, I do not think she would – is the falsity of the “truly global skepticism” May (p. 22) leaves outside of the scope of his book.

If Korsgaard is committed to empirical optimism, then, it does not seem like it is in virtue of her premises. That still leaves the possibility all Kantians become committed to that optimism by giving a more basic role to reason than emotion in their moral theories. Again, however, someone like Korsgaard is going to argue against such a commitment. The starting point of such an argument might be that however sure we are our moral judgments are made for us, we still have to grapple – inside our own heads – with the deliberative task of trying to make those judgments. The significance of rational principles does not rest on our being driven by reason rather than emotion, but rather on the fact we have to reason *even if* we are driven by something like emotion.

Korsgaard (1996a, pp. 162–63; 1996b, pp. 94–97 and cf. 238–42) makes at least two versions of that argument. In the first, she asks us to suppose that we know all our reasoning is guided by a device implanted in our brains. In the second, she imagines that someone can predict everything that she is going to do. In both cases, her claim is that we would still face the deliberative task of trying to make our own choices (cf. Hill 1992, pp. 116–19, 131–38). Within the scope of dealing with that task, there would still be a role for rational concepts.

At most, our moral judgments being determined by our emotions might show that in some sense we are not responsible for them (but cf. Korsgaard 1996a, pp. 188–212). In addition, our being driven by the *wrong* emotions might well put some kinds of moral virtue out of reach. May (pp. 230–37) never sounds more Kantian than when he closes *Regard for Reason in the Moral Mind* with a discussion of moral enhancement. Korsgaard (1996a, p. 324) claims “it is not an accident that the two major philosophers in our tradition who thought of ethics in terms of practical reason – Aristotle and Kant – were also the two most concerned with the methods of moral education.” Neither of them premised their moral thought on an assumption that we were guaranteed or likely to be rational, but they also did not think it followed that when making moral judgments, we need not try to choose rationally.

In fact, Kant does not *exactly* think that we ought to try to be rational (cf. Korsgaard 2009b, pp. 153–58). He thinks we ought to try to act as a free will would, and that this comes to the same thing as trying to be rational. As I understand it, Kant’s (1996, pp. 4:440–55 or 5:28–33) basic idea here is straightforward. It begins with the claim that we only get to determine our actions if we have free will. It follows that there are no ways we can determine ourselves to act which are inconsistent with our having free will. In that sense, when trying to determine how we act, we can take it for granted that we have free will. That claim naturally extends to others when we are trying to make judgments about what they should do (cf. Korsgaard 1996a, pp. 200–12).

As I mentioned, Kant (e.g., 1996, pp. 4:451–52) draws a connection between free will and rationality. On his view, reason is fundamentally just our capacity to be genuinely active, and it has principles because there are conditions of the different ways we might be genuinely active (cf. Korsgaard 2018, pp. 132–34; 2009a, pp. 32–38). When Kant says we can take it for granted that we are able to act with free will, he is also saying we can take it for granted that we are able to be rational. It does not matter how convinced we are that we are driven by forces like our emotions. What matters is that we still face the deliberative task of trying to choose our moral judgments for ourselves. Kantians like Korsgaard argue that there is a proper way of going about that task, and they are philosophical optimists because their arguments do not appeal to our desires or emotions (cf. e.g., Street 2010, pp. 369–70; Velleman 2009, pp. 147–49).

If I am right about Korsgaard’s commitments, at least, then philosophical and empirical optimism are separate. Behind much of the debate in which May is engaging, however, is an assumption that the former requires the latter. The only way that reason can be central to morality *seems to be* if it is central to our moral psychology. That assumption comes quite naturally if we do not separate “speculative” or “theoretical” questions from “practical” ones in quite as radical a way as Kant did (cf. Allison 2004, pp. 47–49; Korsgaard 1996a, pp. 167–76, 201–205). In other words, the assumption that the two kinds of optimism go together is itself anti-Kantian. It supposes there is no deep, perspectival divide between psychology and ethics, or the tasks of trying to explain a part of the world and trying to act in it.

Of course, not everyone is a Kantian, and not every Kantian is like Korsgaard. As I mentioned earlier, the point I want to make here is one about where May’s book fits into the literature. *Regard for Reason in the Moral Mind* is a challenge to pessimists about moral reason, but a challenge made on the pessimist’s terms. If it succeeds, it shows their position is undermotivated

even granting their philosophical assumptions. A Kantian like Korsgaard, however – an optimist in one sense – would not grant those assumptions to the pessimist. They would be unsettled by arguments for the claims May’s granting, but not otherwise for the claims he is challenging.

The space between rationalism and sentimentalism: A perspective from moral development

Joshua Rottman 

Department of Psychology, Franklin and Marshall College, Lancaster, PA 17604.
jrottman@fandm.edu www.joshuarottman.com

doi:10.1017/S0140525X18002698, e165

Abstract

May interprets the prevalence of non-emotional moral intuitions as indicating support for rationalism. However, research in developmental psychology indicates that the mechanisms underlying these intuitions are not always rational in nature. Specifically, automatic intuitions can emerge passively, through processes such as evolutionary preparedness and enculturation. Although these intuitions are not always emotional, they are not clearly indicative of reason.

In *Regard for Reason in the Moral Mind*, May (2018) acknowledges that moral judgments and behaviors are frequently produced by automatic intuitions. May argues that intuitive cognitive processing is best categorized as “reasoning” because it is not heavily dependent upon emotional responses. Thus, May aligns these intuitions with a rationalist (rather than sentimentalist) framework and suggests that these intuitions are not substantively threatened by debunking arguments. However, to successfully vindicate moral cognition on the grounds that it is rooted in reason, it is crucial to determine that intuitive moral cognition truly arises from inferential processes – ideally, those that move from well-justified premises to logically warranted conclusions. Otherwise, moral intuitions can more easily be dismissed, because debunking arguments rely primarily on the *irrationality* or *unreliability* of everyday moral judgments rather than on their *emotionality* (e.g., Sinnott-Armstrong 2011). Therefore, regardless of whether emotions are the primary fuel for moral judgments and actions, it is crucial to determine the extent to which these judgments and actions are aligned with reason to prevent them from being discredited.

Moral cognition, like all cognition, involves information processing. However, the complexity of this processing can vary widely. Some moral evaluations result from careful consideration of clearly represented concepts, whereas others involve no internal representations and are therefore considerably more inflexible and error-prone (e.g., Crockett 2013; Cushman 2013). Therefore, even if moral competence can be described as operating in accordance with certain principles (e.g., intentionally causing harmful outcomes is morally worse than inadvertently allowing harm to occur), this is consistent with a range of psychological

mechanisms ranging from reasoned inference to unreasoned instinct. Although the latter is not necessarily aligned with sentimentalism (as it may not be driven by emotional responding), it is also not clearly aligned with rationalism. Instead, many automatic intuitions defy this binary opposition and instead exist in a liminal space between these philosophical strongholds. A crucial empirical question is therefore raised: Are most moral intuitions produced by processes of inductive or deductive reasoning, or are they formed by less rational means?

Developmental psychology provides a crucial tool for assessing the rationality of moral intuitions, as it can uncover the sources of intuitive responding. Some cognitive developmental processes are clearly aligned with rationalism, whereas others can reveal moral intuitions to be independent of reasoning (see Shweder et al. 1981). As May (2018) proposes, intuitive automaticity could eventually result from an extended rehearsal of conscious reasoning, just like a chess expert is able to spontaneously make adept moves after internalizing the careful thinking that she exerted across many previous games (see also Pizarro & Bloom 2003; Saltzstein & Kasachkoff 2004). However, this seems unlikely, as people are often unable to consciously recover the principles that underlie their moral judgments (e.g., Cushman et al. 2006; Rottman et al. 2014), suggesting that these intuitions may have never been consciously produced to begin with. Alternatively, some developmental psychologists have argued that young children can acquire intuitive frameworks for moral reasoning as a result of rational inference (e.g., Rhodes & Wellman 2017). However, just because intuitions *can* result from rule-governed inferences does not mean that they *typically* do, and recent research on moral development has indicated that babies and children possess a wide array of adaptive moral predispositions that do not appear to be the result of rational inference (see Bloom 2013; Rottman & Young 2015). Therefore, I suspect that children's (and plausibly adults') moral competence can most accurately be described as occupying a middle ground between rationalism and sentimentalism.

From an evolutionary standpoint, it would be maladaptive to rely on one's logical reasoning abilities to reach moral conclusions, as reason would not necessarily converge upon beliefs that successfully promote social status and coordination (see Krebs 2008). Instead, it is likely that moral competence is primarily composed of innately prepared intuitions and learning mechanisms that are modulated by relevant environmental inputs during childhood (Rottman & Young 2015). Recent evidence from research with infants and young children suggests that many morally relevant intuitions are in fact the nascent products of evolutionary adaptation. These intuitions exhibit signatures of evolved psychological traits, for example, being spontaneously acquired in ways that do not rely heavily on protracted learning (see Dunham et al. 2008) and emerging so early in life that it is unlikely that they result from rational inference or relevant experiences (see Hamlin 2013). Young children also think about morality in domain-specific ways, and these features of moral cognition that transcend domain-general reasoning tendencies appear suited to resolve adaptive problems related to sociality (Cummins 1996).

Other moral intuitions rely heavily upon individual learning and enculturation, but it is similarly unlikely that this acquisition process typically involves reasoning. Rather, children are prone to blind conformity in the moral domain and are predisposed to promiscuously moralize a wide range of actions upon brief exposure to normative behaviors (see Chudek & Henrich 2011;

Rakoczy & Schmidt 2013; Tomasello 2016). A recent set of studies has indicated that learning new moral beliefs is not always a rational endeavor (Rottman et al. 2017). This research failed to support a strong sentimentalist view, as incidentally elicited disgust was insufficient for producing moralization. However, children acquired novel moral beliefs in irrational and undiscerning ways. Participants were equally persuaded by "well-fitting" and "poor-fitting" explanations, suggesting that children do not attend to the rationality of the testimony they are provided during the process of forming new moral beliefs, and they often lacked the ability to reconstruct the processes leading to their formation of moral beliefs when learning from emotion-laden testimony. Of course, this research is not conclusive by itself, particularly as it does not align with theoretical perspectives that children should only learn from testimony that they discern to be appropriate and relevant (e.g., Grusec & Goodnow 1994; Nucci 1984; also see Sobel & Kushnir 2013), and considerably more research is needed to more fully understand typical processes of moral acquisition.

Turning from moral thought to moral behavior, children sometimes appear to be motivated by virtue; they are spontaneously prosocial in certain affiliative situations when they can help others at a small cost to themselves (e.g., Warneken & Tomasello 2006). However, this prosociality is selective and strategic (see Martin & Olson 2015). In particular, when children stand to achieve a relative advantage, their behaviors are typically motivated by selfish gains. Even when they clearly understand how they *should* act in moral situations, they often choose to act in self-interested ways instead (see Blake et al. 2014). Children are strongly motivated by *appearing* moral rather than by actually *being* moral (e.g., Engelmann et al. 2013; Leimgruber et al. 2012; Shaw et al. 2014), and it takes many years for them to begin to overcome these egocentric tendencies (to the extent that they succeed at all).

On the whole, a review of recent developmental research uncovers sparse evidence that rationalism successfully accounts for moral cognition in infants and children. Instead, there is reason to conclude that moral intuitions are often *irrational*. Children's moral intuitions are constrained by innate representational biases that are diversified through sociocultural learning, rather than actively formed through reasoned inferences about social interactions. Of course, biased intuitions that are similarly irrational, motivated, and inaccessible to introspection have also been argued to characterize much of adult moral cognition (e.g., Greene 2013; Haidt 2001), but May (2018) argues that the evidence for these biases either falls short or is limited in scope. Developmental evidence has the potential to bolster the pessimists' claims even further, however. First, studying development can rule out some alternative interpretations of automaticity (e.g., that it results from initial judicious deliberation). The processes leading to moral belief formation may be more generally defective than is evident from studies of adult moral cognition, thus surmounting the "Debunker's Dilemma." Second, early development may be a time when motivations are particularly egocentric and situational, and thus poor motivations are sometimes sufficiently pervasive for surmounting the "Defeater's Dilemma."

Overall, although I disagree with many of May's (2018) conclusions, I applaud the many redeeming qualities of this impressive treatise. Throughout its thorough consideration of a wide swath of evidence, this book provides an important counterweight to oppose the strong force of the sometimes overblown claims that morality is wholly driven by emotions and egoism rather than by

reason and virtue. The sentimentalism that has largely taken hold in social psychological approaches to moral psychology (e.g., Haidt 2001) has sometimes obscured the cases in which reason can play at least a limited role in moral cognition (e.g., Holyoak & Powell 2016; Paxton & Greene 2010; Pizarro & Bloom 2003). This emphasis on rapid emotional responding is reflective of a more general tendency to focus on the irrational, motivated, and biased nature of human thought that has prevailed in the field of social psychology as a whole (see Alter 2013; Bargh 2017; Nisbett & Wilson 1977). On the contrary, many developmental psychologists have sought and often found evidence for the rational, scientific, and objective nature of children's thought (see Gopnik 2012; Schulz 2012; Xu & Kushnir 2013). In recent years, dozens of elegant studies have demonstrated that children can use scientist-like reasoning to form and revise beliefs (e.g., Schulz et al. 2007; Sobel & Kirkham 2006), indicating that children rely heavily on reason in certain contexts. This characterization has also reigned in classical theories of moral development, which posit that moral judgments are produced by careful reflection (e.g., Kohlberg 1971; Nucci & Turiel 1978; Piaget 1932; Smetana 2006). However, just as adults are not as asinine as social psychologists often characterize them, children are not as astute as developmental psychologists often characterize them. This may be especially true in the moral domain, for which affiliative motivations tend to reign over truth-seeking and it is difficult (if not impossible) to construct knowledge through individually acting on the world.

Descriptively, there are myriad possibilities for characterizing the nature of moral cognition. As reviewed here, research in moral development has indicated that emotional forces do not ubiquitously drive moral evaluations and behaviors, but neither does careful inductive reasoning. There is an intermediate space between sentimentalism and rationalism that may most accurately characterize everyday moral psychology. Therefore, regardless of whether emotions are shown to be unnecessary or insufficient for moral development to occur, despite some arguments to the contrary (e.g., Eisenberg 2000; Hoffman 1975; Kagan 1987), the veracity of rationalism would not necessarily hinge upon the success of these demonstrations. Even if sentimentalism is found to be empirically false, the unreasoned and heuristic nature of many moral intuitions prompts a cautious pessimism regarding the nature of moral cognition. While this stance is certainly less pleasant than optimism, it may be beneficial for avoiding complacency. A healthy dose of pessimism can serve as motivation for fostering a more humane world, perhaps by investigating ways to encourage future generations to overcome natural moral inclinations. By resisting the tendency to consider moral "truths" to be self-evident and by vigilantly entreating children to apply careful reasoning to crucial moral issues, it may be possible to nurture moral cognition in the direction of rationalism.

Humean replies to *Regard for Reason*

Neil Sinhababu 

NUS Philosophy, National University of Singapore, Singapore 117570.
neiladri@gmail.com
<https://www.neilsinhababu.com>

doi:10.1017/S0140525X18002704, e166

Abstract

First, I argue that the Humean theory is compatible with the commonsense psychological explanations May invokes against it. Second, I explain why desire provides better-integrated explanations than the mental states May describes as sharing its effects. Third, I defend individuating processes by relata, which May rejects in arguing that anti-Humean views are as parsimonious as the Humean theory.

May's (2018) *Regard for Reason in the Moral Mind* is a novel and important defense of the view that reason guides moral thought and motivation in human beings. Although rationalist views of moral psychology have many defenders, few engage in as much detail as May with empirical arguments from situationists, egoists, and Humeans. If rationalists avoid these challenges, they face the criticism that although their theories describe a possible moral psychology, it is not the one that human beings have. May is not afraid to get his hands dirty with the empirical data, and much of his book responds in detail to his opponents' empirical arguments.

Although I doubt situationism and reject egoism, I defend the Humean theory of motivation. May and I agree on how the Humean theory should be formulated: It includes commitments both to the necessity of desire for motivation, and to the impossibility of generating new desires by reasoning from beliefs alone. Desire, then, is not merely an immediate motivator of action that reason can summon up on command. It is the fundamental source of all motivation, and new motivation cannot be generated without it.

Formulated this strongly, the Humean theory is incompatible with the view that moral judgments are beliefs with intrinsic motivational (or desire-generating) force. Creatures with Humean psychologies cannot make moral judgments that fit this cognitivist and internalist model. May regards moral judgments as beliefs that can generate desires this way, and therefore must reject the Humean theory.

In *Humean Nature* (Sinhababu 2017), I argue that the Humean theory is part of the best explanation of how we think, feel, and act. Desire does not just motivate action. It causes pleasant and unpleasant feelings when we have various sorts of thoughts about its object, and it directs our attention toward its object in various ways. Because of its emotional and attentional effects, desire is well-suited to explaining the thoughts and feelings that arise in practical deliberation and various other phenomena like procrastination and daydreaming.

May responds to my arguments at length after describing me as "the best 'philosophical nemesis' one could ask for" (p. xii) in the preface. Here I will try to live up to his praise by defending the Humean theory against three different lines of argument he makes against it.

First, May argues that the Humean theory runs against commonsense explanations of human motivation that we often rely on. He cites examples of people who describe their own moral motivation as the result of a belief that something is the right thing to do. Then he argues that "We often describe one another, and ourselves, this way – as ultimately motivated by beliefs with normative or evaluative content" (p. 180). On May's view, the content of such beliefs enables them to generate new desires to act accordingly, violating the Humean theory. I agree with May that it would be a problem for the Humean theory if our intuitive

folk-psychological theory was committed to the possibility of this kind of fundamentally belief-driven moral motivation. Folk psychology may not always be right, but it works well enough that there is a cost to denying its core commitments. And if people who explain their moral motivation in terms of their moral beliefs really were insisting on an anti-Humean explanation of their motivation, Humeans would have a problem.

