

The Ethical Significance of Cognitive Science

Victor Kumar
Assistant Professor
Philosophy Department
Boston University

The answers to many philosophical questions depend on contingent facts about the human mind. This is especially clear in philosophy of mind and adjoining philosophical neighborhoods. Thus, philosophers draw on empirical research in cognitive science to understand the relationship between the mind and the brain, the nature of reason and emotion, the ability to know the contents of other minds.

Yet, how could cognitive science be relevant to ethics, of all things? Ethics, we are commonly told, is normative rather than purely descriptive. Since empirical research in cognitive science answers only descriptive questions, some philosophers may protest that it cannot yield answers to normative questions in ethics. There is, in short, a well-known gap between “is” and “ought” and no amount of new scientific knowledge about what *is* can bridge the gap that cordons off philosophical knowledge about what *ought*. It would seem, then, that even if its significance for other domains of philosophical inquiry is secure, cognitive science has no ethical significance.

This argument—against the possibility of ethics informed by cognitive science—fails on multiple fronts. And to see this one need not experience any doubt about the existence of the is-ought gap. First of all, many topics in *metaethics* are descriptive rather than normative. Though we wish to know the extent of our obligations to others, we also wish to know what it is to form a moral judgment about an issue like this. First order questions about the good and the right are normative, certainly, but second-order questions about the cognitive status of moral judgment and its link to motivation are plainly descriptive. Thus, even if cognitive science were forever closed off to first order, normative inquiry, it might still provide an important source of evidence for second order, descriptive inquiry in metaethics about the nature of moral judgment.

The ability of cognitive science to help answer second order questions about moral judgment suggests a way in which the field can inform first order inquiry too. Many of our moral beliefs are shaped by unconscious psychological processes. Some of these processes may be unsuited to confer justification. So, if empirical findings in cognitive science uncover the unconscious influences on certain moral beliefs, and if it turns out that these influences are epistemically disreputable, then the findings have a normative upshot: they *debunk* those moral beliefs. Even if cognitive science cannot reveal what we owe to others, it can tell us that moral beliefs we hold about first order issues like this are unjustified and should be abandoned.

Cognitive science is relevant to second order questions in metaethics about the nature of moral judgment and the epistemic evaluation of moral beliefs. In the latter case metaethical conclusions have normative implications. But cognitive science informs ethical theory in a third way too, and here it has more direct normative significance. What we ought to do and who we should aspire to be are constrained by psychological feasibility. The significance our psychological makeup lies in its import for *non-ideal theory*. In normative ethics, non-ideal theory eschews universal criteria of right and wrong and instead seeks more modest, but more useful, generalizations about moral improvement from our present circumstances. By revealing what is psychologically feasible, cognitive science can shed light on the very nature of moral progress.

My task in this chapter is to make good on the three promissory notes issued above. That is, I will show how cognitive science informs three topics of inquiry in ethics—the nature of moral judgment, debunking arguments, and non-ideal theory. My focus in this chapter is on metaethics and normative ethics, and thus I will set aside inquiry that fits more comfortably in the study of free will, moral responsibility, action theory, and applied ethics. In each section I will lay out the relevant philosophical subjects, explain how cognitive science can help address them, and then illustrate by discussing recent work in naturalistic ethics.

1. Moral Judgment

Some areas of metaethics have few obvious connections to cognitive science. The metaphysics of morality is, at best, only tangentially related to the science of moral cognition. However, it is far more plausible that cognitive science offers resources to that branch of metaethics that is concerned with the nature of moral judgment—where “moral judgment” is construed as a psychological state rather than a linguistic utterance. Philosophers working in this area of metaethics are interested primarily in two related topics: (1) whether moral judgment is a cognitive or non-cognitive psychological state and (2) whether moral judgment has a necessary or contingent link with motivation.

Some philosophers explicitly conceive of their target of study as the concept of moral judgment (Smith 1994; Jackson 1998). Others voice no such explicit commitment, but employ an *a priori* methodology that seems best suited to probe the concept. Yet, another possibility is that the target of study is not the concept of moral judgment, but moral judgment *itself*. Is this, however, a coherent thought? Is there any way to get a grip on moral judgment except as mediated by our concept? The answer, I believe, is yes.

Moral judgment is of interest because we are confronted by numerous examples of the phenomenon in everyday life. We make moral judgments about others and about ourselves, as do other people with whom we interact. Paradigm cases of moral judgment thus provide an alternative way of grasping the subject matter: we may study moral judgment itself by studying paradigm cases (cf. Kornblith 2002). And if we want to know what the paradigm cases have uniquely in common, the best way is via empirical study. Thus, if

the target of study is moral judgment itself, we need not begin by analyzing our concept of moral judgment. The concept of moral judgment may inaccurately represent the real world phenomenon.

