Précis of *Regard for Reason in the Moral Mind*

Joshua May

Behavioral and Brain Sciences Vol. 42, No. e146 (2019): 1-60
(along with 21 commentaries and my replies to them)

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.

Keywords: moral psychology, moral judgment, moral motivation, virtue, debunking arguments, rationalization, rationalism, sentimentalism, moral skepticism

Word counts: 8,213 (main text); 1,755 (references); 10,131 (total, incl. abstract)

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 the blood, sweat, and grant money?

If you ask most philosophers and scientists working both within and outside the field of moral psychology, you'll likely hear something like the following. It turns out that Hume was right: emotions are the star of the show, while 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's not far off. Jonathan Haidt, for example, speaks of an "affect revolution" (2003) which apparently explains the "rationalist delusion" (2012) that reason plays a foundational role in moral cognition. Of course, such champions of sentimentalism don't themselves always conceive of this as pessimistic (compare e.g. Nichols 2004; D'Arms & Jacobson 2014), but it's 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're 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. Singer 2005; Greene 2013).

In *Regard for Reason*, 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—e.g. emotion, reason, arbitrary factors, evolutionary pressures. Next, I examine *normative* questions about the status of each element—e.g. 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 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's 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

a. 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 (Chapter 2), before going on to adduce evidence of moral inference (Chapter 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 while 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's 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 Chapter 2) in terms of our well-known susceptibility toward misattributing 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's doubtful that this is enough to attribute significant deficits in moral judgment to them. Moreover, it's often underappreciated that psychopaths exhibit not only "emotional" deficits but clearly rational or inferential ones. Psychopaths are notoriously irrational, particularly imprudent, due to 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 shouldn't 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 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 doesn't mean the rationalism/sentimentalism debate is confused or pointless. What we're

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 aren't 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's 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 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.

b. Moral Inference

Let's now turn briefly to the positive case for moral reasoning. The crucial move here (Chapter 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's 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's going on, but it's not as though we consciously go through all the steps—"Ah, they look to be living in the same dwelling, yet they're 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 needn't 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 vs. omissions, intentional vs. accidental actions, and harming as a means vs. 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. Cushman et al. 2006; Young & Tsoi 2013; Barrett et al. 2016).

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 over 100 studies involving over 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 isn't 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 hadn't 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 patterns of 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 isn't 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. Nussbaum 2004; Kelly 2011), framing effects (e.g. Sunstein 2005; Schwitzgebel & Cushman 2012), evolutionary pressures (e.g. Joyce 2006), and automatic emotional heuristics (e.g. Singer 2005; Greene 2014). All of these challenges are too ambitious for their own good, although more selective debunking arguments may succeed.

a. 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, since 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've 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 aren't 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 isn't 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's self-defeating (Vavova 2015). If we can't 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's briefly examine automatic emotional heuristics. Are our non-utilitarian commitments unwarranted because they're "sensitive to morally irrelevant things, such as the distinction between pushing with one's hands and hitting a switch" (Greene 2013: 328)? That's a fine normative premise, but the corresponding empirical premise is untenable. As Greene acknowledges, experiments demonstrate that our moral intuitions aren't particularly sensitive to pushing alone, but rather pushing that's done with intent or as a means to an end (Greene et al. 2009; Feltz & May 2017). 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 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" (2014: 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).

b. Selective Debunking & Moral Disagreement

Although wide-ranging empirical critiques of moral cognition are flawed, more selective debunking arguments can succeed (Chapter 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 isn't 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 while factory farming isn't, primarily because pets are cute. Similarly, although it's too wide-ranging to critique all non-utilitarian intuitions, we can all agree that it's 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'm usnure 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's 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: 304):

1. In the face of peer disagreement about a claim, one doesn't know that claim.
2. There's 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's 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's 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 shouldn't be overstated. Most people aren't 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 due to 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 isn't pessimistic.

4. Moral Motivation: Sources

Let's 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?

a. Egoism vs. 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's 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 thus 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 isn't 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; Johnston 2010; Flanagan 2017).

The problem with these proposals is that they can't 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 can't, 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, aren't uncommon.

b. Rationalization and Moral Integrity

Pessimists might accept the existence of genuine altruism but argue on empirical grounds that it's 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 isn't 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 don't just behave badly because it's in their interest. When they could just think (probably unconsciously) "I'm 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've done more than my fair share of good deeds this week," or even "I bet it's that sleazy banker's money, and he's 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 doesn't 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 it.

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's 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's right as such.