Fortunately for Humeans, there is no reason to see “I did it because I believed it was the right thing to do” as an anti-Humean explanation of motivation. The belief may have played its motivational role only by combining with a desire to do the right thing. People often explain things by pointing out one particularly salient explanatory factor, and in doing so they are not denying the presence of other explanatory factors. If someone explains that she did not eat the mushroom because she believed it was poisonous, she is not committing herself to an anti-Humean psychology where beliefs about poison have intrinsic motivational force. It is perfectly consistent with her explanation that a pre-existing desire not to be poisoned combined with her belief and motivated her not to eat the mushroom. When we explain things to each other in ordinary conversation, we do not usually name all the causal factors – that would take too long and bore our audience. We name some, and let our audience infer the others. If that is all people are doing when they mention only their moral beliefs in explaining motivation, such explanations leave plenty of room for desires to do the right thing, and thus provide no evidence against the Humean theory.

Second, May argues that many mental states other than desire can produce the phenomena that I credit the Humean theory with explaining. I take desire to motivate action, cause pleasant and unpleasant feelings, and direct attention, with all these effects taking greater magnitude when the desire’s object is vividly represented. My argument for the Humean theory is that desire, so conceived, provides the best explanation of a variety of psychological phenomena. May responds that many other psychological states can do these things as well (pp. 192–95). He argues that habits can cause behavior without amounting to desires as I conceive them, sensory stimulation that does not involve desire can generate pleasant or unpleasant feelings, attention can be directed by psychological associations that are not grounded in desire, and vivid representations of obscenities can make a Tourette’s sufferer more likely to tic by saying the obscenities. I accept much of this. If other mental states have many of the effects that desire does, why are Humean explanations of our action, thought, and feeling superior to anti-Humean explanations?

My answer is that Humean explanations provide a better-integrated explanation of motivation, feeling, attention-direction, and the effects of vividness than anti-Humean explanations do. Suppose we are trying to explain why a very hungry person was pleased to be told he would soon be served a delicious meal, why his hunger prevented him from paying close attention to the boring dinner-table conversation around him, why he became especially excited to eat when the food was brought out before him, and why he ate with enthusiasm. Desire for food has the hedonic, attentional, vividness-related, and motivational effects to explain all of this at once. The other sources of these effects that May discusses would not fit into such a well-integrated explanation. Should we posit a sudden pleasant sensory experience when he was offered food, a set of psychological associations distracting him from the dinner-table conversation, something like the ticcing of a Tourette’s sufferer when the food is brought out, and a habit of eating whatever is on a plate before him?

Often we will not have any reason to posit such a disconnected hodgepodge of psychological factors. The fact that the things we are motivated to pursue attract our attention and cause pleasant and unpleasant feelings, and that all of these effects are amplified by vivid representations, is best explained in terms of a unified psychological state with all these effects. We cannot count on the other factors May cites to come together frequently enough to explain the phenomena.

Third, May criticizes my frequent appeals to parsimony. I claim that the Humean theory is more parsimonious in treating instrumental reasoning, where a desire for an end combines with a means-end belief to generate a desire for the means, as the one and only way that reasoning can generate a desire. Continuing a debate that we have had in previous work, May argues that the Humean theory may not actually be more parsimonious. As he notes, I am individuating processes by their relata. This makes instrumental reasoning (where the relata are a desire for an end, a means-end belief, and a new desire for the means) a different process from anti-Humean desire generation (where the relata are a normative belief and a new desire for the normatively favored course of action). He argues that we do not always individuate processes this way: “we don’t posit two kinds of baking or two kinds of corrosion just because the relationship can hold between different entities. A human or a robot can bake a cake (or a quiche); water or acid can corrode a pipe (or a rock) [...] We needn’t posit two kinds of motivational process just because one is initiated by a desire while the other is initiated by a belief” (p. 197). The upshot is that Humean theory is not more parsimonious than opposing views, as instrumental desire-creation and anti-Humean desire generation can be treated as the same process. We just have to give up the assumption that processes are individuated by their relata.

Here I will defend the idea that processes are to be individuated by relata, and that May’s baking and corrosion examples do not provide good analogies to the psychological issues at hand, because the differing relata he mentions are not essential to characterizing the processes. The reason we might not divide up baking into separate processes depending on whether the baker is a human or a robot, or the specific type of food, is that these are not essential to characterizing baking. What makes something an instance of baking (rather than say, frying or applying no heat) are a general way of applying heat and general sorts of effects on the food, not the identity of the baker or the precise nature of the dish. Similarly, what makes something an instance of corrosion is the interaction between metallic particles and ions, not the specific nature of the substance that contains the metallic particles or the liquid that contains the ions. Once we are sufficiently precise about the nature of the processes, we see that we do individuate them by their relata.

Maybe there is some more general level of explanation on which we could likewise treat instrumental reasoning and anti-Humean desire generation as instances of the same general process of reasoning, with the relata being very general – perhaps, “some antecedent psychological states” and “a new desire.” But admitting these general relata takes us away from the level on which my psychological debate with May is being conducted. We are advocating psychological explanations that invoke different intentional states, and the fact that the differing relata we invoke could be lumped together at some other level than psychology that invokes specific intentional states is neither here nor there. To illustrate the point: Is reasoning the same process as telepathy, because both are ways for psychological states to affect other psychological states? Maybe one could find an

explanatory level where generalizing the related and lumping these processes together makes sense. But at the level of psychological explanation, we need to distinguish these processes, and admit reasoning while rejecting telepathy. Similarly, psychology needs to distinguish between instrumental reasoning and anti-Humean desire generation. Whether we should believe only in the former, or also admit the latter, is at the heart of the debate between *Humean Nature* and *Regard for Reason in the Moral Mind*. To have that debate, we need to recognize the differences between these processes, rather than treating them as the same.

Rationalization, controversy, and the entanglement of moral-social cognition: A “critical pessimist” take

Robin Zheng 

Philosophy Department, Yale-NUS College, Singapore 13853.

robin.zheng@yale-nus.edu.sg

<http://robin-zheng.me>

doi:10.1017/S0140525X18002637, e167

Abstract

I raise two worries about the Debunker’s and Defeater Dilemmas, respectively, and I argue that moral cognition is inextricable from social cognition, which tends to rationalize deep social inequality. I thus opine that our moral-social capacities fare badly in profoundly unjust social contexts such as our own.

Joshua May’s (2018) new book is a tremendous and much-needed intervention in the field of moral psychology. May argues that, notwithstanding a voluminous literature on the various foibles, “bugs,” and biases to which we are prone, our moral capacities are basically sound. He walks us carefully through the many halls of this literature, executing a suite of reliability checks that show these findings to be less sensational than they appear in headlines. May performs a great service by systematically cataloging these pessimistic challenges all in one place, and his comprehensive, ably argued, many-sided defense will no doubt reinvigorate key debates over rationalism versus sentimentalism, consequentialism versus deontology, and egoism versus altruism.

In this commentary, however, I consider the perspective of an ordinary, self-reflective moral agent with no philosophical horse in these races – just a “concerned (moral) citizen,” as it were, wondering how worried she should be. I offer what might be called a *critical pessimism*: critical in the sense of being explicitly concerned with ameliorating injustice, and in the tradition of “normative reflection that is historically and socially contextualized” (Young 1990) – that is, sensitive to how moral reasoning is always embedded within a specific sociopolitical and historical moment. Writing on the edge of the 2018 U.S. mid-term elections, I suggest there is still reason for a certain pessimism about our moral capacities, even if they are not fundamentally arational and egoist.

I begin with two worries about the twin centerpiece of May’s case for optimism: the Debunker’s Dilemma (for skeptics of

moral cognition) and the Defeater Dilemma (for skeptics of moral motivation).

May uses the Defeater Dilemma to show that motivation can be genuinely moral, that is, it is not fundamentally egoistic. He concedes that a particular form of self-interest – desiring to see oneself as morally good – is indeed a main basis of wide-ranging moral behavior, as demonstrated by studies of motivated moral reasoning, moral licensing, and moral hypocrisy. Yet he argues that we could just as well consider such a desire to be a form of moral integrity, because it reflects an agent’s authentic concern for morality. Thus, rationalizations of bad moral behavior are themselves evidence of moral integrity, because self-serving outcomes are made to seem appealing *via* the enlistment of bona fide moral reasoning. I find this a compelling and ingenious argument against the egoist who believes true moral motivation is rare or non-existent.

However, accepting the widespread influence of this species of moral integrity seems to cut against arguments defending the general reliability of moral *cognition*. May writes, for instance: “You can criticize a politician who passes anti-discrimination legislation only to acquire votes, but not if she does it primarily because she wants to see herself as doing the right thing and would otherwise feel guilty” (p. 204). But here May is only assessing moral motivation, in a case where the desire to view oneself as moral works in the right direction. If we consider a moral rationalization case where it leads a self-righteous politician to pass voter-suppression laws despite initial qualms, we can and should criticize her – precisely *because* her moral integrity has corrupted her moral judgment.

May might respond by pointing out that this is not unique to moral cognition; indeed, he uses desire’s distorting effect on prudential reasoning to illuminate his account of rationalization. He only wants to demonstrate that moral reasoning is not inherently different from non-moral reasoning in being dependent on emotions, as sentimentalists would have it. However, there is an important disanalogy here. Prudential rationalization runs up against natural limits: those oft-cited indulgent desserts will actually prevent weight loss, and too many splurges add up to an empty bank account. In other words, bad prudential reasoning will as a matter of course work against an agent’s prudential interests, because it represents a mistaken theoretical understanding of the world. This is far more tenuous for moral rationalization, however, because moral reasoning is practical rather than theoretical: it concerns the world as it ought to be, not as it actually is. Hence, there is no natural “check” on bad moral judgments, making it possible for the moral rationalizer to dig in his heels against deleterious consequences. Others’ suffering, for example, can be rationalized by painting them to be sufficiently unlike oneself, be it less competent, less deserving, or less human. Thus, moral cognition can plunge into horrific abysses, as with May’s (and today’s) sincere Nazis (p. 176), because it is uniquely susceptible to distortion by the very desire to be moral – that is, by moral integrity itself. The more widespread moral rationalization is, the less reliable our moral cognition.

I am not, however, trying to develop this into a sweeping argument subject to the Debunker’s Dilemma, which demonstrates that there is no single factor – emotions, evolutionary influence, or framing effects – that threatens our moral reasoning as a whole. But note that skepticism does not require there be some “One True Debunker,” that is, the *wide-scope* claim that there exists some debunker universally undermining justification for moral

judgments across all contexts. We would also be in trouble if we accepted the *narrow-scope* claim that for all contexts, there lurks some distinct debunker undermining justification in that particular context. I do not think our concerned citizen need be a global skeptic of this sort, because as May demonstrates, it is doubtful that the skeptic will be able to substantiate the strong narrow-scope claim. However, this disambiguation points us toward a different flavor of skepticism that can still be quite concerning. The worry here is not: “Are *all* of my moral judgments unjustified because they are influenced by some universal debunker?” but rather: “Does *this* particular token moral judgment involve a specific context that is prey to a narrowly debunking influence?” May does endorse such targeted debunking arguments. But in the absence of systematic principles for determining whether or not a given moral judgment falls within the restricted range of these “danger zones,” we might lose justification for it.

Now, May does specify one kind of context where we should be wary: peer disagreement can serve to debunk *controversial* moral judgments. But the obvious problem here is that these are contexts where we might be most worried about the veracity of specific token judgments (“Am I right to believe that violent resistance is never justified? that it is right to protect the spotted owl at the expense of loggers’ jobs?” etc.). This is where our concerned citizen would most want to depend on the reliability of moral cognition: yet it is precisely here that May’s optimism reaches its limits.

On a very different note: Is it really advisable to be less confident in controversial moral beliefs (p. 128), given that these are often cases in which it is most important to stand up firmly for them? Consider “abnormal moral contexts” (Calhoun 1989), where controversy arises precisely because moral knowledge within a specialized community – obvious examples include trans and disabled communities – develops faster than it can be disseminated externally. Here, standing in solidarity with an oppressed group might require me to *defer* to their judgment even when I disagree (Kolers 2016). The fact is that when it comes to contested moral terrain, there is usually more at stake than just the accuracy of my beliefs, that is, whether I end up with a correct or incorrect moral judgment. What we believe in controversial cases is not merely an epistemic but also a *political* matter: of whom we trust, to whose testimony we are willing to defer, and whose interests we value.

This brings me to a wider critical pessimism about the way that moral cognition is typically understood. In the actual world, moral reasoning is never really about the socially abstracted, stick-figure characters that feature in experimental vignettes. This means that we should be careful about optimistic results obtained through generating moral intuitions about underspecified individuals. Studies of gender attribution, for instance, demonstrate that in the absence of substantial cues, people assume the “default” person to be male (Hamilton 1991; Merritt & Kok 1995). Similarly, some fans’ outrage over fictional characters being revealed as Black or gay suggest that their default is to imagine them as White and heterosexual (Hetter 2015; Holmes 2012). Such simplifying abstractions might be necessary for scientific investigation, but moral cognition in the wild always concerns full-blooded individuals replete with traits attributed on the basis of multiple social categories. It would be naive and dangerous, for example, to think that differences between liberals and conservatives come down only to different emphases on moral foundations (cf. Graham et al. 2009), rather than being deeply intertwined with racial resentment,

sexism, and reactionary White male rage. Yet social categories are rarely manipulated in studies of moral cognition, where “liberal” and “conservative” function as the primary contrast classes. May discusses only two (from the same publication), to which I will return shortly.

This disconnect between research on moral versus social cognition is, of course, not unique to May. But I worry that May underestimates the influence of social difference on moral cognition as a result of an overly narrow construal of moral judgment as explicit beliefs concerning the rightness and wrongness of acts. May circumscribes the threat of implicit biases, for example, by using evidence that they only weakly predict measured behaviors (p. 217). Elsewhere, he distinguishes between judgments about the wrongness of an act versus judgments ascribing blameworthiness (pp. 60–61) or negative moral traits to agents (p. 35), excluding the latter from his arguments. But why shouldn’t these count as genuine and important instances of moral cognition? Much moral thinking and acting proceeds through attunement to morally relevant properties other than rightness/wrongness. So I am not sure we can so cleanly separate out moral from social cognition, as May seems to do. (Discussing social psychological effects of group membership on moral cognition, he writes that “these flaws can largely be attributed to cognitive biases present in other domains, not to something particular about moral cognition itself” [p. 129].)

Indeed, I would argue that, in the real world, moral cognition *is* in some sense social cognition, and vice versa, because we live in a world so deeply and unjustly stratified by social difference. Our biggest moral quandaries today are not so much difficult exercises in moral reasoning as they are failures to unite masses of people who should be making common cause against systems of domination that benefit a tiny minority at their expense – but who instead remain divided by race, gender, class, and so on. In the two studies that did consider the effects of racial difference on consequentialist versus deontological reasoning, the results were significant: liberals were less willing to sacrifice a Black (vs. White) man in a trolley scenario, and conservatives were less willing to sacrifice innocent American (vs. Iraqi) civilians (Uhlmann et al. 2009). May and the authors interpret these findings as demonstrating the role of ideology in motivated moral reasoning. But they also highlight the pervasive (albeit unmeasured) entanglement of moral and social cognition. Social psychologists have long demonstrated a host of cognitive phenomena – fundamental attribution error (Ross 1977), defensive attribution (Shaver 1970), just world hypothesis (Lerner 1980), systems justification theory (Jost et al. 2004), to name a few – that underwrite people’s rationalizations of a deeply unjust status quo. Thus, May’s optimism might be dampened if we kept in view how these factors are always ineluctably in play within actual moral cognition.

To sum up, I think that there is reason for critical pessimism about how our moral-social capacities fare in a deeply unequal, highly segregated, hyper-partisan society. Under such conditions it is unsurprising that we continue to commit or allow moral atrocities against others, despite sharing the same, basically sound systems of moral cognition and motivation. I do not think that this critical pessimism should lead us to defeatism. But it should remind us, as others have argued (e.g., Vargas 2013a), that our moral agency is profoundly dependent on scaffolding from our social contexts. We cannot reason or act morally well in a badly ordered society – and none of us, moral psychologists included, should rest easy with this.

Author's Response

Defending optimistic rationalism: A reply to commentators

Joshua May 

Department of Philosophy, University of Alabama, Birmingham, AL 35294-1260.
joshmay@uab.edu
<https://www.joshdmay.com>

doi:10.1017/S0140525X19000967, e168

Abstract

In response, I elaborate on my conception of moral reasoning, as well as clarify the structure of debunking arguments and how my cautious optimism is only of the “glass half full” sort. I also explain how rationalism can capture insights purportedly only explained by sentimentalist and Humean views. The reply concludes by clarifying and admitting some limits of the book's scope.

R1. Introduction

Regard for Reason in the Moral Mind (May 2018) attempts to provide a more optimistic view of moral psychology in light of the science, partly by highlighting the centrality of reasoning. The book's topic is broad and its method multidisciplinary, so I am particularly grateful for having thoughtful commentaries from a wide range of philosophers and scientists. Each raises some important concerns or pushes the discussion further in significant ways.

My reply is divided up into three main themes that arise in the commentaries. Although the book and the précis begin with empirical questions about what influences us and only then asks about the normative status of these influences, my reply takes the reverse order. I start with my optimism about moral psychology and defend the idea that it is not in disrepair (even if the proverbial glass is at best only half full). Next, I discuss moral inference and address those critics who believe I fail to accord emotions or passions (conceived as distinct from reason) their special place in moral cognition or motivation. Finally, I concede some limitations and omissions in the book but explain the need for narrowing my focus for this particular project.