Metaethical naturalists need not start from scratch, methodologically, and can take a cue from decades of work in philosophy of mind (Fodor 1981, 1987; Griffiths 1997; Schroder 2004; Prinz 2004; Holton 2009; Weiskopf 2009). Traditionally, philosophers who study mental state categories have been interested in the concept of belief, or the concept of desire, etc. But nearly since the birth of cognitive science, it's been understood that ordinary concepts of mental state categories might be radically mistaken—hence the extensive debate about eliminativism about folk psychology (Churchland 1981; Stich 1983, 1996). If the topic of study is mental states themselves, what we should do is empirically study them. To take this approach is to treat mental state categories (corrigibly) as *natural psychological kinds* and to adopt the methodological approach suited to investigation of putative natural kinds (Kumar 2016-a). As philosophers of language suggest, a natural kind term refers directly to the corresponding kind, usually via ostension to paradigm cases, even if the intensional concept associated with a natural kind term picks out a different extension, even if it denotes the empty set (Putnam 1975; Kripke 1980; see also Kumar 2014).

Like other mental states, moral judgment has two ingredients: attitude and content (see Kumar 2015). The bulk of attention in metaethics has been accorded to the attitude that is constitutive of moral judgment, rather than its content. The default view, perhaps, is that moral judgments are ordinary beliefs. To judge that someone's disrespectful behavior is wrong is simply to believe that it is wrong. Historically, this cognitivist view of moral judgment came under attack for both psychological and metaphysical reasons. On the one hand, moral judgments, unlike beliefs, seem to have an essential tie to motivation. Someone who says that it is wrong to cheat at cards, but goes on to do so without the slightest hesitation, seems not to be speaking sincerely (Smith 1994). On the other hand, if moral judgments are beliefs then they must have truth-conditions, and the existence of moral properties has seemed dubious, as has our ability to gain epistemic access to them (Mackie 1977). Better to save the legitimacy of moral thought—to avoid the threat of massive and widespread error—by understanding moral judgments as non-cognitive attitudes with no ontological purport.

A number of different non-cognitivist proposals have been put forward. The earliest view characterized moral judgments as emotions or states of approval and disapproval (Stevenson 1937; Ayer 1952). Boo to cheating at cards. A more recent proposal casts moral judgments as non-cognitive attitudes of acceptance toward a norm (Gibbard 1990). To judge that cheating at cards is wrong is, on this view, to accept a norm that forbids or discourages cheating at cards. Non-cognitivist views such as these are able to secure an essential link between moral judgment and motivation, and able too to avoid the embarrassing result that moral thought is perpetually and irredeemably in error.

Cognitivism, however, remains an attractive position. First of all, moral judgments seem to be truth-apt, and arguably only cognitive attitudes can be true or false. Relatedly, non-cognitivists have a difficult time accounting for the validity of inferences in which one of the premises is a moral judgment. This is especially true in embedded contexts, where a moral proposition is embedded in a logically complex construction (Geach 1965; Schroeder 2008). Furthermore, even though moral properties are metaphysically and epistemically puzzling, it still seems as if we have moral knowledge, halting and fallible though it certainly is (Campbell 2007). For instance, we know that Jim Crow laws in the U.S. were morally wrong. Of course, knowledge seems to be a cognitive mental state. What's more, much progress has been made in understanding the existence of moral properties and the possibility of moral knowledge, both in naturalistic and non-naturalistic terms (see, e.g., Railton 1986; Shafer-Landau 2003). All of these considerations buttress cognitivist theories of moral judgment.

The philosophical literature on moral judgment is vast and the foregoing review of the cognitivism/non-cognitivism debate elides many complications, many of which involve each side attempting to accommodate considerations that have been thought to support the other side. Still, it is enough to gain an entry point. From the naturalistic perspective outlined above, if we want to know how psychological considerations—rather than, say, metaphysical considerations—bear on the debate, what we should do is set aside thought experiments and instead examine directly whether paradigm cases of moral judgment behave like cognitive or like non-cognitive states.

A further methodological hurdle must be cleared if empirical evidence can be infused into the cognitivism/non-cognitivism debate. The mind is complex, to say the least. How are we to individuate moral judgment from other mental states that are concomitants of it? For example, a wealth of evidence suggests that emotions are critical elements of moral cognition. But this does not show, by itself, that moral judgment itself consists—partly or wholly—in emotion. For all this empirical evidence says, emotions may be related to moral judgments only causally, rather than constitutively.

The solution to this problem is to examine not just what is present in paradigm cases of moral judgment, but also the *causal/explanatory role* of moral judgment in psychological generalizations (Kumar 2016-a). What we want to know, then, is what sort of causal role paradigm cases of moral judgment regularly play in our psychology and whether cognitive or non-cognitive states underwrite that role. To broaden our attention in this way is to adhere more closely the methodology reflected in scientific study of other natural kinds. Water is H₂O, and not some more complex formula that includes common solubles found in paradigm cases, because it is H₂O alone that underwrites the chemical explanations in which water figures.