When we do what's right, then, we aren't 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's 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).

c. 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're only ultimately motivated to act because you happen to care about being respectful or about doing what's 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're antecedently motivated to do what's right as such ("de dicto") or to promote particular moral values, such as kindness, respect, and fairness ("de re"). But this needn't 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's 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's 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'll 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). Some philosophers believe these patients support the Humean thesis that moral (or otherwise normative) beliefs can't motivate by themselves (cf. Roskies 2003; Schroeder et al. 2010).

But that's based on a misunderstanding of the opposition. Of course one's moral beliefs don't *always* generate the corresponding desire, but when they do they needn't rely on an *antecedent* desire. Instead, the necessary element could be, say, a lack of full understanding—e.g. a patient believes she ought to thank the host, but she doesn't entirely

appreciate that she's 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 can't play this same role without either being desires or being an unnecessary additional posit in psychology.

However, I show that desires don't 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's 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's 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 don't have any empirical reason to *always* do so. Reason isn't destined to be a slave to the passions.

5. Moral Motivation: Status Update

So far in the book I argue that our moral beliefs aren't 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 aren't 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. Nelkin 2005; Vargas 2013; Doris 2015) 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.

a. 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's 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'd 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's 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 don't 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 don't 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* just is 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's 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 wouldn't have worried themselves about the morality of their choices.

The Defeater Dilemma is evident here. It's 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're not motivated by the wrong reasons if we're motivated to do what's right. There's no doubt that we're motivated by egoism as well, but we shouldn't overstate its power and prevalence, and likewise we shouldn't ignore the power and prevalence of moral integrity (even when it's due to motivated reasoning). The same dilemma arises for the challenge from situationism.

b. Situational Forces

Countless studies support the situationist thesis that we're 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 wasn't 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 aren't 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. Payne 2001; Bertrand & Mullainathan 2004). 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 2013). However, the more fundamental worry is that we're 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 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" (2005: 204). Thus, the situationist literature might seem to fund a wide-ranging critique of what motives 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's inappropriate to only help because you feel like it, but it's *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's 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 don't 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 can't 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 don't predict much behavior (Greenwald et al. 2009; Oswald et al. 2013; Forscher et al. 2017). Importantly, and I can't stress this enough, that doesn't mean implicit biases can't explain large-scale problems in society. Indeed, implicit biases may add up to explain the powerful discrimination any one individual experiences due to slights from many people, however well-meaning these people are. But the evidence to date doesn't suggest that *most* ordinary people base *many* of their morally relevant decisions *primarily* on their implicit biases. Some do, for sure, but we're 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's brandishing a toy gun, it's 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's 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'm right, moral psychology is in an important sense continuous with other domains of human psychology. The heart of the rationalist view is that morality isn't special; emotions aren't 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, isn't deeply flawed. A full defense of optimistic rationalism would require responding to challenges to reason itself. But that's for another day. *Regard for Reason* already discusses a wide range of literature in just ten chapters. It certainly hasn't 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 due to 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's 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 (Chapter 10). It's 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's 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.

References

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): 484–497.

Alfano, M. 2013. *Character as Moral Fiction*. Cambridge University Press.

Aquino, K., & Reed, A. 2002. "The Self-Importance of Moral Identity." *Journal of Personality and Social Psychology* 83(6): 1423–1440.

Ariely, D. 2012. *The Honest Truth About Dishonesty*. HarperCollins.

Arpaly, N. 2003. *Unprincipled Virtue*. Oxford University Press.

Arpaly, N., & Schroeder, T. (2014). *In Praise of Desire.* Oxford University Press.

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.

Barrett, H. C., Bolyanatz, A. et al. 2016. "Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment." *Proceedings of the National Academy of Sciences* 113 (17): 4688–4693.

Batson, C. D. 2011. *Altruism in Humans*. Oxford University Press.

Batson, C. D. 2016. *What's Wrong with Morality?* Oxford University Press.

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): 1335–1348.

Batson, C. D., Thompson, E. R., & Chen, H. 2002. "Moral Hypocrisy: Addressing Some Alternatives." *Journal of Personality and Social Psychology* 83(2): 330–339.

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.

Betancourt, H. (1990). An attribution-empathy model of helping behavior. *Personality and Social Psychology Bulletin* 16(3):573–91.

Blanken, I., van de Ven, N., & Zeelenberg, M. 2015. "A Meta-Analytic Review of Moral Licensing." *Personality and Social Psychology Bulletin* 41(4): 540–558.