R2. A glass half full

Regard for Reason in the Moral Mind contrasts itself with an existing literature that is more pessimistic about ordinary moral thought and motivation. Early on when developing the book project, I saw the drawbacks to dividing the dialectical terrain between “optimists” and “pessimists” about moral psychology. As **Doris** rightly points out, many of my opponents would eschew the “pessimist” label. Now that may be because they have in mind some ordinary uses of the term that fall outside of my quasi-technical use (e.g., the optimism/pessimism dichotomy is not meant to concern “the possibility of progress in moral inquiry”). Nevertheless, I admit that some theorists are only borderline pessimists in my quasi-technical sense of the term. Part of the

problem is that “optimism” and “pessimism” can seem to label discrete categories, but they are better thought of as the ends of a spectrum. Unfortunately, no label does the job perfectly. Given the diversity of my opponents, I had to adopt terminology that would be useful for framing the project at the cost of being fully apt for all theories or theorists under discussion.

Terminological issues aside, some of the commentators charged me with being too optimistic about our moral minds (**D. Haas; Landy; McDonald, Yin, Weese, & Sinnott-Armstrong [McDonald et al.]; Rottman; Zheng**). I must admit that my optimism has been somewhat tempered by events that have unfolded around the world in the past few years. In November of 2016, I was putting the final touches on this book manuscript, amid one of the most significant moral and political events in decades. During the intervening years, people around the world, not just in the United States, have become tremendously incensed and polarized, with ghastly forms of hatred, disrespect, and violence oozing to the surface. Although my position was only ever characterized as cautious optimism, my tone in the book likely rings Panglossian, post-2016.

Now, I do not think we have enough evidence to confidently say how full the glass is exactly. But, if you will continue to indulge the metaphor, I will try in this section to defend the cliché that it is at least half full.

R2.1. Irrationality

For now, let us set aside particular kinds of bad influences (more on that later) and consider more general worries about irrationality in ethics. **Landy** admits, at least for the sake of argument, that “better reasoners are more likely to arrive at well-founded moral beliefs than are worse reasoners.” But he suggests that many of us are bad reasoners in general and thus lack well-founded moral beliefs. It is tempting to dismiss this as a radical form of skepticism about our general reasoning capacities, which I set aside as outside the scope of the book (see especially p. 22). But **Landy** astutely pitches his argument as one “for a moderate amount of pessimism,” which “cannot be dismissed as merely radical skepticism.”

My response is two-pronged. First, since pessimism comes in degrees and my optimism is only of the cautious “glass half full” sort, I am happy to concede a good deal of pessimism. However, second, let me say a few things to suggest that the pessimism about reasoning generally only goes so far.

Landy's pessimism is based primarily on evidence of our general poor performance on cognitive tests, particularly the Cognitive Reflection Test (CRT). There are multiple versions of the CRT now, but the one used most often by far involves what (in the United States at least) are called word problems in grade school math class. For example, one of the three questions on the original CRT is “If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?” (**Frederick 2005**). The mix of numbers, repetition, and counter-intuitive tricks in such word problems makes most people's eyes glaze over, and consequently they do not exert the cognitive control to carry out the right inference. The struggle with word problems, however, does not necessarily reflect irrationality generally or a general unwillingness to reason.

Consider another version of the CRT mentioned in the book (p. 71), which downplays the numbers and deals with more ordinary scenarios. For example, one question reads: “Emily's father has three daughters. The first two are named April and May.

What is the third daughter's name?" (Thomson & Oppenheimer 2016, p. 101). The intuitive answer is "June," but the right answer after a little reflection is "Emily." Online participants correctly answered on average a little more than two of the four questions, which is only slightly higher than the original CRT. Nevertheless, it does suggest that people reason better with more ordinary problems. And yet the data that **Landy** draws on regarding performance on the CRT rely almost exclusively on the original CRT with the mind-numbing number problems.

Another reason for optimism lies in a comparison with research on the famous Wason selection task. This problem, which originally involved two-sided cards, requires one to determine when a conditional statement is false, such as "If a card has an even number on its face, then it is red on the other side." Which card(s) should you turn over to determine whether the conditional is false? Few people are successful at this seemingly simple task. As it turns out, however, most people can solve it easily if the context is one we are more equipped to deal with, particularly when it involves conditional social rules, such as "If a man eats cassava root, then he must have a tattoo on his face" (Cosmides & Tooby 1992). Even children as young as three can navigate a Wason-style selection task if framed in terms of a norm, such as "If a mouse is squeaky, it must stay in the house" (Cummins 1996). So, even if we are no good at tackling certain mathematical word problems, we may be much better when reasoning about their social or normative counterparts.

Finally, the CRT tests for *reflective* inference, in which one must slow down and consciously apply the steps in a process of reasoning. However, this only represents one kind of reasoning. As I emphasize in the book (especially Ch. 3), much reasoning is unconscious, automatic, and intuitive. Perhaps Greene (2014) is right that the resolution of controversial moral problems requires engaging reflective "System 2" reasoning. But even Greene acknowledges that in everyday life we should by and large rely on intuitive moral reasoning. And we do have evidence that intuitive reasoning is often both sophisticated and reliable (see, e.g., Gigerenzer 2008; Nichols et al. 2016). Indeed, in ethics, moral judgments associated with more reflective reasoning are correlated with more anti-social and egoistic moral views (Kahane et al. 2015). So I caution against using (only) reflective reasoning as a test of good reasoning abilities, especially given that being adept at reflection and consciously articulating a justification often just helps one to better rationalize a bankrupt position (see, e.g., Kahan et al. 2012; Kunda 1990).

Indeed, reflective reasoning is sometimes good, sometimes bad. Often we are unable to accurately identify and articulate our reasons, which can lead to post hoc rationalization. **Clark & Winegard** are not themselves pessimists, but suggest that without the right emotional intuitions post hoc rationalization can make moral debate full of self-deception. However, this is not always the case. Initially inaccurate attempts to explain one's own choices can ultimately lead to insights and improve one's behavior (Summers 2017). For example, I might not fully understand why I stop eating meat, but that forces me to research factory farming and to be consistent in whatever analysis I come up with (compare similar points made by **Holyoak & Powell**).

In general, it is important to keep in mind that an inability to identify and articulate one's reasons does not imply a lack of reasoning or poor reasoning. Children seem to reason in sophisticated ways well before they are able to consciously identify and articulate their reasons (see, e.g., Gopnik 2012). Once when my daughter was 5 years old she asked, "Dad, did you pack my

toothpaste [flavored] jelly beans?" Trying to be dead-serious, I said "Well, no ... I ate them all. Sorry." She replied: "No, you didn't! You don't like them." The fact that she added an explanation was noteworthy. When she was younger, she would have called my bluff by just saying "No you didn't!" without offering a reason. The ability to articulate the reason came later.

In sum, poor performance on tricky math problems does not necessarily warrant pessimism about reasoning in general. Moreover, we have positive evidence that we are often even better at logical reasoning in moral contexts involving rules. Finally, reflective reasoning is only one form of inference, and it is not generally unreliable or reliable, for it can lead to both good and bad reasoning. Still, I take many of **Landy's** points which contribute to my optimism only going so far. In particular, it is true that laziness can prevent many of us from engaging in effortful reasoning which is important in the moral domain, where one must exert great effort to carefully and charitably consider the opposition.

R2.2. Rationalizing our biases

Even if our general reasoning capacities are fairly sound, one might worry about our starting points. In adolescence or adulthood, our independent reasoning capacities may be up against powerful biases we have absorbed from our cultures and institutions, from misogyny to materialism. Reason can promote consistency in ethics (Ch. 5; Kumar & May 2019), a point well-illustrated by **Holyoak & Powell**, and I appreciate their proposal that coherence-based reasoning helps explain the difficulties with debunking most moral cognition or motivation. But, a skeptic may retort: "Put garbage in and you will get garbage out." We have to consider the materials that mature moral reasoners begin with.

This is the sort of message one might take away from several of the commentaries. **Rottman**, for example, recognizes all of the evidence that children are like little scientists, unconsciously reasoning and experimenting to learn about the causal and social structure of the environment. However, he points out that as children we are, despite some bouts of altruism, "particularly egocentric" creatures who start with some psychological dispositions that are adaptive even if not moral, such as tribalism and prejudice. Moreover, children are primed to uncritically learn from their cultures and as a result can, as Rottman puts it, form "novel moral beliefs in irrational and undiscerning ways."

Rottman and I may not disagree greatly. He acknowledges that children often engage in unconscious reasoning that is quite sophisticated. Rottman only defends a "middle ground" that maintains a "healthy dose of pessimism," something I welcome, as well, especially because it can foster moral progress and combat complacency. Because children are not yet mature moral agents, I focused on adulthood. Still, I take the point that our developmental origins shape adult moral reasoning (see also **Carpendale & Wallbridge; Narvaez**), and often not for the better.

We may find evidence of this in **Zheng's** commentary. Her concern is that social cognition generally is so riddled with bias that most people are likely to use reason to justify morally reprehensible attitudes and actions, such as racism, sexism, and xenophobia, especially given our unjust social conditions. Many people come into adulthood with homophobic tendencies, for example, that were likely inculcated by their culture, society, and perhaps evolutionary history. For many people, forming enlightened attitudes about homosexuality requires great effort and many do not

succeed. Similarly, there are unfortunately powerful amounts of racism, xenophobia, and dogmatism that good reasoning has to work against. Moreover, Zheng rightly worries that the pervasiveness of motivated reasoning means we are likely to just rationalize many of the morally reprehensible attitudes we start with.

Like other calls for some pessimism, I concede it as far as it goes; and, again, it certainly goes farther than the tone of the book suggests. But it is important to remember that my optimism is only about the role of reason in moral psychology and our basic modes of moral thought and motivation. Of course, as **Zheng** makes clear, reason can justify terribly inaccurate ideologies, even ones that are woefully unjustified. However, whether our concern is moral judgment or social cognition generally, I do argue that many illicit forces, including implicit biases, are not as powerful as they are often thought to be (Chs. 4 and 9). Of course, as I say, little biases can add up to a powerful effect on society as a whole, just not so much on individual beliefs and desires. So I agree with Zheng that a badly ordered society impairs reasoning and acting well, but then moral reasoning is not rotten to the core but rather situated within unjust societies.

Second, I suspect that correcting structural injustices will require improvements in individual rational capacities, not mere feelings (conceived as distinct from reason). To combat misogyny, for example, we cannot just get men to feel more positive toward women. As Kate Manne (2017) has recently pointed out, sexists can still care sincerely for their mothers, sisters, friends, and coworkers, all while perpetuating patriarchal attitudes and social orders at home and beyond. A loving husband can still expect his spouse, daughters, and women politicians to be “attentive, loving subordinates” (Manne 2017, p. 49) and mete out punishment if they fail to live up to those expectations. The problem here is not mere feeling or affection but unconscious patriarchal ideologies and social norms. Now, such problems may have less to do with individual moral cognition than entrenched cultural norms and institutions. But individuals may bear some responsibility for the structural injustices, which one cannot combat without reasoning that the ideologies and structures are flawed (Madva 2016; Zheng 2018).

In sum, as a cautious optimist, I am a good deal pessimistic. The more one studies the science, the clearer it is that we need to be humbler in our controversial moral views and recognize our ability to rationalize our egocentrism and prejudice. I agree with **Zheng** that these are the most troubling and pressing sources of moral irrationality, but as she also emphasizes this is well known from history. As far as the psychological science, I remain doubtful that we can ground much pessimism in incidental disgust, evolutionary pressures, framing effects, automatic emotional heuristics, and the like. Perhaps now then is a good time to defend my optimism about those potentially bad influences that have received so much attention in moral psychology.

R2.3. *Bad influences*

My optimism about moral cognition and motivation is due primarily to the failure of wide-ranging attempts to debunk them. I framed the skeptical challenges in the simplest of terms to highlight their key empirical and normative premises:

1. **Empirical premise:** Moral cognition/motivation is mainly based on some factor.
2. **Normative premise:** That factor is morally irrelevant, extraneous, etc.

3. **Normative conclusion:** Moral cognition/motivation is improperly influenced (and so unwarranted, improper, etc.).

One drawback of this oversimplification is that it obscures an important issue, which **Demaree-Cotton** nicely draws out. It is not necessarily problematic that a belief or motivation has a bad *influence*, even if it is substantial. A pleasing smell or energetic background music may be irrelevant to my task of determining the appropriate punishment for a defendant, but those same factors could help focus my attention on morally relevant facts of the case. As I point out in the book (e.g., pp. 13, 28, 31, 48, 71), emotions in particular can help draw one’s attention toward (or away) from relevant information. Demaree-Cotton is quite right to note, however, that my frequent talk of “substantial influence” can suggest (wrongly) that debunkers need only point to an extraneous cause whether or not it also ignites a reliable process of forming one’s moral belief or motivation. The better phrase – which I do employ in the official statements of the Debunker’s and Defeater’s Dilemmas – is that the relevant judgment or motive is “mainly based” on the bad influence. That phrase at least suggests more clearly that the resulting judgment or motive is based on the bad influence and not also a good influence.

As **Demaree-Cotton** notes, if we emphasize this point, then it is even harder to draw skeptical conclusions from the data. Even if, for example, incidental disgust substantially influenced a wide range of moral beliefs, debunkers would still have to show that this emotion does not draw one’s attention to morally relevant information. Still, it seems reasonable to assume that many if not all bad influences – incidental disgust, genuine framing effects, implicit biases, and so on – do not generally beget good influences. I at least tended to grant my opponents this assumption for the sake of argument.

In the book, whenever there was a concern about nearby good influences, I typically addressed the issue in relation to the normative premise of the debunking argument, which essentially concerns whether the influence is actually bad or good. Consider, for example, my discussion of framing effects in chapter 4 (more on that soon). The order in which information is presented can seem like a bad influence on moral belief, but it is perfectly appropriate if it leads to rationally updating on new evidence (see Horne & Livengood 2017). Rather than saying this is a bad influence that generates a good one, I typically treated the seemingly problematic influence as not so bad after all.

Speaking of problematic framing effects, **McDonald et al.** take issue with my claim that meta-analyses suggest that such effects on moral belief are generally small. Part of their worry is that I rely primarily on only one meta-analysis by Demaree-Cotton (2016), which McDonald et al. believe is flawed. Although they do acknowledge that I also rely on another meta-analysis (Kühberger 1998), they object that it concerns framing effects “on risky choices rather than morality.” But many of the risky choices examined in that meta-analysis are morally relevant ones, such as tax evasion, contract negotiations, game-theoretic social dilemmas (e.g., public goods games), and public health policies such as the famous “Asian disease problem” (from Tversky & Kahneman 1981). Moreover, my reliance on Kühberger’s meta-analysis is important because it suggests that one of the most substantial framing effects (the disease study) is actually an outlier, not representative of similar decisions.

McDonald et al. also object that **Demaree-Cotton’s** meta-analysis is limited and that they are currently preparing a larger analysis of framing effects on moral judgment. That is very welcome

news, and I look forward to seeing the final results. Until then, we must work with the data available. What can we conclude so far?

McDonald et al. urge us to recognize that framing effects on moral judgment do not just arise for order of presentation and the like. They point to studies showing that moral judgments are influenced by videos, sleep deprivation, and social milieu. But these are often morally relevant factors. It is important to realize, as we saw previously, that a factor can seem morally relevant in the abstract but is not problematic when understood in context. Consider sleep deprivation and compare its effects on mathematical judgment. Suppose, as is plausible, that studies show that sleep deprivation dramatically affects one's solutions to math problems. There is no doubt that sleep deprivation tends to negatively affect mathematical judgment, but the question is whether this general factor of *amount of sleep* is a defective way of forming mathematical beliefs. Again, it may seem so in the abstract but amount of sleep – in particular, *sufficient* sleep – does plausibly ignite mathematically relevant cognitive processes. Indeed, it is a truism that quality and quantity of sleep directly impacts general cognitive functions, including those relevant to quantitative reasoning. Similarly, when properly described, many factors affect one's perception of morally relevant considerations, including videos, amount of sleep, and social milieu.

Another issue raised by **McDonald et al.** concerns **Demaree-Cotton's** finding that roughly 80% of moral judgments remain unchanged even when subjected to framing effects. The commentators object that 20% is large enough for skepticism to creep in. Even if the effects are small, they can add up for any single individual and ultimately sway some of their moral judgments for arbitrary reasons. The exact structure of this idea could take different forms.

On one reading, this is a version of what, in chapter 5, I call the “generalization worry” (p. 112), which can be found in the work of others, as well (see, e.g., Doris 2015; Rini 2016). I address the generalization worry in chapter 5 but also in chapter 9, where I discuss this move among those who attempt to debunk or “defeat” moral motivation. In both places, my response to the generalization worry is essentially that it fails to demonstrate evidence of unreliability (I take it that 80%, e.g., is reliable), which is necessary for a debunking argument.

One could retreat to merely raising the hypothetical *possibility* of unreliability. **McDonald et al.** do worry that “we as individuals often do not know whether we are in the unreliable 20% or the reliable 80%.” But such “skeptical hypothesis arguments” are notoriously weak (May 2013b). If swallowed, they lead to radical skepticism well beyond the moral domain. Some brilliant philosophers have defended such forms of argument and some have embraced the radical skepticism that follows. But few of my opponents aim to go that route. Indeed, if they did, the empirical evidence would be entirely irrelevant, for as Descartes showed us long ago one can spin out skeptical hypothesis arguments day and night while meditating in the armchair. At any rate, my concern is whether moral cognition is unreliable, not whether unreliability is merely possible.