Cognitive science informs metaethical investigation of moral judgment when the target of study is moral judgment itself, rather than the concept, because cognitive science facilitates empirical study of paradigm cases of moral judgment and their caus-

al/explanatory role. Jesse Prinz (2007) and Victor Kumar (2015, 2016-a) are two philosophers who pursue this methodology.

Prinz argues for a view that he calls “emotionism,” not to be confused with “emotivism,” the non-cognitivist view that moral judgments are emotions. Prinz claims, rather, that moral judgments are beliefs. However, he argues that the moral concepts that are constituents of moral judgments are composed of emotions. More precisely, to have a moral concept is to have a disposition to undergo certain types of emotional episodes. To conceptualize an action as morally wrong is, roughly, to be disposed to experience guilt if one performs it and resentment if others do. Thus, Prinz’s view is cognitivist but also says that emotions are constitutive of moral judgment.

In support of emotionism, Prinz (2007: 21-3) cites empirical evidence that paradigm cases of moral judgment are invariably accompanied by emotion. However, as he makes clear, this evidence might show only that emotions are necessary causes of moral judgments, rather than constitutive of them (23, 29). The clincher, according to Prinz, is evidence that emotions are not only necessary but also sufficient for moral judgment. In one study, participants judge quite innocent actions to be morally wrong when they are hypnotized to experience negative emotions in response to morally neutral words (Wheatley and Haidt 2005). In a study of “moral dumbfounding,” participants maintain moral judgments about emotionally evocative activities even when they can cite no reasons, suggesting that emotions are sufficient for moral judgment in the absence of states that might otherwise be proposed as constituents (Murphy et al. 2000). A number of criticisms might be raised against this argument, but the most pertinent is that it does not seem to adhere to the naturalistic standards outlined above and that Prinz himself accepts. Emotions might be necessary and sufficient causes of moral judgment, for all the cited empirical evidence says, and yet they may generate ordinary beliefs. The evidence Prinz cites does not distinguish cause from constitution.

Kumar (2016-a) argues for a “hybrid” theory of moral judgment. This is a theory that marries cognitivism with non-cognitivism. According to Kumar, moral judgment consists in a moral belief and a separable state of moral emotion (see also Campbell 2007). To judge that someone acted wrongly is to believe that she acted wrongly and also feel, say, resentment toward her. Typical moral judgments have both elements, but atypical cases may have only one or the other.

Kumar begins his case for a hybrid theory by considering two types of psychological explanations in which moral judgment figures. First, moral judgment explains reasoning not just about moral matters but also in many other domains. For example, moral judgments seem to influence mental state ascriptions in the Knobe effect (Knobe 2010). Second, moral judgments also explain behavior—in particular, cooperative, uncooperative, and punitive behavior. Thus, in economic games, people make moral judgments about others and this influences whether they are likely to cooperate with fellow participants in prisoners’ dilemmas, whether they punish fellow participants in the ultimatum game

(Fehr and Fischbacher 2004). So, collectively, this evidence suggests that moral judgment plays the causal/explanatory roles of cognitive states *and* non-cognitive states.

Kumar argues next that the non-cognitive state that is partly constitutive of moral judgment is moral emotion. His main source of evidence is a large body of work which suggests that two types of psychological processing are at play in moral cognition, one generating beliefs and the other generating emotions (for review and discussion see Haidt 2001; Greene 2008; Campbell and Kumar 2012). Kumar also cites evidence suggesting that these two processes are tightly integrated in moral cognition (see esp. Rozin 1999), and thus that belief and emotion are bound together in a causal unity—a homeostatic property cluster (Boyd 1991). Kumar seems to adhere more closely than Prinz to the naturalistic methodology suited to investigation of natural kinds. The principal challenge for his hybrid theory, arguably, is whether other, competing accounts of moral judgment might also have the ability to underwrite its causal/explanatory role. This remains an open, empirical question.

I have focused on naturalistic investigation of the attitude that is constitutive of moral judgment. But another central question about moral judgment also seems capable of benefiting from a naturalized methodology. Is the link between moral judgment and motivation necessary or contingent? Prinz and Kumar's theories suggest similar answers to this question. For Prinz, a disposition to experience motivating emotions is necessary, but this disposition is *ceteris paribus* and may be defeated. For Kumar, motivating emotions are present in typical, hybrid cases of moral judgment, but in some atypical cases only moral belief is present without any corresponding emotion or motivation. Thus, both views entail that motivation is linked constitutively with moral judgment, without being strictly necessary (see also Kumar 2016-b).