Bloom, P. 2016. *Against Empathy: The Case for Rational Compassion*. New York: Ecco.

Campbell, R. & Kumar, V. 2012. "Moral Reasoning on the Ground." *Ethics* 122 (2):273-312.

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.

Cialdini, Robert B., S. L. Brown, B. P. Lewis, C. Luce, & S. L. Neuberg 1997. "Reinterpreting the Empathy- Altruism Relationship: When One Into One Equals Oneness" *Journal of Personality and Social Psychology* 73(3): 481-494.

Cushman, F., Young, L., and M. Hauser, 2006. "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm." *Psychological Science* 17(12): 1082–1089.

D'Arms, J. & Jacobson, D. 2014. "Sentimentalism and Scientism." In *Moral Psychology and Human Agency*, ed. by J. D'Arms & D. Jacobson. Oxford University Press.

Damasio, A. 1994/2005. *Descartes' Error*. Penguin Books. (Originally published by Putnam.)
Demaree-Cotton, J. 2016. "Do Framing Effects make Moral Intuitions Unreliable?" *Philosophical Psychology* 29(1): 1-22.
Doris, J. 2015. *Talking to Our Selves: Reflection, Ignorance, and Agency*. New York: Oxford University Press.
Feltz, A. & May, J. 2017. "The Means/Side-Effect Distinction in Moral Cognition: A Meta-Analysis." *Cognition* 166: 314–327.
Flanagan, O. 2017. *The Geography of Morals: Varieties of Moral Possibility*. Oxford University Press.
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.
Glenn, A. L., & Raine, A. 2014. *Psychopathy: An Introduction to Biological Findings and Their Implications*. New York University Press.
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* 47: 55–130.
Greene, J. 2013. *Moral Tribes*. Penguin Press.
Greene, J. D. 2014. "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics." *Ethics* 124(4): 695-726.
Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., and Cohen, J. D. 2009. "Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment." *Cognition* 111 (3): 364 –371.
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.
Haidt, J. 2003. "The Moral Emotions." In *Handbook of Affective Sciences*, ed. by R. J. Davidson, K. R. Scherer, & H. H. Goldsmith. Oxford University Press.
Haidt, J. 2012. *The Righteous Mind*. New York: Pantheon.
Henrich, J. 2015. *The Secret of Our Success*. Princeton University Press.
Holton, R. 2009. *Willing, Wanting, Waiting*. Oxford: Clarendon Press.
Huebner, B. 2015. "Do Emotions Play a Constitutive Role in Moral Cognition?" *Topoi* 34 (2): 427-440.
Hurka, T. 2014. "Many Faces of Virtue." *Philosophy and Phenomenological Research* 89(2): 496–503.
Johnston, M. 2010. *Surviving Death*. Princeton University Press.
Joyce, R. 2006. *The Evolution of Morality*, Cambridge, MA: MIT Press.
Kahane, G. 2011. "Evolutionary Debunking Arguments." *Noûs* 45(1):103-125.
Kelly, D. 2011. *Yuck!: The Nature and Moral Significance of Disgust*. MIT Press.
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.
Kühberger, A. 1998. "The Influence of Framing on Risky Decisions: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 75(1): 23–55.
Kumar, V. 2017. "Foul Behavior." *Philosophers' Imprint* 17(15): 1-16.
Kumar, V. & Campbell, R. 2012. "On the Normative Significance of Experimental Moral Psychology." *Philosophical Psychology* 25 (3): 311-330.
Kumar, V. & May, J. 2018. "How to Debunk Moral Beliefs." In *Methodology and Moral Philosophy*, ed. by J. Suikkanen & A. Kauppinen, Routledge.
Kunda, Z. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108(3): 480-98.