McDonald et al. might be offering instead an argument from disagreement. They write that “one cannot know that one is a reliable moral judge, whereas others are not, unless one has some reason to believe that one is special in some relevant way.” I entirely agree, provided we do restrict the skeptical conclusion to whether “moral judgments *in controversial cases* can be justified” (my emphasis), as I do in chapter 5 when embracing a *selective* debunking argument from *peer* disagreement (more on that in the next section).

It is quite important to distinguish the different forms of argument here, which point to either: the mere possibility of error, likelihood of error, or the prospect of error from peer disagreement. An analogy may help. Consider my various beliefs about the capitals of countries, such as the belief that Lima is the capital of Peru. Of course, it is merely possible that I am wrong, but that is no reason to suspend judgment. Now imagine I learn that 20% of my geographical beliefs are formed in a faulty way, such that if those 20% are accurate it is a kind of fluke. Should I suspend judgment about Peru's capital? I do not think so. Suspension would be required if I learned that 80% of my geographical beliefs are suspect, but not 20%. Matters change greatly, however, if we introduce epistemic peers. Suppose instead that I learn that 20% of my *peers* disagree with *that particular belief* about Peru. If these are well-informed people who I have no reason to believe are less likely to be right, then suspension of judgment is appropriate.

Ultimately, **McDonald et al.** may disagree with me less than it appears. They do not conclude that “moral judgments are not justified” but only that “they are not justified by intuition alone.” My concern is only to examine empirical evidence that bears on the question of whether many of our moral beliefs are justified, whether by intuition or inference. As I emphasize in chapter 3, moral judgment involves a great deal of (unconscious) inference, even if the resulting moral judgments seem (consciously) to the individual to be automatic gut feelings.

R2.4. Disagreement about disagreement

In chapter 5, I argued that moral disagreement among epistemic peers should give most of us pause. Many *controversial* moral beliefs, of the average person at least, do not amount to knowledge and are less justified than they like to admit. This is one of the few skeptical challenges that I acknowledge is powerful, albeit limited in scope.

I focus my discussion on disagreement within a culture, partly because that is more tractable but also because I did not want to assume that moral truths are universal across cultures. **D. Haas** accepts my narrowed focus on disagreements among liberals and conservatives but argues that moral foundations theory cannot help identify foundational moral beliefs. One worry is that moral foundations merely express general values (e.g., loyalty, fairness), not beliefs in foundational moral statements, which are the stuff of disagreement and propositional knowledge (or lack thereof). It is true that the standard 30-item Moral Foundations Questionnaire (see Graham et al. 2009, p. 1032) asks only, “When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking?” The options include only general moral values, such as “Whether or not someone acted unfairly” and “Whether or not someone showed a lack of loyalty” (Graham et al. 2009, Appendix A).

However, Haidt and his collaborators (Graham et al. 2009, Study 2, Appendix B) have explicitly asked participants for their attitudes toward foundational moral propositions, such as:

“It can never be right to kill a human being.”
 “People should not do things that are revolting to others, even if no one is harmed.”

Yet the researchers found similar results when asking participants to indicate their agreement with such moral statements (Graham et al. 2009, pp. 1034–35). This suggests that the many studies

employing foundational values are also tapping into foundational moral beliefs.

D. Haas also objects to my claim that liberals and conservatives share many foundational moral values/beliefs. Even if all moral foundations are recognized by most liberals and conservatives, differences remain among the groups' relative weights of the values. If that is right, then there is more foundational moral disagreement than I admit, which would make the skeptical challenge affect many more moral beliefs. However, as I say in the book, differences in foundation weighting should not be overblown. Most people are not on the ends of the ideological spectrum, and yet those toward the middle express little variation in the foundational value.

In response, **D. Haas** argues that, although many liberals do value loyalty, authority, and sanctity, they do not "consciously acknowledge this and ... typically will disown using these foundations." That is true (for summary, see Graham et al. 2013, p. 96), but I suggest we take liberals' intuitive responses as more representative of their fundamental moral values/beliefs. Compare implicit versus explicit racial biases. We should discount one's explicit disavowal of racial biases when one nevertheless exhibits them implicitly, especially given the motivation to appear unbiased. Similarly, moderate liberals want to appear clearly liberal, so they will be motivated to express greater commitment to foundations valued further on the left. (Perhaps we do not observe this effect as much in moderate conservatives because, although extreme conservatives value all foundations more equally, a motivation to appear more conservative need not involve discounting any particular foundations.) At any rate, often we can better determine a person's fundamental attitudes by investigating one's intuitive responses, because one's implicit reactions can be a better guide to their true feelings (see Cameron et al. 2017).

In sum, the research on moral foundations does seem to provide evidence of people's foundational moral beliefs. And the differences between most (moderate) liberals and conservatives, although real, should not be overstated, which limits the scope of the skeptical challenge. Although I agree with **D. Haas** that the Moral Foundations Questionnaire is not perfect, I believe it is a good start – better at least than only armchair speculation about people's foundational moral beliefs.

McAuliffe & McCullough instead take issue with my claim that the average person should regard many of their opponents as epistemic peers. They object that social-scientific research can show that, in the aggregate at least, some opponents tend to be *epistemic inferiors* – people who are less likely to have morally reasonable or correct views. If that is right, I may be conceding too much to skeptics, which would make me overly pessimistic about controversial moral judgments.

The research **McAuliffe & McCullough** cite is indeed useful but limited. Many of the studies concern reflection, which earlier I noted is only one kind of reasoning that does not always lead to morally sound views (see the response to **Landy** in sect. R2.1). More reflective does not always mean more rational. Other research cited by **McAuliffe & McCullough** concerns general intelligence. But again I would caution that some measures of intelligence, such as mathematical reasoning and "book smarts," do not always yield a virtuous person. Moreover, as **McAuliffe & McCullough** make clear, there are many aspects of morality and rationality, but that makes it difficult to know whether an opponent lacks the rationality relevant to the controversial issue at hand. A degree in medicine or the ability to quickly solve a Rubik's cube will not necessarily make one an expert on the ethics of international trade agreements.

Another concern of **McAuliffe & McCullough's** is that we should be able to identify epistemic inferiors if there is to be moral progress. They mention that abolitionists provided cogent arguments that were met with self-serving and empirically spurious counter-arguments in defense of slavery. Shouldn't we be able to identify moral trailblazers and disregard defenders of the status quo as epistemic inferiors?

I believe some of us can. As I say in chapter 5, we can successfully debunk particular kinds of moral beliefs, just not large swathes of moral cognition. But it is not easy to identify epistemic inferiors – whether individuals or groups – and doing so definitely is not just a matter of "counting the number of learned people who hold a certain moral point of view." Indeed, I concur with **McAuliffe & McCullough** that one must "examine the reasoning and evidence that each side has." When I say that most people lack some moral knowledge about controversial moral issues, I am assuming that most people form their controversial moral beliefs without engaging in an honest and thorough examination of the opposition. That is part of the path to moral knowledge amid controversy. The denial of knowledge at one time does not preclude gaining it at another (cf. **McDonald et al.** on acquiring "independent confirmation"). My position here is not far from Greene's (2013), who counsels us to be cautious and reflective when it comes to forming moral beliefs about controversial issues. The difference is that I do not believe the reflective examination of moral arguments relies preferentially on utilitarian considerations or otherwise requires abandoning our automatic moral intuitions.

Zheng raises a related issue about how disagreement affects moral progress. She helpfully points out that it is often precisely amid moral controversies that we should "stand up firmly" in support of our beliefs. **Zheng** also points out that we should often defer to marginalized groups, for they are epistemic superiors regarding their own oppression. This is an important point, but I can concede it as far as it goes. My claim is only that many ordinary people lack knowledge about controversial moral issues on which reasonable people can disagree. But knowledge is not required for all actions, such as voting to abolish slavery, listening to the concerns of those in a minority group, and protesting the inhumane treatment of oneself or one's group. Indeed, greater intellectual humility can yield deference and open-mindedness.

Of course, if such open-mindedness is required of those in the oppressed group too, then should they no longer confidently assert their grievances? Not necessarily. First, even in the absence of moral knowledge I admit a good deal of moral justification, which may be sufficient for action (cf. **Locke** 2015). Even if we should only act on what we know (**Hawthorne & Stanley** 2008), marginalized people do know what their experiences are. Although reasonable people might disagree, say, about whether a woman with white parents should identify as black, members of the black community know their concerns and should voice them (and others should listen). Moreover, a lack of knowledge about controversial issues need not paralyze one politically. One can, for example, exercise one's right to vote. In my view, voting is precisely about expressing one's concerns by backing the candidates or ballot initiatives one favors, regardless of whether one should be confident about their preferences.

In sum, my claim is not that most people lack any justification for their controversial moral beliefs; it is that they cannot claim to know. That is compatible with a good deal of justification for belief and for relevant action. My position is just incompatible with the kind of excessive confidence and intellectual arrogance one often sees in individuals.

R3. Reasoning with passion

Some of the commentators were concerned with my optimism about the role of reason in moral psychology. In this section, I respond to these more sentimentalist and Humean concerns.

R3.1. The primacy of emotion?

A common complaint is that I only target extreme, rather than sophisticated, forms of sentimentalism. On more modest versions (that at least still make empirical claims), even if emotions do not cause every moral judgment, they ultimately explain the moral judgments most people make. **Kauppinen** argues that there is a “striking parallel” between our “adaptive emotional tendencies and widespread patterns of moral judgment.” **Kurth** argues that emotions are to moral judgments as color experiences are to judgments of color. The mere feeling associated with shame, for example, “shapes and constrains” our concept of shame and its associated thoughts or content. Similarly, **Clark & Winegard** maintain that ultimately “we must build our moral judgments and arguments from the raw materials of our moral intuitions,” which they conceive of as “pre-rational preferences.”

In chapter 2, I do argue against such forms of sentimentalism by pointing to evidence that our emotional responses are typically consequences of our moral judgments, not vice versa. **Kurth** responds that the “experiments focused on just one emotion (disgust),” although for the record I also mentioned studies on compassion (p. 38). **Kurth** does question the value of one of the studies I draw on (Yang et al. 2013). The worry is that the experimenters only measured *judgments* of disgust, not *feelings* of disgust. But presumably the former typically relies on the latter. One typically judges something to be disgusting by feeling disgust toward it, or at least being disposed to feel that way – thus expressing a sentiment toward it (Prinz 2007).

Kurth and **Kauppinen** also raise a challenge for rationalism. As **Kauppinen** puts it, what a coincidence if “reason just happens to tell us to disapprove of the very things we in any case tend to feel negatively about.” Sophisticated sentimentalism nicely explains this striking parallel, for it is no surprise that we tend to praise acts that promote what we antecedent care deeply about and condemn acts that do not.

However, there is no coincidence if moral reasoning takes as its materials the things we antecedently care about. Imagine, for example, that utilitarianism is true and we can grasp the truth of the principle of utility by reason alone (Singer 2005). Then moral reasoning will take as input what people care about, namely what makes them happy and sad. The same goes for other prominent moral theories. On Kantian rationalism, for example, we take our antecedent goals, plans, or maxims and test them against morality’s demand that they be universalizable (Korsgaard 1996a). On contractualism, we similarly test existing plans to see whether they can be reasonably rejected (Scanlon 1998). On such forms of rationalism, it is no mere coincidence that we happen to tend to judge wrong what we antecedently dislike. Genuine moral judgment makes heavy use of these starting materials in reasoning to moral conclusions.

Indeed, rationalism nicely explains why we do not *always* judge to be moral that which promotes our given desires, preferences, or goals – because only some of them are rational. Of course, rationalists must then maintain that our preferences are not themselves moral judgments. **Clark & Winegard** seem to think otherwise: “Moral judgments, by their very nature, must

be grounded in intuitions about what is good and what is bad.” But judgments of good and bad are not sufficient for moral judgment. As Nichols (2004, p. 15) has pointed out, we do not prefer natural disasters, and we think they are bad, but we do not judge them to be immoral.

R3.2. The structure of moral reasoning

Setting aside my critique of sentimentalism, some commentators raised challenges for my account of moral reasoning. That account was certainly under-described. More specific models of reasoning generally might help to illuminate moral cognition, such as the erotetic theory discussed by **Alfano** (based on work by Philipp Koralus). (Indeed, I suspect the erotetic theory provides a useful analysis of what it is to form a belief “on the basis of” another belief.) However, the already broad scope of the book demands certain limits on which issues I can address (more on that in sect. R4).

Take **Kauppinen**’s example of a simple moral argument: “Clinton lied. Lying is wrong. So Clinton did something wrong.” How, **Kauppinen** asks, do we reason to the moral principle that lying is wrong? In the book, I point to more fundamental moral principles, such as the principle of agential involvement (p. 69), but **Marshall** rightly argues that this cannot be the final foundation given its “all else being equal” clause. **Marshall** predicts that I would not want to posit a single fundamental moral principle, such as Kant’s categorical imperative or the principle of utility, that is devoid of exceptions. I am actually open to that possibility. But my aim has only been to argue for rule-based moral reasoning that need not rely on emotions (conceived as distinct from such reasoning). I try to remain neutral on what the (rationalist-friendly) foundations of moral reasoning are, largely because I do not think we have enough empirical evidence to settle that psychological question.

With that said, let me say a bit more about what I take to be the options for rationalism. Now, the rationalist thesis at issue here is merely psychological, not normative, but moral epistemology provides various prescriptive theories that can also serve as psychological models (for review, see Zimmerman 2010). On infinitist theories, moral reasoning can just go on indefinitely, with the better justified moral views being the ones that are supported by more and better reasons. On coherentist theories, moral reasoning involves developing an internally consistent set of beliefs, perhaps through the venerable method of reflective equilibrium (compare **Holyoak & Powell**). On foundationalist theories, moral reasoning bottoms out in non-inferential states, such as moral perceptions or self-evident moral truths.

Which of these views most accurately represents human moral reasoning? I am not sure; we currently lack enough evidence. But all of the models are rationalist-friendly. Now, each model must ultimately address a foundational question: What exactly makes a belief distinctively moral? (Infinitists, for example, still need to determine when one has a *moral* reason.) That is another question I will not attempt to fully answer, because it is not necessary for carving out options for empirically sound rationalism. The rationalist need only deny that a distinctively moral judgment requires emotions (conceived as distinct from reason), which is precisely the aim of chapter 2. Once that is in place, the rationalist can adopt a coherentist, foundationalist, or infinitist model of belief formation in ethics (or even some combination of these).

Of course, there are sentimentalist versions of these models. For example, drawing on the analogy with color perception,

sentimentalists might argue that the non-inferential foundations of moral reasoning are moral perceptions that are necessarily emotional in nature (compare **Clark & Winegard**; **Kauppinen**; **Kurth**). But my challenge to these and other sentimentalist views is to explain the sorts of issues taken up in the previous section, such as: (a) Why don't incidental emotions generally seem to change moral judgments (even if they need not always do so according to sentimentalists)? (b) Why do emotions often seem to be the consequences of, or dependent on, our prior moral judgments?

J. Haas raises a rather different concern about my view of moral reasoning. She argues that, instead of rule-based inference, we can explain moral judgment in terms of domain-general valuation mechanisms familiar from work on reinforcement learning. These mechanisms are certainly important and are beginning to contribute insights into moral psychology. I omitted discussion of them in the book partly for reasons of space, so I appreciate the opportunity here to say why at this stage I do not find them as useful for understanding moral judgment.

Consider how **J. Haas** proposes to account for the relevance of harmful outcomes to moral cognition. Such outcomes can be represented by multiple domain-general valuation mechanisms, but just consider the model-based mechanism which, as she defines it, "explicitly represents possible choices and determines the sequence of actions that maximizes value." However, such a mechanism just models which outcomes promote what the individual takes to be of value, and even consequentialists agree that representations of value do not suffice for moral judgment. Consider two choices that generate a morally valuable state of affairs because they promote happiness. On a utilitarian view, only the one that brings about the most happiness will be the *right* thing to do – a *deontic*, not merely evaluative, status (cf. **Mikhail 2011**, p. 85). A model-based mechanism can only account for moral judgment if it factors in that deontic step. Yet I have suggested that assigning the deontic status to an act involves categorization through the application of a moral rule or rules (the specific content of which I remain neutral on). We probably do not have enough evidence to settle this dispute, but until then I believe **J. Haas's** proposal is incomplete.

Let me say briefly why my bet is on a rule-based explanation. Domain-general valuation mechanisms may often appear to appropriately model moral cognition because, as **Kauppinen** and **Kurth** point out, there is that striking parallel (though not equivalence) between the things we judge right/wrong and the things we like/dislike, value/disvalue. Key to moral judgment, however, is not only its deontic status, but also its applicability to third-parties that have no relation to one's own personal choices. For example, I make a paradigm moral judgment when I read about atrocities abroad (or in fiction) and think "That's just wrong" – despite there being no connection to a choice I make to achieve some personal goal. I take it that, although many animals have evaluative attitudes, they lack this sort of core moral judgment. Yet the human brain shares domain-general valuation mechanisms with many other animal brains. Indeed, much of the work on reinforcement learning involves experiments with rats and monkeys in order to model the non-moral decisions they make to personally acquire what they value, such as food. For at least this reason, I suspect the relevant models of moral judgment will always be incomplete. The relevant mechanisms are developed to explain personal decisions underwritten by attitudes with only evaluative, not deontic, content.

Of course, we may eventually find that domain-general valuation mechanisms provide a complete explanation of core moral judgment. That is not necessarily a problem for rationalism anyhow. Even if the valuation mechanisms all rely ultimately on valenced flashes of affect (cf. **Railton 2017**), they do not conflict with the principal rationalist idea that moral cognition is continuous with other forms of cognition and reasoning, that there is no special moral module whose operations are fundamentally different from reasoning generally (cf. Ch. 1, p. 11).

R3.3. Slaves of the passions?

Reasoning may generate new moral beliefs, but can it generate new motives or desires? As a Humean about motivation, **Sinhababu** is unconvinced; reasoning, he contends, only tells us how to satisfy our antecedent desires. One of his concerns is that I place too much weight on our ordinary explanations of action that do not appear to always posit antecedent desires, such as: *Yongming chose salad over pizza because he thinks the carbs will make him too sleepy on the long drive home*. Such explanations occur in everyday communication, which often omits the obvious, including background desires. Surely the explanation of Yongming's temperance, for example, tacitly assumes a desire to stay alert while driving.

Fair enough, but my reliance on ordinary explanations was merely meant to do dialectical duty. My aim was to demonstrate how anti-Humean explanations can work and to shift the burden of proof onto Humeans who must always posit an antecedent desire. The examples I draw on also help to illustrate how the Humean explanation is not any simpler, for it requires positing that extra antecedent desire.

Another key contention is precisely such claims about parsimony. **Sinhababu** maintains that the Humean theory is simpler because it posits only one causal relation for motivation, one in which *desires cause beliefs*, never the other way around. In the book, I suggest that the anti-Humean theory does not posit an additional causal process, because we should carve up the causal relation more broadly – whether the cause is a desire or a belief, it is *mental causation* among propositional attitudes. For comparison, I give the example of *X baked Y*, which remains the same process or relation, even when the "relata" (*X* and *Y*) vary. **Sinhababu** replies that we should individuate a process by its relata, at least when they are "essential to characterizing the processes." But why should we treat desires or beliefs as essential to the causal processes in this debate? **Sinhababu** answers that otherwise Humeans and anti-Humeans cannot have a dispute in the first place. By comparison, he suggests that we should treat reasoning and telepathy as different processes in order to make sense of disputes about the existence of telepathy.

However, to recognize the differences between our two theories of human motivation, we do not need to treat the different causes of desires as distinct processes. We need only recognize their different implications for ethics (e.g., whether beliefs can ever motivate). To illustrate, go back to the baking example. Suppose we have two different theories. One says both humans and robots bake, whereas the other says only humans bake. Imagine further there is an ethical upshot: if robots also bake, then more humans will soon be out of a job. This empirical dispute remains, even if we assume baking is baking whether it is done by a human or a robot. Indeed, it would be odd to suggest that in order to have a dispute here we would have to posit two

different types of processes: human-baking and robot-baking. (Such a move reminds me of those opposed to marriage equality who insist on distinguishing marriage from civil unions because of the *relata* in the relationship.)

In sum, I concede that these issues are difficult to settle, and I do not pretend to have proven definitively that we are not ultimately “slaves of the passions.” In chapter 8, I am largely on defense, aiming to show that when it comes to theories of human motivation, the Humean theory is not the only empirically sound game in town.

R4. Scope

In closing, let me clarify and admit some limits of the book’s scope.

R4.1. *The moral domain*

Despite being concerned with distinctively moral judgment, I never provide a complete analysis of morality itself. As usual, however, if possible I avoid taking a stance on such controversial issues. My account is quite compatible with a number of specific views about the content and contours of moral norms – e.g., whether they concern mutual recognition (**Carpendale & Wallbridge**), the Kantian construction of one’s practical identity (**Roberts**), or universalization of one’s maxims (**Marshall**). On all of these specific proposals, moral judgment can be fundamentally a matter of reasoning.

Other commentators worry that such omissions have dire consequences. **Chater, Zeitoun, & Melkonyan (Chater et al.)** rightly stress that moral reasoners are “not lone and omnipotent decision makers” but embedded in social groups (cf. also **Zheng**). Yet most of the research I draw on measures individual moral judgments about dilemmas in which all relevant information is purportedly present. Moreover, I focus on mature moral thought and action among adults, without delving greatly into moral development and cross-cultural variation (**Carpendale & Wallbridge; Narvaez; Rottman**). As Narvaez puts it, “a wider examination of human behavior across time and societies is needed when discussing human moral potential.”

To be fair, I do discuss children’s development of altruistic motivation to help others (Ch. 6). I also discuss moral knowledge among groups, particularly the masses (Ch. 5), as well as the social nature of moral cognition that lies in our evolutionary past (Ch. 4). Moreover, my position is supported by some appeal to cross-cultural research (see, e.g., Ch. 3). Although these discussions are all too brief, my aims were largely to demonstrate that optimistic rationalism is empirically defensible. To do that, I did not have to cover all of the scientific evidence or show that the picture I sketch perfectly describes all human beings and all the ways they interact. Indeed, I deliberately avoided committing to such claims about universality or innateness (see, e.g., Ch. 3, pp. 77–78).

R4.2. *Moral truth and objectivity*

Gibson worries about how my optimism is restricted to merely justified belief and appropriate moral motivation. A truly anti-skeptical defense should “provide a picture according to which thought and action can be meaningfully connected to the normative.” Whether a moral judgment amounts to knowledge, for example, depends not only on whether it is justified, but also

whether it is *true*. An ancient Greek may be justified in believing that slavery is ethical, all while being ultimately motivated to uphold the practice for morally relevant (even if empirically misguided) reasons, such as justice and fairness. If that is the status of the moral mind, how can we be optimistic?

It certainly is part of a complete anti-skeptical project to explain how we are in touch with moral reality (cf. **Marshall 2018**). But it is possible for justification to have some connection to reality. If we adopt a reliabilist picture, one’s moral beliefs are only justified if they are reliably accurate. I remained neutral on such disputes about the nature of justification, but this shows how justification and reality are not necessarily separable. And a similar approach can be taken to virtuous motivation.

In any event, I set aside such metaphysical issues in the book because most of the “empirical pessimists” I engage with do. Evolutionary debunkers, for example, often argue that, even if our moral beliefs are true, they are only accidentally true and so unjustified (e.g., **Joyce 2006**). Similarly, sentimentalists are “pessimists” about reason at least, but they are not defending a disconnect between moral judgment and reality. Quite the opposite: They believe emotions are precisely what put us in touch with such reality, even if that reality is something we ultimately construct (e.g., **Prinz 2007**). In fact, let me emphasize that (contra **Carpendale & Wallbridge**) I neither assume nor endorse cultural relativism or any theory which takes moral truths to be subjective. My aim was to *remain neutral* about such issues in moral metaphysics because I did not need to take a stance on them to address my topics of justification and proper motivation.

R4.3. *On the armchair*

The book focuses heavily on empirical evidence, yet some important philosophical arguments in moral psychology are decidedly non-empirical. **Kauppinen** even contends that “what defines the various sentimentalist views are the conclusions of the a priori arguments.” However, my aim of course is only to address versions of sentimentalism that make empirical claims about human psychology, not any theory worthy of the label “sentimentalism.”

In the end, the book is ultimately “on the pessimist’s terms,” as **Roberts** nicely puts it, which are empirical. Rationalists too, though, rely sometimes extensively or exclusively on non-empirical premises to defend their views. Roberts points out that some Kantians, such as Christine Korsgaard, would insist that there is a “deep, perspectival divide between psychology and ethics.” We cannot settle the issue here, but I am of course dubious of philosophical views that make assumptions about human psychology – even just its possibilities and limits – and yet take empirical results to be irrelevant. Roberts’s comments also make clear that optimistic rationalism is not a specifically Kantian view, let alone a version that posits an unbridgeable gap between ethics and empirical psychology. I hope the book, as well as the insightful commentaries in this journal issue show how to at least start building that bridge while respecting the philosophical terrain.

References

[The letters “a” and “r” before author’s initials stand for target article and response references, respectively]

Aglioti A., Goodale M. & Sousa J. (1995) Size contrast illusions deceive the eye but not the hand. *Current Biology* 5:679–85. [CK]

- Aharoni E., Sinnott-Armstrong W. & Kiehl K. A. (2012) Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology* **121**(2):484–97. [aJM, WHBM]
- Alfano M. (2013) *Character as moral fiction*. Cambridge University Press. [aJM]
- Alfano M. (2016) *Moral psychology: An introduction*. Polity. [MA]
- Alfano M. (2017) Twenty-first century perspectivism: The role of emotions in scientific inquiry. *Studi di Estetica* **7**(1): 65–79. [MA]
- Allison H. E. (2004) *Kant's transcendental idealism*. Yale University Press. [AJR]
- Alter A. (2013) *Drunk tank pink: And other unexpected forces that shape how we think, feel, and behave*. Penguin. [JR]
- Aquino K. & Reed A. (2002) The self-importance of moral identity. *Journal of Personality and Social Psychology* **83**(6):1423–40. [aJM]
- Arain M., Haque M., Johal L., Mathur P., Nel W., Rais A., Sandhu R. & Sharma S. (2013) Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment* **9**:449–61. [DN]
- Ariely D. (2012) *The honest truth about dishonesty*. HarperCollins. [aJM]
- Arpaly N. (2003) *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press. [QHG, aJM]
- Arpaly N. & Schroeder M. (1998) Praise, blame, and the whole self. *Philosophical Studies* **93**:161–88. [QHG]
- Arpaly N. & Schroeder T. (2014) *In praise of desire*. Oxford University Press. [aJM]
- Audi R. (2013) *Moral perception*. Princeton University Press. [AK]
- Avnur Y. & Scott-Kakures D. (2015) How irrelevant influences bias belief. *Philosophical Perspectives* **29**:7–39. [JD-C]
- Ayars A. (2016) Can model-free reinforcement learning explain deontological moral judgments? *Cognition* **150**:232–42. [JH]
- Bargh J. (2017) *Before you know it: The unconscious reasons we do what we do*. Touchstone. [JR]
- Baron R. A. (1997) The sweet smell of ... helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. *Personality and Social Psychology Bulletin* **23**(5):498–503. [aJM]
- Barrett H. C., Bolyanatz A., Crittenden A. N., Fessler D. M. T., Fitzpatrick S., Gurven M., Henrich J., Kanovsky M., Kushnick G., Pisor A., Scelza B. A., Stich S., von Rueden C., Zhao W. & Laurence S. (2016) Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences* **113**(17):4688–93. [aJM]
- Batson C. D. (2011) *Altruism in humans*. Oxford University Press. [aJM]
- Batson C. D. (2016) *What's wrong with morality?* Oxford University Press. [aJM]
- Batson C. D., Kobrynowicz D., Dinnerstein J. L., Kampf H. C. & Wilson A. D. (1997) In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology* **72**(6):1335–48. [aJM]
- Batson C. D., Thompson E. R. & Chen H. (2002) Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology* **83**(2): 330–39. [aJM]
- Battaglia P. W., Hamrick J. B. & Tenenbaum J. B. (2013) Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, USA* **110**(45):18327–332. Available at: <https://doi.org/10.1073/pnas.1306572110>. [KJH]
- Bekkers R. & Wiepking P. (2011) Who gives? A literature review of predictors of charitable giving, part one: Religion, education, age and socialisation. *Voluntary Sector Review* **2**(3):337–65. [WHBM]
- Berns G. S., Bell E., Capra C. M., Prietula M. J., Moore S., Anderson B., Ginges J. & Atran S. (2012) The price of your soul: Neural evidence for the non-utilitarian representation of sacred values. *Philosophical Transactions of the Royal Society B* **367**(1589):754–62. [JH]
- Bertrand M. & Mullainathan S. (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* **94**(4):991–1013. [aJM]
- Betancourt H. (1990) An attribution-empathy model of helping behavior. *Personality and Social Psychology Bulletin* **16**(3):573–91. [aJM]
- Blake P. R., McAuliffe K. & Warneken F. (2014) The developmental origins of fairness: The knowledge-behavior gap. *Trends in Cognitive Sciences* **18**(11):559–61. [JR]
- Blanken I., van de Ven N. & Zeelenberg M. (2015) A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin* **41**(4):540–58. [aJM]
- Bloom P. (2013) *Just babies: The origins of good and evil*. Crown. [JH, JR]
- Bloom P. (2016) *Against empathy: The case for rational compassion*. Ecco. [aJM]
- Boehm C. (2012) *Moral origins. The evolution of virtue, altruism, and shame*. Basic Books. [AK]
- Boghossian P. (2014) What is inference? *Philosophical studies* **169**(1):1–18. [AK]
- Bolhuis J. J., Brown G. R., Richardson R. C. & Laland K. N. (2011) Darwin in mind: New opportunities for evolutionary psychology. *PLoS Biology* **9**(7):e1001109. Available at: <http://doi.org/10.1371/journal.pbio.1001109>. [KJH]
- Brady M. (2013) *Emotional insight: The epistemic role of emotional experience*. Oxford University Press. [MA]
- Brandt M. J. & Crawford J. T. (2016) Answering unresolved questions about the relationship between cognitive ability and prejudice. *Social Psychological and Personality Science* **7**(8): 884–92. [WHBM]
- Broome J. (2013) *Rationality through reasoning*. Oxford University Press. [AK]
- Brosnan S. F., Schiff H. C. & de Waal F. B. M. (2005) Tolerance for inequity may increase with social closeness in chimpanzees. *Proceedings of the Royal Society B: Biological Sciences* **272**:253–58. Available at: <http://doi.org/10.1098/rspb.2004.2947>. [KJH]
- Cajete G. (2000) *Native science: Natural laws of interdependence*. Clear Light. [DN]
- Calhoun C. (1989) Responsibility and reproach. *Ethics* **99**(2):389–406. [RZ]
- Cameron C. D., Payne B. K. & Doris J. M. (2013) Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology* **49**(4):719–25. [JMD]
- Cameron C. D., Payne B. K., Sinnott-Armstrong W., Scheffer J. A. & Inzlicht M. (2017) Implicit moral evaluations: A multinomial modeling approach. *Cognition* **158**:224–41. [rJM]
- Campbell R. & Kumar V. (2012) Moral reasoning on the ground. *Ethics* **122**(2): 273–312. [aJM]
- Campitelli G. & Gerrans P. (2014) Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory and Cognition* **42**:434–47. [JFL]
- Carlson M., Charlin V. & Miller N. (1988) Positive mood and helping behavior: A test of six hypotheses. *Journal of Personality and Social Psychology* **55**(2):211–29. [aJM]
- Carpendale J. I. M. (2000) Kohlberg and Piaget on stages and moral reasoning. *Developmental Review* **20**:181–205. [JIMC]
- Carpendale J. I. M. (2009) Piaget's theory of moral development. In: *The Cambridge companion to Piaget*, ed. U. Müller, J. I. M. Carpendale & L. Smith, pp. 270–86. Cambridge University Press. [JIMC]
- Carpendale J. I. M. (2018) Communication as the coordination of activity: The implications of philosophical preconceptions for theories of the development of communication. In: *Advancing developmental science: Philosophy, theory, and method*, ed. A. S. Dick & U. Müller, pp. 145–56. Routledge. [JIMC]
- Carpendale J. I. M., Hammond S. I. & Atwood S. (2013) A relational developmental systems approach to moral development. In: *Embodiment and epigenesis: Theoretical and methodological issues in understanding the role of biology within the relational developmental system: Advances in child development and behavior*, vol. 45, ed. R. M. Lerner & J. B. Benson, pp. 105–33. Academic Press. [JIMC]
- Carpendale J. I. M. & Krebs D. L. (1995) Variations in level of moral judgment as a function of type of dilemma and moral choice. *Journal of Personality* **63**:289–313. [JIMC]
- Carpendale J. I. M., Sokol B. & Müller U. (2010) Is a neuroscience of morality possible? In: *Developmental social cognitive neuroscience*, ed. P. Zelazo, M. Chandler & E. Crone, pp. 289–311. Psychology Press. [JIMC]
- Cheng P. W. (1997) From covariation to causation: A causal power theory. *Psychological Review* **104**:367–405. [KJH]
- Chockler H. & Halpern J. Y. (2004) Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research* **22**: 93–115. [NC]
- Chudek M. & Henrich J. (2011) Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences* **15**(5):218–26. [JR]
- Cialdini R. B., Brown S. L., Lewis B. P., Luce C. & Neuberg S. L. (1997) Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology* **73**(3):481–94. [aJM]
- Clark C. J., Baumeister R. F. & Ditto P. H. (2017) Making punishment palatable: Belief in free will alleviates punitive distress. *Consciousness and Cognition* **51**:193–211. [CJC]
- Clark C. J., Chen E. & Ditto P. H. (2015) Moral coherence processes: Constructing culpability and consequences. *Current Opinion in Psychology* **6**:123–28. [CJC, KJH]
- Cohen J. (2003) On the structural properties of the colors. *Australasian Journal of Philosophy* **81**:78–95. [CK]
- Comunian A. L. & Gielen U. P. (1995) Moral reasoning and prosocial action in Italian culture. *The Journal of Social Psychology* **135**(6): 699–706. [WHBM]
- Cosmides L. & Tooby J. (1992) Cognitive adaptations for social exchange. In: *The adapted mind*, ed. J. Barkow, L. Cosmides & J. Tooby, pp. 163–228. Oxford University Press. [rJM]
- Crawford J. & Brandt M. J. (2019) Who is prejudiced, and toward whom? Big five traits and inclusive generalized prejudice. Available at: <https://doi.org/10.31234/osf.io/6vqwk>. [WHBM]
- Crockett M. J. (2013) Models of morality. *Trends in Cognitive Sciences* **17**(8):363–66. [JH, JR]
- Crockett M. J. (2016a) Computational modeling of moral decisions. In: *The social psychology of morality*, ed. J. P. Forgas, L. Jussim & P. A. M. Van Lange, pp. 87–106. Psychology Press. [JH]
- Crockett M. J. (2016b) How formal models can illuminate mechanisms of moral judgment and decision making. *Current Directions in Psychological Science* **25**(2):85–90. [JH]
- Cummins D. D. (1996) Evidence of deontic reasoning in 3- and 4-year-old children. *Memory and Cognition* **24**(6):823–29. [rJM, JR]
- Cushman F. (2013) Action, outcome and value: A dual-system framework for morality. *Personality and Social Psychology Review* **17**(3): 273–92. [MA, JH, JR]
- Cushman F. (2015) From moral concern to moral constraint. *Current Opinion in Behavioral Sciences* **3**:58–62. [JH]