Landy, J. F. & Goodwin, G. P. 2015. "Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence." *Perspectives on Psychological Science* 10(4): 518-536.
Latané, B., & Nida, S. 1981. "Ten Years of Research on Group Size and Helping." *Psychological Bulletin*, 89(2): 308-324.
Maibom, H. L. 2005. "Moral Unreason: The Case of Psychopathy." *Mind and Language* 20(2): 237-57.
Mallon, R. & Nichols, S. 2010. "Rules." In *The Moral Psychology Handbook*, ed. J. M. Doris and The Moral Psychology Research Group, New York: Oxford University Press, 297–320.
Marsh, A. A., & Blair, R. J. R. 2008. "Deficits in Facial Affect Recognition among Antisocial Populations: A Meta-analysis." *Neuroscience & Biobehavioral Reviews* 32(3): 454–65.
May, J. 2011. "Egoism, Empathy, and Self-Other Merging." *Southern Journal of Philosophy* 49(S1): 25–39, Spindel Supplement: Empathy & Ethics, ed. R. Debes.
May, J. 2013a. "Because I Believe It's the Right Thing to Do." *Ethical Theory & Moral Practice* 16(4): 791–808.
May, J. 2013b. "Skeptical Hypotheses and Moral Skepticism." *Canadian Journal of Philosophy* 43(3): 341–359.
May, J. 2014. "Does Disgust Influence Moral Judgment?" *Australasian Journal of Philosophy* 92(1): 125–141.
May, J. 2018. *Regard for Reason in the Moral Mind*. Oxford University Press.
Mazar, N., Amir, O., & Ariely, D. 2008. "The Dishonesty of Honest People: A Theory of Self-concept Maintenance." *Journal of Marketing Research* 45(6): 633–644.
Mazar, N., & Zhong, C. B. 2010. "Do Green Products Make Us Better People?" *Psychological Science*, 21(4): 494–498.
McGrath, S. 2008. "Moral Disagreement and Moral Expertise." In *Oxford Studies in Metaethics*, Vol. 3, ed. R. Shafer-Landau. Oxford University Press.
Mikhail, J. 2011. *Elements of Moral Cognition*. Cambridge University Press.
Miller, C. B. 2013. *Moral Character: An Empirical Theory*. Oxford University Press.
Nelkin, D. K. 2005. "Freedom, Responsibility and the Challenge of Situationism." *Midwest Studies in Philosophy* 29(1): 181–206.
Nichols, S. 2002. "Norms with Feeling: Towards a Psychological Account of Moral Judgment." *Cognition* 84(2): 221–236.
Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.
Nichols, S. 2014. "Process Debunking and Ethics." *Ethics*, 124: 727-49.
Nussbaum, M. C. 2004. *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton, NJ: Princeton University Press.
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): 171–192.
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): 181–192.
Perry, J. 1979. "The Problem of the Essential Indexical." *Noûs* 13(1): 3–21.
Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford University Press.
Prinz, J. 2016. "Sentimentalism and the Moral Brain." In *Moral Brains: The Neuroscience of Morality*, ed. S. Matthew Liao. Oxford University Press.
Railton, P. 2017. "Moral Learning: Conceptual Foundations and Normative Relevance." *Cognition* 167: 172-190.
Roskies, A. 2003. "Are Ethical Judgments Intrinsically Motivational? Lessons From 'Acquired Sociopathy.'" *Philosophical Psychology* 16 (1): 51–66.

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.
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.
Schnall, S., Haidt, J., Clore, G. L., and A. H. Jordan 2008. "Disgust as Embodied Moral Judgment." *Personality and Social Psychology Bulletin* 34(8): 1096–1109.
Schroeder, T. 2004. *Three Faces of Desire*. New York: Oxford University Press.
Schroeder, T. Roskies, A. & Nichols, S. 2010. "Moral Motivation." In *The Moral Psychology Handbook*, ed. by J. Doris & The Moral Psychology Research Group, pp. 72-110. Oxford University Press.
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): 513-523.
Schwitzgebel, E., & Cushman, F. A. 2012. "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers." *Mind & Language* 27(2): 135–153.
Singer, P. 2005. "Ethics and Intuitions." *The Journal of Ethics* 9: 331–52.
Sinhababu, N. 2017. *Humean Nature: How Desire Explains Action, Thought, and Feeling*. Oxford University Press.
Sunstein, C. R. 2005. "Moral Heuristics." *Behavioral and Brain Sciences* 28(4): 531–42.
Tversky, A., & Kahneman, D. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–458.
Vargas, M. 2013. "Situationism and Moral Responsibility." In *Decomposing the Will*, ed. by A. Clark, J. Kiverstein, & T. Vierkant, pp. 325–50. Oxford University Press.
Vavova, K. 2014. "Moral Disagreement and Moral Skepticism." *Philosophical Perspectives* 28: 302–333.
Vavova, K. 2015. "Evolutionary Debunking of Moral Realism." *Philosophy Compass* 10(2): 104-116.
Warneken, F. 2013. "Young Children Proactively Remedy Unnoticed Accidents." *Cognition* 126(1): 101–108.
Woodward, J. 2016. "Emotion versus Cognition in Moral Decision-Making." In *Moral Brains: The Neuroscience of Ethics*, ed. by S. Matthew Liao. Oxford University Press.
Young, L. & Tsoi, 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.