- 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):1082–89. [aJM, JR]
- Damasio A. (1994/2005) *Descartes' error*. Penguin. (Originally published by Putnam.) [aJM]
- D'Arms J. (2005) Two arguments for sentimentalism. *Philosophical Issues* 15:1–21. [CK]
- D'Arms J. & Jacobson D. (2014) Sentimentalism and scientism. In: *Moral psychology and human agency: Philosophical essays on the science of ethics*, ed. J. D'Arms & D. Jacobson, pp. 253–78. Oxford University Press. [JMD, aJM]
- D'Arms J. & Jacobson D. (forthcoming) *Rational sentimentalism*. Oxford University Press. [JMD]
- Darwin C. (1871/1981) *The descent of man*. Princeton University Press. (Original work published in 1871.) [DN]
- Davidson P., Turiel E. & Black A. (1983) The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. *British Journal of Developmental Psychology* 1(1):49–65. [WHBM]
- Dayan P. & Abbott L. F. (2001) *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press. [JH]
- Dayan P. & Niv Y. (2008) Reinforcement learning: The good, the bad and the ugly. *Current opinion in neurobiology* 18(2):185–96. [JH]
- Deary I. J., Batty G. D. & Gale C. R. (2008) Bright children become enlightened adults. *Psychological Science* 19(1):1–6. [WHBM]
- Deloria V. (2006) *The world we used to live in*. Fulcrum. [DN]
- Demaree-Cotton J. (2016) Do framing effects make moral intuitions unreliable? *Philosophical Psychology* 29(1):1–22. [JD-C, arJM, KM]
- Derakshan N., Eysenck M. & Myers L. (2007) Emotional information processing in repressors: The vigilance–avoidance theory. *Cognition and Emotion* 21:1585–1614. [CK]
- Descola P. (2013) *Beyond nature and culture*, trans. J. Lloyd. University of Chicago Press. [DN]
- Doris J. M. (2002) *Lack of character: Personality and moral behavior*. Cambridge University Press. [JMD]
- Doris J. M. (2015) *Talking to our selves: Reflection, ignorance, and agency*. Oxford University Press. [JMD, QHG, arJM]
- Doris J. M. (2018) Collaborating agents: Values, sociality, and moral responsibility. *Behavioral and Brain Sciences* 41:E65. doi:10.1017/S0140525X17001935. [JMD]
- Doris J. M. & Murphy D. (2007) From My Lai to Abu Ghraib: The moral psychology of atrocity. *Midwest Studies in Philosophy* 31:25–55. [JMD]
- Dunham Y., Baron A. S. & Banaji M. R. (2008) The development of implicit intergroup cognition. *Trends in Cognitive Sciences* 12(7):248–53. [JR]
- Eisenberg N. (2000) Emotion, regulation, and moral development. *Annual Review of Psychology* 51:665–97. [JR]
- Engelmann J. M., Clift J. B., Herrmann E. & Tomasello M. (2017) Social disappointment explains chimpanzees' behaviour in the inequity aversion task. *Proceedings of the Royal Society B: Biological Sciences* 284(1861):20171502. Available at: <http://doi.org/10.1098/rspb.2017.1502>. [KJH]
- Engelmann J. M., Over H., Herrmann E. & Tomasello M. (2013) Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science* 16(6):952–58. [JR]
- Ermer E. & Kiehl K. A. (2010) Psychopaths are impaired in social exchange and precautionary reasoning. *Psychological Science* 21(10):1399–1405. [WHBM]
- Evans J. St. B. T. (2009) How many dual-process theories do we need: One, two or many? In: *In two minds: Dual processes and beyond*, ed. J. St. B. T. Evans & K. Frankish, pp. 31–54. Oxford University Press. [KJH]
- Fehr E., Fischbacher U. & Kosfeld M. (2005) Neuroeconomic foundations of trust and social preferences: Initial evidence. *American Economic Review* 95(2):346–51. [JH]
- Feltz A. & May J. (2017) The means/side-effect distinction in moral cognition: A meta-analysis. *Cognition* 166:314–27. [aJM]
- Fessler D. (2007) From appeasement to conformity: Evolutionary and cultural perspectives on shame, competition, and cooperation. In: *The self-conscious emotions: Theory and research*, ed. J. Tracy, R. Robins & J. Tangney, pp. 174–93. Guilford Press. [CK]
- Festinger L. (1957) *A theory of cognitive dissonance*. Stanford University Press. [CJC]
- Fisher S. E. (2006) Tangled webs: Tracing the connections between genes and cognition. *Cognition* 101:270–97. [JIMC]
- Flanagan O. (2017) *The geography of morals: Varieties of moral possibility*. Oxford University Press. [aJM]
- Foot P. (1967) The problem of abortion and the doctrine of the double effect. *Oxford Review* 5:5–15. [NC]
- Forscher P. S., Lai C. K., Axt J. R., Ebersole C. R., Herman M., Devine P. G. & Nosek B. A. (2017) A meta-analysis of change in implicit bias. Unpublished manuscript. [aJM]
- Four Arrows (2016) *Point of departure: Returning to our more authentic, worldview for education and survival*. Information Age. [DN]
- Frederick S. (2005) Cognitive reflection and decision making. *Journal of Economic Perspectives* 19:25–42. [JFL, rJM]
- Gibbs J. C. (2013) *Moral development and reality: Beyond the theories of Kohlberg, Hoffman, and Haidt*. Oxford University Press. [WHBM]
- Gibson Q. H. (2017) *On the fringes of moral responsibility: Skepticism, self-deception, delusion, and addiction*. Doctoral dissertation, University of California, Berkeley. [QHG]
- Gigerenzer G. (2008) *Gut feelings: The intelligence of the unconscious*. Penguin. [rJM]
- Glenn A. L. & Raine A. (2014) *Psychopathy: An introduction to biological findings and their implications*. New York University Press. [aJM]
- Glimcher P. W., Camerer C. F., Fehr E. & Poldrack R. A. (2009) *Neuroeconomics: Decision making and the brain*. Elsevier. [JH]
- Göckeritz S., Schmidt M. F. H. & Tomasello M. (2014) Young children's creation and transmission of social norms. *Cognitive Development* 30:81–95. [JIMC]
- Gómez-Robles A., Hopkins W. D., Schapiro S. J. & Sherwood C. C. (2015) Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proceedings of the National Academy of Sciences* 12:14799–804. Available at: <http://dx.doi.org/10.1073/pnas.1512646112>. [DN]
- Gopnik A. (2012) Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science* 337(6102):1623–27. [rJM, JR]
- Gottlieb G. (2007) Probabilistic epigenesis. *Developmental Science* 10:1–11. [JIMC]
- Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S. P. & Ditto P. H. (2013) Moral foundations theory: The pragmatic validity of moral pluralism. In: *Advances in experimental social psychology*, vol. 47, ed. P. Devine & A. Plant, pp. 55–130. Academic Press. [arJM]
- 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):1029–46. [rJM, RZ]
- Greene J. (2008) The secret joke of Kant's Soul. In: *Moral psychology*, vol. 3, ed. W. Sinnott-Armstrong, pp. 35–117. MIT Press. [CM]
- Greene J. D. (2013) *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin. [JMD, arJM, JR]
- Greene J. D. (2014) Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics* 124(4):695–726. [JMD, arJM]
- Greene J. D., Cushman F. A., Stewart L. E., Lowenberg K., Nystrom L. E. & Cohen J. D. (2009) Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111(3):364–71. [aJM]
- Greenough W. & Black J. (1992) Induction of brain structure by experience: Substrate for cognitive development. In: *Minnesota symposia on child psychology*, vol. 24: *Developmental behavioral neuroscience*, ed. M. R. Gunnar & C. A. Nelson, pp. 155–200. Erlbaum. [DN]
- Greenwald A. G., Poehlman T. A., Uhlmann E. L. & Banaji M. R. (2009) Understanding and using the implicit association test: III. *Journal of Personality and Social Psychology* 97(1):17–41. [aJM]
- Griffiths P. (2004) Toward a “Machiavellian” theory of emotional appraisal. In: *Emotion, evolution and rationality*, ed. D. Evans & P. Cruse, pp. 89–105. Oxford University Press. [CK]
- Griffiths T. L., Kemp C. & Tenenbaum J. B. (2008) Bayesian models of cognition. In: *Cambridge handbook of computational cognitive modeling*, ed. R. Sun, pp. 59–100. Cambridge University Press. [KJH]
- Grusec J. E. & Goodnow J. J. (1994) Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology* 30(1):4–19. [JR]
- Habermas J. (1983/1990) *Moral consciousness and communicative action*. MIT Press. (Original work published 1983) [JIMC]
- Haidt J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108:814–34. [MA, CJC, JIMC, JFL, JR]
- Haidt J. (2003) The moral emotions. In: *Handbook of affective sciences*, ed. R. J. Davidson, K. R. Scherer & H. H. Goldsmith, pp. 852–70. Oxford University Press. [MA, aJM]
- Haidt J. (2012) *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books/Vintage Books. [QHG, DH, aJM]
- Haidt J. (2016) Comfortably dumbfounded. In: *A very bad wizard: Morality behind the curtain*, 2nd ed., ed. Tamler Sommers, 235–52. Routledge. [DH]
- Hamilton M. C. (1991) Masculine bias in the attribution of personhood: People = Male, Male = People. *Psychology of Women Quarterly* 15(3):393–402. [RZ]
- Hamlin J. K. (2013) Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science* 22(3):186–93. [JR]
- Hawthorne J. & Stanley J. (2008) Knowledge and action. *Journal of Philosophy* 105(10):571–90. [rJM]
- Helzer E. G. & Pizarro D. A. (2011) Dirty liberals! Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science* 22(4):517–22. [KM]
- Henrich J. (2015) *The secret of our success*. Princeton University Press. [aJM]
- Henrich J., Heine S. J. & Norenzayan A. (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3):61–83. [DN]
- Herman B. (1993) *The practice of moral judgement*. Harvard University Press. [CM]
- Hetter K. (2015) J. K. Rowling's reply to critic of gay Dumbledore. CNN, March 25. Available at: <https://edition.cnn.com/2015/03/25/entertainment/feat-rowling-dumbledore-gay-tweet/index.html>. [RZ]
- Hewlett B. S. & Lamb M. E. (2005) *Hunter-gatherer childhoods: Evolutionary, developmental and cultural perspectives*. Aldine. [DN]

- Hill T. E. (1992) *Dignity and practical reason in Kant's moral theory*. Cornell University Press. [AJR]
- Hoffman M. L. (1975) Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology* 11(5):607–22. [JR]
- Hofmann S., Ellard K. & Siegle G. (2012) Neurobiological correlates of cognitions in fear and anxiety. *Cognition and Emotion* 26:282–99. [CK]
- Holmes A. (2012) White until proven black: Imagining race in hunger games. *The New Yorker* (March 30). Available at: <https://www.newyorker.com/books/page-turner/white-until-proven-black-imagining-race-in-hunger-games>. [RZ]
- Holton R. (2009) *Willing, wanting, waiting*. Clarendon. [aJM]
- Holyoak K. J. (2019) *The spider's thread: Metaphor in mind, brain, and poetry*. MIT Press. [KJH]
- Holyoak K. J. & Powell D. (2016) Deontological coherence: A framework for common-sense moral reasoning. *Psychological Bulletin* 142(11):1179–1203. [KJH, WHBM, JR]
- Holyoak K. J. & Simon D. (1999) Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General* 128:3–31. [KJH]
- Horne Z. & Livengood J. (2017) Ordering effects, updating effects, and the specter of global skepticism. *Synthese* 194(4):1189–218. [rJM]
- Horne Z. & Powell D. (2016) How large is the role of emotion in judgments of moral dilemmas? *PLOS ONE* 11(7): e0154780. Available at: <http://doi.org/10.1371/journal.pone.0154780>. [KJH]
- Horne Z., Powell D. & Hummel J. (2015) A single counterexample leads to moral belief revision. *Cognitive Science* 39:1950–64. [KJH]
- Huebner B. (2015) Do emotions play a constitutive role in moral cognition? *Topoi* 34(2):427–40. [aJM]
- Huebner B. (2016) Implicit bias, reinforcement learning, and scaffolded moral cognition. In: *Implicit bias and philosophy: Vol. 1, Meta- physics and epistemology*, ed. M. Brownstein & J. Saul, pp. 47–79. Oxford University Press. [JH]
- Hume D. (2006) *Moral philosophy*, ed. G. Sayre-McCord. Hackett. [AK]
- Hurka T. (2014) Many faces of virtue. *Philosophy and phenomenological research* 89(2):496–503. [aJM]
- Ingold T. (2005) On the social relations of the hunter-gatherer band. In: *The Cambridge encyclopedia of hunters and gatherers*, ed. R. B. Lee & R. Daly, pp. 399–410. Cambridge University Press. [DN]
- Intergovernmental Panel on Climate Change (IPCC) (2014) *Climate change 2014: Synthesis report*. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change, ed. R. K. Pachauri & L. A. Meyer (core writing team). IPCC, Geneva, Switzerland. [DN]
- Iurino K., Robinson B., Christen M., Stey P. & Alfano M. (2018) Constructing and validating a scale of inquisitive curiosity. In: *The moral psychology of curiosity*, ed. I. Inan, L. Watson, D. Whitcomb & S. Yigit, pp. 157–81. Rowman & Littlefield. [MA]
- Izard C. (2007) Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science* 2:260–80. [CK]
- Jablonka E. & Lamb M. J. (2005) *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT Press. [DN]
- Johnson-Laird P. (2008) Mental models and deductive reasoning. In: *Reasoning: Studies in human inference and its foundations*, ed. L. Rips & J. Adler, pp. 207–22. Cambridge University Press. [MA]
- Johnston M. (1992) How to speak of the colors. *Philosophical Studies* 68: 221–63. [CK]
- Johnston M. (2010) *Surviving death*. Princeton University Press. [aJM]
- Jost J. T. (2017) Ideological asymmetries and the essence of political psychology. *Political Psychology* 38(2):167–208. [WHBM]
- Jost J. T., Banaji M. R. & Nosek B. A. (2004) A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology* 25(6): 881–919. [RZ]
- Joyce R. (2006) *The evolution of morality*. MIT Press. [arJM]
- Jussim L. (2017) Précis of *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-Fulfilling Prophecy*. *Behavioral and Brain Sciences* 1–65. [WHBM]
- Kagan J. (1987) Introduction. In: *The emergence of morality in young children*, ed. J. Kagan & S. Lamb, pp. ix–xx. The University of Chicago Press. [JR]
- Kahan D. M., Peters E., Wittlin M., Slovic P., Ouellette L. L., Braman D. & Mandel G. (2012) The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change* 2(6): 732–35. [rJM]
- Kahane G. (2011) Evolutionary debunking arguments. *Noûs* 45(1):103–25. [aJM]
- Kahane G., Everett J. A. C., Earp B. D., Farias M. & Savulescu J. (2015) “Utilitarian” judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition* 134(C):193–209. [rJM]
- Kahneman D. (2011) *Thinking, fast and slow*. Farrar, Straus and Giroux. [KJH, JFL]
- Kant I. (1785/2002) *Groundwork for the metaphysics of morals*. Yale University Press. (Original work published in 1785.) [KJH]
- Kant I. (1996) *Practical philosophy*, trans. M. Gregor. Cambridge University Press. [CM, AJR]
- Kant I. (1998) *Critique of pure reason*, trans. P. Guyer & A. Wood. Cambridge University Press. [CM]
- Kauppinen A. (2013a) A Humean theory of moral intuition. *Canadian Journal of Philosophy* 43: 360–81. [CK]
- Kauppinen A. (2013b) Moral sentimentalism. In: *Stanford encyclopedia of philosophy, winter 2018 edition*, ed. Edward N. Zalta. Stanford, CA: Stanford University, Center for the Study of Language and Information. Available at: <https://plato.stanford.edu/archives/win2018/entries/moral-sentimentalism>. [AK]
- Kauppinen A. (2014) Empathy, emotion regulation, and moral judgment. In: *Empathy and morality*, ed. H. Maibom, pp. 97–121. Oxford University Press. [AK]
- Kelly D. (2011) *Yuck!: The nature and moral significance of disgust*. MIT Press. [aJM]
- Kelly D., Stich S., Haley K. J., Eng S. J. & Fessler D. M. (2007) Harm, affect, and the moral/conventional distinction. *Mind and Language* 22(2):117–31. [KM]
- Kennett J. & Fine C. (2008) Internalism and the evidence from psychopaths and “acquired sociopaths.” In: *Moral psychology, vol. 3*, ed. W. Sinnott-Armstrong, pp. 173–90. MIT Press. [aJM]
- Khemlani S., Orenes I. & Johnson-Laird P. (2012) Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology* 24(5): 541–59. [MA]
- Kidner D. W. (2001) *Nature and psyche: Radical environmentalism and the politics of subjectivity*. State University of New York. [DN]
- Killgore W. D. S., Kahn-Greene E. T., Lipizzi E. L., Newman R. A., Kamimori G. H. & Balkin T. J. (2007) Sleep deprivation reduces perceived emotional intelligence and constructive thinking skills. *Sleep Medicine* 9(5):517–26. [KM]
- Knowlton B. J., Morrison R. G., Hummel J. E. & Holyoak K. J. (2012) A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences* 16:373–81. [KJH]
- Knudsen E. I. (2004) Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience* 16(8):1412–25. [DN]
- Kochanska G. (2002) Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science* 11(6):191–95. Available at: <http://doi.org/10.1111/1467-8721.00198>. [DN]
- Kohlberg L. (1971) From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In: *Cognitive development and epistemology*, ed. T. Mischel, pp. 151–235. Academic Press. [JR]
- Kohlberg L. (1981) *Essays in moral development: The philosophy of moral development, vol. 1*. Harper & Row. [JIMC]
- Kokis J. V., Macpherson R., Toplak M. E., West R. F. & Stanovich K. E. (2002) Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology* 83:26–52. [JFL]
- Kolbert E. (2014) *The sixth extinction: An unnatural history*. Holt. [DN]
- Kolers A. (2016) *A moral theory of solidarity*. Oxford University Press. [RZ]
- Koralus P. & Alfano M. (2017) Reasons-based moral judgment and the erotetic theory. In: *Moral inferences*, ed. J.-F. Bonnefon & B. Trémolière, pp. 77–106. Routledge. [MA]
- Koralus P. & Mascarenhas S. (2013) The erotetic theory of reasoning: Bridges between formal semantics and the psychology of propositional deductive inference. *Philosophical Perspectives* 27:312–65. [MA]
- Korsgaard C. M. (1996a) *Creating the kingdom of ends*. Cambridge University Press. [rJM, AJR]
- Korsgaard C. M. (1996b) *The sources of normativity*. Cambridge University Press. [AJR]
- Korsgaard C. M. (2008) *The constitution of agency: Essays on practical reason and moral psychology*. Oxford University Press. [AJR]
- Korsgaard C. M. (2009a) The activity of reason. *Proceedings and Addresses of the APA* 83(2):23–43. [AJR]
- Korsgaard C. M. (2009b) *Self-Constitution: Agency, identity, and integrity*. Oxford University Press. [AJR]
- Korsgaard C. M. (2018) *Fellow creatures: Our obligations to the other animals*. Oxford University Press. [AJR]
- Kounios J. & Beeman M. (2015) *The eureka factor: Aha moments, creative insight, and the brain*. Random House. [KJH]
- Krebs D. L. (2008) Morality: An evolutionary account. *Perspectives on Psychological Science* 3(3):149–72. [JR]
- Krebs D. L. & Denton K. (2005) Toward a more pragmatic approach to morality: A critical evaluation of Kohlberg's model. *Psychological Review* 112(3):629–49. [WHBM]
- Kubricht J. R., Lu H. & Holyoak K. J. (2017) Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences* 21:749–59. [KJH]
- Kühberger A. (1998) The influence of framing on risky decisions: A meta-analysis. *Organizational behavior and human decision processes* 75(1):23–55. [arJM, KM]
- Kumar V. (2017) Foul behavior. *Philosophers' Imprint* 17(15):1–16. [aJM]
- Kumar V. & Campbell R. (2012) On the normative significance of experimental moral psychology. *Philosophical Psychology* 25(3):311–30. [aJM]
- Kumar V. & May J. (2019) How to debunk moral beliefs. In: *Methodology and moral philosophy*, ed. J. Suikkanen & A. Kauppinen, pp. 25–48. Routledge. [arJM]
- Kunda Z. (1990) The case for motivated reasoning. *Psychological Bulletin* 108(3):480–98. [arJM]
- Kurth C. (2016) Anxiety, normative uncertainty, and social regulation. *Biology and Philosophy* 31:1–21. [CK]

- Kurth C. (2018) *The anxious mind: An investigation into the varieties and virtues of anxiety*. MIT Press. [CK]
- Lagnado D. A. & Gerstenberg T. (2017) Causation in legal and moral reasoning. In: *Oxford handbook of causal reasoning*, ed. M. R. Waldmann, pp. 562–602. Oxford University Press. [KJH]
- Lakoff G. & Johnson M. (1999) *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Harper Collins. [DN]
- Landy J. F. (2016) Representations of moral violations: Category members and associated features. *Judgment and Decision Making* 11(5): 496–508. [JFL, WHBM]
- Landy J. F. & Bartels D. M. (2018) An empirically-derived taxonomy of moral concepts. *Journal of Experimental Psychology: General* 147:1148–61. [JFL]
- 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):518–36. [MA, JFL, aJM]
- Landy J. F. & Royzman E. B. (2018) The moral myopia model: Why and how reasoning matters in moral judgment. In: *The new reflectionism in cognitive psychology: Why reason matters*, ed. G. Pennycook, pp. 70–92. Psychology. [JFL]
- Latané B. & Nida S. (1981) Ten years of research on group size and helping. *Psychological Bulletin* 89(2):308–24. [aJM]
- Lawrence J. A. (1987) Verbal processing of the Defining Issues Test by principled and non-principled moral reasoners. *Journal of Moral Education* 16(2):117–30. [WHBM]
- Lee R. B. (1979) *The !Kung San: Men, women, and work in a foraging community*. Cambridge University Press. [DN]
- Lee R. B. & Daly R., eds. (2005) *The Cambridge encyclopedia of hunters and gatherers*. Cambridge University Press. [DN]
- Leimgruber K. L., Shaw A., Santos L. & Olson K. R. (2012) Young children are more generous when others are aware of their actions. *PLOS ONE* 7(10):e48292. [JR]
- Lepore J. (2018) *These truths: A history of the United States*. Norton. [WHBM]
- Lerner M. (1980) *Belief in a just world: A fundamental delusion*. Plenum. [RZ]
- Levenson R., Soto J. & Pole N. (2007) Emotion, biology, and culture. In: *Handbook of cultural psychology*, ed. S. Kitayama & D. Cohen, pp. 780–96. Guilford Press. [CK]
- Lickliter R. & Honeycutt H. (2009) Rethinking epigenesis and evolution in light of developmental science. In: *Oxford handbook of developmental behavioral neuroscience*, ed. M. S. Blumberg, J. H. Freeman & S. R. Robinson, pp. 30–49. Oxford University Press. [JIMC]
- Lickliter R. & Honeycutt H. (2015) Biology, development, and human systems. In: *Theory and method: Handbook of child psychology and developmental science, vol. 1, 7th ed.*, ed. R. Lerner (editor-in-chief), W. F. Overton & P. C. M. Molenaar, pp. 162–207. Wiley Blackwell. [JIMC]
- Locke D. (2015) Practical certainty. *Philosophy and Phenomenological Research* 90(1): 72–95. [rJM]
- Lourenço O. (2003) Making sense of Turiel's dispute with Kohlberg: The case of the child's moral competence. *New Ideas in Psychology* 21(1):43–68. [WHBM]
- Lupien J., McEwen B. S., Gunnar M. R. & Heim C. (2009) Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience* 10(6):434–45. [DN]
- Machery E. (2010) The bleak implications of moral psychology. *Neuroethics* 3(3):223–31. [JMD]
- Mackie J. L. (1977) *Ethics: Inventing right and wrong*. Penguin. [JFL]
- Mackintosh N. J. (1983) *Conditioning and associative learning*. Oxford University Press. [JH]
- Madva A. (2016) A Plea for anti-anti-individualism: How oversimple psychology misleads social policy. *Ergo* 3(27):701–728. [rJM]
- Maibom H. L. (2005) Moral unreason: The case of psychopathy. *Mind and Language* 20(2):237–57. [aJM]
- Maibom H. L. (2010) What experimental evidence shows us about the role of emotions in moral judgement. *Philosophy Compass* 5(11):999–1012. [QHG]
- Mallon R. & Nichols S. (2010) Rules. In: *The moral psychology handbook*, ed. J. M. Doris & The Moral Psychology Research Group, pp. 297–320. Oxford University Press. [aJM]
- Mameli M. & Bateson P. (2006) Innateness and the sciences. *Biology and Philosophy* 21:155–88. [JIMC]
- Manne K. (2017) *Down girl: The logic of misogyny*. Oxford University Press. [rJM]
- Marsh A. A. & Blair R. J. R. (2008) Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews* 32(3):454–65. [aJM]
- Marshall C. (2018) *Compassionate moral realism*. Oxford University Press. [rJM]
- Marshall J., Watts A. L., Frankel E. L. & Lilienfeld S. O. (2017) An examination of psychopathy's relationship with two indices of moral judgment. *Personality and Individual Differences* 113:240–45. [WHBM]
- Martin A. & Olson K. R. (2015) Beyond good and evil: What motivations underlie children's prosocial behavior? *Perspectives on Psychological Science* 10(2):159–75. [JR]
- May J. (2011) Egoism, empathy, and self-other merging, Spindel supplement: Empathy and ethics, ed. R. Debes. *Southern Journal of Philosophy* 49(S1):25–39. [aJM]
- May J. (2013a) Because I believe it's the right thing to do. *Ethical Theory and Moral Practice* 16(4):791–808. [aJM]
- May J. (2013b) Skeptical hypotheses and moral skepticism. *Canadian Journal of Philosophy* 43(3):341–59. [arJM]
- May J. (2014) Does disgust influence moral judgment? *Australasian Journal of Philosophy* 92(1):125–41. [aJM]
- May J. (2018) *Regard for reason in the moral mind*. Oxford University Press. [MA, CJC, JIMC, NC, JD-C, JMD, QHG, DH, JH, KJH, AK, CK, JFL, CM, arJM, KM, WHBM, DN, JR, AJR, NS, RZ]
- Mazar N., Amir O. & Ariely D. (2008) The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6):633–44. [aJM]
- Mazar N. & Zhong C. B. (2010) Do green products make us better people? *Psychological Science* 21(4):494–98. [aJM]
- McCullough M. E. (forthcoming) *Why we give a damn*. Basic Books. [WHBM]
- McGilchrist I. (2009) *The master and his emissary: The divided brain and the making of the western world*. Yale University Press. [DN]
- McGrath S. (2008) Moral disagreement and moral expertise. In: *Oxford studies in meta-ethics, vol. 3*, ed. R. Shafer-Landau, pp. 87–107. Oxford University Press. [aJM]
- Mead G. H. (1934) *Mind, self and society: From the standpoint of a social behaviorist*. University of Chicago Press. [JIMC]
- Meaney M. J. (2010) Epigenetics and the biological definition of gene x environment interactions. *Child Development* 81:41–79. [JIMC]
- Melkonyan T., Zeitoun H. & Chater N. (2018) Collusion in Bertrand versus Cournot competition: A virtual bargaining approach. *Management Science* 64(12):5461–59. Available at: <https://doi.org/10.1287/mnsc.2017.2878>. [NC]
- Merchant C. (2003) *Reinventing Eden: The fate of nature in Western culture*. Routledge. [DN]
- Mercier H. & Sperber D. (2011) Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34:57–74. [CJC]
- Merritt R. D. & Kok C. J. (1995) Attribution of gender to a gender-unspecified individual: An evaluation of the people = male hypothesis. *Sex Roles* 33(3–4):145–57. [RZ]
- Mikhail J. (2011) *Elements of moral cognition*. Cambridge University Press. [arJM]
- Millennium Ecosystem Assessment (2005) *Ecosystems and human well-being: Synthesis*. Island. [DN]
- Miller C., Furr M. R., Knobel A. & Fleeson W., eds. (2015) *Character: New directions from philosophy, psychology, and theology*. Oxford University Press. [JMD]
- Miller C. B. (2013) *Moral character: An empirical theory*. Oxford University Press. [aJM]
- Misyak J. B., Melkonyan T., Zeitoun H. & Chater N. (2014) Unwritten rules: Virtual bargaining underpins social interaction, culture, and society. *Trends in Cognitive Sciences* 18(10):512–19. [NC]
- Montgomery M. A., Kappes A. & Crockett M. J. (2017) Compassion is not always a motivated choice: A multiple decision systems perspective. *Moral psychology: Vol. 5, Virtue and character*, ed. W. Sinnott-Armstrong & C. B. Miller, pp. 409–18. MIT Press. [JH]
- Moors A., Ellsworth P. C., Scherer K. R. & Frijda N. H. (2013) Appraisal theories of emotion: State of the art and future development. *Emotion Review* 5:119–24. [KJH]
- Muller J. Z. (2018) *The tyranny of metrics*. Princeton University Press. [DN]
- Murphy D. & Doris J. M. (forthcoming) Skepticism about evil: Atrocity and the limits of responsibility. In: *The Oxford handbook of moral responsibility*, ed. D. Nelkin & D. Pereboom. Oxford University Press. [JMD]
- Nadelhoffer T. & Feltz A. (2008) The actor–observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. *Neuroethics* 1(2):133–44. [KM]
- Nahmias E. (2007) Autonomous agency and social psychology. In: *Cartographies of the mind: Philosophy and psychology in intersection*, ed. M. Marraffa, M. Caro & F. Ferretti, pp. 169–85. Springer. [JMD]
- Narvaez D. (2010) Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science* 5(2):163–81. [DN]
- Narvaez D. (2013) The 99% – Development and socialization within an evolutionary context: Growing up to become “a good and useful human being.” In: *War, peace and human nature: The convergence of evolutionary and cultural views*, ed. D. Fry, pp. 643–72. Oxford University Press. [DN]
- Narvaez D. (2014) *Neurobiology and the development of human morality: Evolution, culture and wisdom*. Norton. [DN]
- Narvaez D. (2016a) *Embodied morality: Protectionism, engagement and imagination*. Palgrave-Macmillan. [DN]
- Narvaez D. (2016b) Kohlberg Memorial Lecture 2015: Revitalizing human virtue by restoring organic morality. *Journal of Moral Education* 45(3):223–38. [DN]
- Narvaez D. (2017) Are we losing it? Darwin's moral sense and the importance of early experience. In: *Routledge handbook of evolution and philosophy*, ed. R. Joyce, pp. 322–32. Routledge. [DN]
- Narvaez D., ed. (2018a) *Basic needs, wellbeing and morality: Fulfilling human potential*. Palgrave-Macmillan. [DN]
- Narvaez D. (2018b) Ethogenesis: Evolution, early experience and moral becoming. In: *The atlas of moral psychology*, ed. J. Graham & K. Gray, pp. 451–64. Guilford Press. [DN]

- Narvaez D., Four Arrows, Halton E., Collier B. & Enderle G., eds. (2019) *Indigenous sustainable wisdom: First Nation know-how for global flourishing*. Lang. [DN]
- Narvaez D., Gleason T., Wang L., Brooks J., Lefever J., Cheng A. & Centers for the Prevention of Child Neglect (2013a) The evolved development niche: Longitudinal effects of caregiving practices on early childhood psychosocial development. *Early Childhood Research Quarterly* 28(4):759–73. Available at: <http://doi.org/10.1016/j.ecresq.2013.07.003>. [DN]
- Narvaez D., Panksepp J., Schore A. & Gleason T., eds. (2013b) *Evolution, early experience and human development: From research to practice and policy*. Oxford University Press. [DN]
- Narvaez D., Wang L. & Cheng A. (2016) The evolved developmental niche in childhood: Relation to adult psychopathology and morality. *Applied Developmental Science* 20(4):294–309. Available at: <http://dx.doi.org/10.1080/10888691.2015.1128835>. [DN]
- Narvaez D. & Witherington D. (2018) Getting to baselines for human nature, development and wellbeing. *Archives of Scientific Psychology* 6(1):205–13. [DN]
- Nelkin D. K. (2005) Freedom, responsibility and the challenge of situationism. *Midwest Studies in Philosophy* 29(1):181–206. [JMD, aJM]
- Nelson M. K. (2008) *Original instructions: Indigenous teachings for a sustainable future*. Bear. [DN]
- Nichols S. (2002) Norms with feeling: Towards a psychological account of moral judgment. *Cognition* 84(2):221–36. [aJM]
- Nichols S. (2004) *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press. [JMD, AK, arJM]
- Nichols S. (2014) Process debunking and ethics. *Ethics* 124:727–49. [aJM]
- Nichols S., Kumar S., Lopez T., Ayars A. & Chan H. Y. (2016) Rational learners and moral rules. *Mind and Language* 31(5):530–54. [JMD, arJM]
- Nisbett R. E. & Wilson T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3):231–59. [CJC, JR]
- Nucci L. P. (1984) Evaluating teachers as social agents: Students' ratings of domain appropriate and domain inappropriate teacher responses to transgressions. *American Educational Research Journal* 21(2):367–78. [JR]
- Nucci L. P. & Turiel E. (1978) Social interactions and the development of social concepts in preschool children. *Child Development* 49:400–07. [JR]
- Nussbaum M. C. (2004) *Hiding from humanity: Disgust, shame, and the law*. Princeton University Press. [aJM]
- Oaksford N. & Chater N. (2013) Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking and Reasoning* 19:346–79. [KJH]
- Öhman A. (2008) Fear and anxiety. In: *Handbook of emotions*, ed. M. Lewis, J. Haviland-Jones & L. F. Barrett, pp. 127–156. Guilford Press. [CK]
- O'Kane A., Fawcett D. & Blackburn R. (1996) Psychopathy and moral reasoning: Comparison of two classifications. *Personality and Individual Differences* 20:505–14. [WHBM]
- Oliner S. & Oliner P. (1988) *The altruistic personality: Rescuers of Jews in Nazi Europe*. Free Press. [WHBM]
- Olsen O. K., Pallesen S. & Eid J. (2010) The impact of partial sleep deprivation on moral reasoning in military officers. *Sleep* 33(8):1086–90. [KM]
- Oswald F. L., Mitchell G., Blanton H., Jaccard J. & Tetlock P. E. (2013) Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology* 105(2):171–92. [aJM]
- Overton W. F. (2013) A new paradigm for developmental science: Relationism and individual-developmental-systems. *Applied Developmental Science* 17(2):94–107. [DN]
- Oyama S., Griffiths P. E. & Gray R. D. (2001) *Cycles of contingency: Developmental systems and evolution*. MIT Press. [DN]
- Paxton J. M. & Greene J. D. (2010) Moral reasoning: Hints and allegations. *Topics in Cognitive Science* 2(3):511–27. [JR]
- Payne B. K. (2001) Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology* 81(2):181–92. [aJM]
- Pelham B. W., Mirenberg M. C. & Jones J. T. (2002) Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology* 82(4):469–87. [QHJ]
- Penn D. C., Holyoak K. J. & Povinelli D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31:109–30. [KJH]
- Pennycook G., Cheyne J. A., Koehler D. J. & Fugelsang J. A. (2016) Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods* 48:341–48. [JFL]
- Pennycook G. & Ross R. M. (2016) Commentary: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology* 7:9. [JFL]
- Perales F. (2018) The cognitive roots of prejudice towards same-sex couples: An analysis of an Australian national sample. *Intelligence* 68:117–27. [WHBM]
- Perry J. (1979) The problem of the essential indexical. *Noûs* 13(1):3–21. [aJM]
- Pessoa L. & Pereira M. G. (2013) Cognition–emotion interactions: A review of the functional magnetic resonance imaging literature. In: *Handbook of cognition and emotion*, ed. M. D. Robinson, E. Watkins & E. Harmon-Jones, pp. 55–68. Guilford Press. [KJH]
- Petrinovich L. & O'Neill P. (1996) Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology* 17(3):145–71. [KM]
- Piaget J. (1932/1965) *The moral judgement of the child*, trans. M. Gabain. Free Press/Harcourt. (Original work published in 1932.) [JIMC, WHBM, JR]
- Pinker S. (2011a) *The better angels of our nature: The decline of violence in history and its causes*. Penguin. [WHBM]
- Pinker S. (2011b) *The better angels of our nature: Why violence has declined*. Viking. [CJC]
- Pinker S. (2018) *Enlightenment now: The case for reason, science, humanism, and progress*. Viking. [WHBM]
- Pizarro D. A. & Bloom P. (2003) The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review* 110(1):193–96. [JR]
- Polanyi K. (2001) *The great transformation: The political and economic origins of our time*, 2nd ed. Beacon. [DN]
- Prinz J. (2007) *The emotional construction of morals*. Oxford University Press. [JMD, AK, arJM]
- Prinz J. (2016) Sentimentalism and the moral brain. In: *Moral brains: The neuroscience of morality*, ed. S. Matthew Liao, pp. 45–73. Oxford University Press. [MA, aJM]
- Rai T. S. & Holyoak K. J. (2010) Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science* 34:311–21. [KJH]
- Railton P. (2017) Moral learning: Conceptual foundations and normative relevance. *Cognition* 167:172–90. [arJM]
- Rakoczy H. & Schmidt M. F. H. (2013) The early ontogeny of social norms. *Child Development Perspectives* 7(1):17–21. [JR]
- Rangel A., Camerer C. & Montague P. R. (2008) A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience* 9(7):545. [JH]
- Rawls J. (1971) *A theory of justice*. Harvard University Press. [KJH]
- Rest J. R., Narvaez D., Bebeau M. & Thoma S. (1999) *Postconventional moral thinking: A neo-Kohlbergian approach*. Erlbaum. [WHBM, DN]
- Rhodes M. & Wellman H. (2017) Moral learning as intuitive theory revision. *Cognition* 167:191–200. [JR]
- Rini R. A. (2016) Debunking debunking: A regress challenge for psychological threats to moral judgment. *Philosophical Studies* 173(3):675–97. [rJM]
- Rips L. (1994) *The psychology of proof*. MIT Press. [MA]
- Rosenthal D. (2005) *Consciousness and Mind*. Oxford University Press. [MA]
- Roskies A. (2003) Are ethical judgments intrinsically motivational? Lessons from “acquired sociopathy.” *Philosophical Psychology* 16(1):51–66. [aJM]
- Ross L. (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology* 10:173–220. [RZ]
- Ross R. (2006) *Returning to the teachings: Exploring aboriginal justice*. Penguin Canada. [DN]
- Rottman J., Kelemen D. & Young L. (2014) Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition* 130(2):217–26. [JR]
- Rottman J. & Young L. (2015) Mechanisms of moral development. In: *The moral brain: A multidisciplinary approach*, ed. J. Decety & T. Wheatley, pp. 123–42. MIT Press. [JR]
- Rottman J., Young L. & Kelemen D. (2014) The impact of testimony on children's moralization of novel actions. *Emotion* 17(5):811–27. [JR]
- Royzman E., Atanasov P., Landy J. F., Parks A. & Gepty A. (2014a) CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the divinity code. *Emotion* 14:892–907. [JFL]
- Royzman E. B., Kim K. & Leeman R. F. (2015a) The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making* 10(4):296–313. [NC]
- Royzman E. B., Landy J. F. & Goodwin G. P. (2014b) Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision Making* 9(3):176–90. [JFL, WHBM]
- Royzman E. B., Landy J. F. & Leeman R. F. (2015b) Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science* 39(2):325–52. [JFL, WHBM]
- Royzman E. B., Leeman R. F. & Baron J. (2009) Unsentimental ethics: Towards a content-specific account of the moral–conventional distinction. *Cognition* 112(1):159–74. [aJM]
- 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. [aJM]
- Saltzstein H. D. & Kaschhoff T. (2004) Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology* 8(4):273–82. [JR]
- Scanlon T. (1998) *What we owe to each other*. Harvard University Press. [QHJ, rJM]
- Schein C. & Gray K. (2018) The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review* 22:32–70. [CJC]

- Schnall S., Haidt J., Clore G. L. & Jordan A. H. (2008) Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34(8):1096–09. [JFL, aJM, KM]
- Schore A. N. (2003a) *Affect dysregulation and disorders of the self*. Norton. [DN]
- Schore A. N. (2003b) *Affect regulation and the repair of the self*. Norton. [DN]
- Schore A. N. (2017) All our sons: The developmental neurobiology and neuroendocrinology of boys at risk. *Infant Mental Health Journal* 38(1):15–52. Available at: <http://doi.org/10.1002/imhj.21616>. [DN]
- Schroeder T. (2004) *Three faces of desire*. Oxford University Press. [aJM]
- Schroeder T., Roskies A. & Nichols S. (2010) Moral motivation. In: *The moral psychology handbook*, ed. J. Doris & The Moral Psychology Research Group, pp. 72–110. Oxford University Press. [aJM]
- Schulz L. (2012) The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences* 16(7):382–89. [JR]
- Schulz L. E., Bonawitz E. B. & Griffiths T. L. (2007) Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental psychology* 43(5):124–39. [JR]
- Schwarz N. & Clore G. L. (1983) Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45(3):513–23. [aJM]
- Schwitzgebel E. & Cushman F. A. (2012) Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27(2):135–53. [aJM]
- Shafir E. (1993) Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition* 21(4):546–56. [MA]
- Shaver K. G. (1970) Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology* 14(2):101–13. [RZ]
- Shaw A., Montinari N., Piovesan M., Olson K. R., Gino F. & Norton M. I. (2014) Children develop a veil of fairness. *Journal of Experimental Psychology: General* 143(1):363–75. [JR]
- Shenhav A. & Greene J. D. (2010) Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron* 67(4):667–77. [JH]
- Shermer M. (2015) *The moral arc: How science and reason lead humanity toward truth, justice, and freedom*. Macmillan. [WHBM]
- Shonkoff J. P. & Phillips D. A., eds. (2000) *From neurons to neighborhoods: The science of early childhood development*. National Academies Press. [DN]
- Shweder R. A., Turiel E. & Much N. C. (1981) The moral intuitions of the child. In: *Social cognitive development: Frontiers and possible futures*, ed. J. H. Flavell & L. Ross, pp. 288–305. Cambridge University Press. [JR]
- Silvia P. J., Wigert B., Reiter-Palmon R. & Kaufman J. C. (2012) Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts* 6(1):19–34. [MA]
- Simon D. (2012) *In doubt: The psychology of the criminal justice process*. Harvard University Press. [KJH]
- Simon D. & Holyoak K. J. (2002) Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review* 6:283–94. [KJH]
- Simon D., Stenstrom D. M. & Read S. J. (2015) The coherence effect: Blending hot and cold cognitions. *Journal of Personality and Social Psychology* 109:369–94. [KJH]
- Simonsohn U. (2011) Spurious? Name similarity effects (implicit egotism) in marriage, job, and moving decisions. *Journal of Personality and Social Psychology* 101(1):1–24. [QHG]
- Singer P. (1999) The Singer solution to world poverty. *The New York Times Magazine* (September 5):60–63. Available at: <https://www.nytimes.com/1999/09/05/magazine/the-singer-solution-to-world-poverty.html>. [JMD]
- Singer P. (2005) Ethics and intuitions. *The Journal of Ethics* 9:331–52. [JMD, arJM]
- Singer P. (2015) *The most good you can do: How effective altruism is changing ideas about living ethically*. Text Publishing. [JMD]
- Sinhababu N. (2017) *Humean nature: How desire explains action, thought, and feeling*. Oxford University Press. [aJM, NS]
- Sinnott-Armstrong W. (2008) Framing moral intuitions. In: *Moral psychology, vol. 2: The cognitive science of morality: Intuition and diversity*, ed. W. Sinnott-Armstrong, pp. 47–76. MIT Press. [JD-C]
- Sinnott-Armstrong W. (2011) Emotion and reliability in moral psychology. *Emotion Review* 3(3):288–89. [JR]
- Smetana J. G. (2006) Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In: *Handbook of moral development*, ed. M. Killen & J. Smetana, pp. 119–53. Erlbaum. [JR]
- Smith A. (2002) *The theory of moral sentiments*, ed. Knut Haakonssen. Cambridge University Press. [AK]
- Smith M. (1994) *The moral problem*. Blackwell. [AK]
- Sobel D. M. & Kirkham N. Z. (2006) Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology* 44:1103–15. [JR]
- Sobel D. M. & Kushnir T. (2013) Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review* 120(4):779–97. [JR]
- Sober E. & Wilson D. S. (1998) *Unto others. The evolution and psychology of unselfish behavior*. Harvard University Press. [AK]
- Sousa P. & Piazza J. (2014) Harmful transgressions qua moral transgressions: A deflationary view. *Thinking and Reasoning* 20(1):99–128. [WHBM]
- Spellman B. A., Ullman J. B. & Holyoak K. J. (1993) A coherence model of cognitive consistency. *Journal of Social Issues* 4:147–65. [KJH]
- Spinoza B. (1988) *Collected works, vol. 1*, trans. E. Curley. Princeton University Press. [CM]
- Stams G. J., Brugman D., Deković M., Van Rosmalen L., Van Der Laan P. & Gibbs J. C. (2006) The moral judgment of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology* 34(5):692–708. [WHBM]
- Stanovich K. E., West R. F. & Toplak M. E. (2013) Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science* 22(4):259–64. [WHBM]
- Stiles J., Brown T. T., Haist F. & Jernigan T. L. (2015) Brain and cognitive development. In: *Cognitive processes: Handbook of child psychology and developmental science, vol. 2, 7th ed.*, ed. R. Lerner (editor-in-chief), L. Liben & U. Müller, pp. 9–62. Wiley Blackwell. [JIMC]
- Street S. (2006) A Darwinian dilemma for realist theories of value. *Philosophical Studies* 127(1):109–66. [AK]
- Street S. (2010) What is constructivism in ethics and metaethics? *Philosophy Compass* 5(5):363–84. [AJR]
- Sugden R. (2018) *The community of advantage: A behavioural economist's defence of the market*. Oxford University Press. [NC]
- Summers J. S. (2017) Post hoc ergo propter hoc: Some benefits of rationalization. *Philosophical Explorations* 20(suppl. 1):21–36. [rJM]
- Sunstein C. R. (2005) Moral heuristics. *Behavioral and Brain Sciences* 28(4):531–42. [aJM]
- Sutton R. S. & Barto A. G. (1998) *Introduction to reinforcement learning, vol. 135*. MIT Press. [JH]
- Tanner C. & Medin D. L. (2004) Protected values: No omission bias and no framing effects. *Psychonomic Bulletin and Review* 11(1):185–91. [KM]
- Tappolet C. (2010) Emotion, motivation and action: The case of fear. In: *Oxford handbook of philosophy of emotion*, ed. P. Goldie, pp. 325–45. Oxford University Press. [MA]
- Tappolet C. (2016) *Emotions, values, and agency*. Oxford University Press. [AK, CK]
- Thompson R. (2012) Whither the pre-conventional child? Toward a life-span moral development theory. *Child Development Perspectives* 6:423–29. [DN]
- Thomson J. J. (1985) The trolley problem. *Yale Law Journal* 94(6): 1395–415. [NC]
- Thomson K. S. & Oppenheimer D. M. (2016) Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making* 11(1): 99–113. [rJM]
- Tomasello M. (2016) *A natural history of human morality*. Harvard University Press. [JR]
- Toplak M. E., West R. F. & Stanovich K. E. (2011) The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition* 39:1275–89. [JFL]
- Trevathan W. R. (2011) *Human birth: An evolutionary perspective, 2nd ed.* Aldine de Gruyter. [DN]
- Turnbull C. M. (1984) *The human cycle*. Simon and Schuster. [DN]
- Tversky A. & Kahneman D. (1981) The framing of decisions and the psychology of choice. *Science* 211(4481):453–58. [arJM]
- Tyber J., Lieberman D., Kurzban R. & DeScioli P. (2013) Disgust: Evolved function and structure. *Psychological Review* 120:65–84. [CK]
- Uhlmann E. L., Pizarro D. A., Tannenbaum D. & Ditto P. H. (2009) The motivated use of moral principles. *Judgment and Decision Making* 4(6):476–91. [RZ]
- Valdesolo P. & DeSteno D. (2006) Manipulations of emotional context shape moral judgment. *Psychological Science* 17(6):476–47. [KM]
- Van Langen M. A., Wissink I. B., Van Vugt E. S., Van der Stouwe T. & Stams G. J. J. M. (2014) The relation between empathy and offending: A meta-analysis. *Aggression and Violent Behavior* 19(2):179–89. [WHBM]
- Van Vugt E., Gibbs J., Stams G. J., Bijleveld C., Hendriks J. & van der Laan P. (2011) Moral development and recidivism: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology* 55(8):1234–50. [WHBM]
- Vargas M. (2013a) *Building better beings: A theory of moral responsibility*. Oxford University Press. [RZ]
- Vargas M. (2013b) Situationism and moral responsibility. In: *Decomposing the will*, ed. A. Clark, J. Kiverstein & T. Vierkant, pp. 325–50. Oxford University Press. [JMD, aJM]
- Vavova K. (2014) Moral disagreement and moral skepticism. *Philosophical Perspectives* 28:302–33. [aJM]
- Vavova K. (2015) Evolutionary debunking of moral realism. *Philosophy Compass* 10(2):104–16. [aJM]
- Velleman J. D. (2009) *How we get along*. Cambridge University Press. [AJR]
- Vonasch A. J., Reynolds T., Winegard B. M. & Baumeister R. F. (2017) Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science* 9(5): 604–13. Available at: <https://doi.org/10.1177/1948550617720271>. [CJC]

- Waldmann M. R. & Dieterich J. H. (2007) Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science* 18:247–53. [KJH]
- Walker L. J., Frimer J. A. & Dunlop W. L. (2010) Varieties of moral personality: Beyond the banality of heroism. *Journal of Personality* 78(3):907–42. [WHBM]
- Walsh C. & Johnson-Laird P. (2004) Coreference and reasoning. *Memory and Cognition* 32:96–106. [MA]
- Warneken F. (2013) Young children proactively remedy unnoticed accidents. *Cognition* 126(1):101–108. [aJM]
- Warneken F. & Tomasello M. (2006) Altruistic helping in human infants and young chimpanzees. *Science* 311(5765):1301–303. [JR]
- Watson L. (2018) Educating for curiosity. In: *The moral psychology of curiosity*, ed. I. Inan, L. Watson, D. Whitcomb & S. Yigit, pp. 293–309. Rowman & Littlefield. [MA]
- WindEagle & RainbowHawk (2003) *Heart seeds: A message from the ancestors*. Beaver's Pond. [DN]
- Winnicott D. W., Winnicott C., Shepherd R. & Davis M. (1989) *Psycho-analytic explorations*. Harvard University Press. [DN]
- Woodward J. (2016) Emotion versus cognition in moral decision-making. In: *Moral brains: The neuroscience of ethics*, ed. S. Matthew Liao, pp. 87–117. Oxford University Press. [aJM]
- Wright D. (1982) Piaget's theory of moral development. In: *Jean Piaget: Consensus and controversy*, ed. S. Modgil & C. Modgil, pp. 207–17. Holt, Rinehart & Winston. [JIMC]
- Wringe B. (2015) The contents of perceptions and the contents of emotions. *Nous* 49:275–97. [CK]
- Wynne C. D. I. & Bolhuis J. J. (2008) Minding the gap: Why there is still no theory in comparative psychology. *Behavioral and Brain Sciences* 31:152–53. [KJH]
- Xu F. & Kushnir T. (2013) Infants are rational constructivist learners. *Current Directions in Psychological Science* 22(1): 28–32. [JR]
- Yang Q., Yan L., Luo J., Li A., Zhang Y., Tian X. & Zhang D. (2013) Temporal dynamics of disgust and morality. *PLOS ONE* 8:e65094. [CK, rJM]
- Young I. M. (1990) *Justice and the politics of difference*. Princeton University Press. [RZ]
- Young L. & Tsai L. (2013) When mental states matter, when they don't, and what that means for morality. *Social and personality psychology compass* 7(8):585–604. [aJM]
- Zamir E. & Medina B. (2010) *Law, economics, and morality*. Oxford University Press. [KJH]
- Zheng R. (2018) What is my role in changing the system? *Ethical Theory and Moral Practice* 21(4):869–85. [rJM]
- Zimmerman A. (2010) *Moral epistemology*. Routledge. [rJM]