Application in metaethics of the naturalistic methodology so common in philosophy of mind has only begun, but it seems to have promise. We may better understand what moral judgment is—cognitive or non-cognitive, necessarily or contingently motivational—if we draw on research in cognitive science that uncovers what is common to paradigm cases of moral judgment and supports their causal/explanatory role.

2. Debunking Arguments

If cognitive science can offer insight into the nature of moral judgment, the central methodological barrier it must cross is to successfully apply empirical criteria that distinguish cause from constitution. However, the causes of moral judgment are, indeed, philosophically significant in their own right. A leading view in epistemology is that the causes of a belief—both what initiate and sustain it—determine its epistemic status. That is, a belief is justified or warranted insofar as it is caused by epistemically reputable processes. It would seem, then, that research in cognitive science on the causes of moral beliefs might help us epistemically evaluate them.

But what precisely counts as an epistemically reputable process for arriving at moral beliefs? Unfortunately, this question is fraught by disagreement at the level of first order ethics. Utilitarianism suggests that a moral belief is justified if it is based on processes that are sensitive to aggregate happiness or well-being. Deontology suggests that a moral belief is justified if it is based on processes that apply legitimate principles about rights and duties. Naturally, other first order ethical theories imply other, competing views about the processes sufficient to confer justification on moral beliefs. The trouble, then, is that even if cognitive science isolates the processes influencing moral beliefs, there seems to be no uncontroversial way to infer that they are epistemically reputable.

Cognitive science may not be able to endow epistemic credit upon our moral beliefs. However, it *can* discredit them. Utilitarians and deontologists do not agree about whether a belief forming process attuned to happiness or well-being confers justification. But they must certainly agree that a wide range of other psychological processes are *epistemically defective*. For example, suppose we were to discover that certain moral beliefs are the product of wishful thinking or logical fallacies. In that case, no matter which first order ethical theory is true, these beliefs would be *debunked*. Herein lies the possibility of ethical debunking arguments that rely on research in cognitive science (Nichols 2014). If any such arguments are sound, we should conclude that the targeted moral beliefs are unjustified, that we should abandon them, and that any moral beliefs conflicting with them are now on firmer ground (see Kumar and May forthcoming).

One source of skepticism about ethics informed by cognitive science is the existence of the is-ought gap. As we have seen, the nature of moral judgment is a descriptive topic, and thus metaethical naturalists who study it need not mind the is-ought gap. However, conclusions about the epistemic status of our moral beliefs are patently normative. Does this spell trouble? Not at all. Psychological debunking arguments in ethics require empirical premises derived from research on the causes of moral beliefs. But they also require one or more normative premises, i.e., that some psychological process is epistemically disreputable. Debunking arguments are effective only when the normative premises upon which they rely are more plausible than the moral beliefs they attempt to debunk (Kumar and Campbell 2012). Moreover, then, debunking arguments explicitly depend on normative assumptions and do not even attempt to jump the is-ought gap. As we shall see over the course of this section, the critical challenge for debunking arguments is to articulate empirical premises that can bear the load placed upon them.

The most well known type of debunking argument in ethics draws on evolutionary psychology (Ruse 1986; Joyce 2006; Street 2006; Rosenberg 2011). According to many scientists, human morality is fundamentally a biological and cultural adaptation for social creatures that faced a host of ecological challenges requiring cooperation and coordination. Philosophers argue that if morality is designed by natural selection, then it is unlikely to have hit upon objective moral truths. Evolved moral beliefs are constrained only by expediency, not by accuracy.

Evolutionary debunking arguments face a number of challenges. Most pressing, the required empirical premises are highly speculative. Thus, many evolutionary debunkers are eager to disavow any categorical commitment to evolutionary accounts of morality and resort to conclusions that are conditional in form: *if* morality is an adaptation, *then* moral beliefs are unjustified. Another problem is that it is not at all obvious that natural selection is an epistemically disreputable process (Kumar and May forthcoming). To make this claim stick one needs to assume controversial premises in first order ethics. For if, say, what grounds morality is desire-satisfaction (consequentialism) or social interaction upon reasonable terms (contractualism), then it may well be that natural selection has been, as a contingent matter, sensitive to these grounds.

Much research in cognitive science is experimental, and some of it targets the psychological processes underlying moral cognition. This program of research is maturing and continues to face methodological challenges. Nonetheless, it is on firmer evidential ground than evolutionary psychology. Furthermore, it is uncontroversial that certain psychological premises are epistemically disreputable. Thus, it seems, research on the more proximal—rather than ultimate—causes of moral beliefs provides a more credible source of evidence upon which to found empirical debunking arguments.

To better understand how experimental research can debunk moral beliefs, and to understand too the challenges that debunking arguments face, it will help to examine them more concretely. Notice, first, that these arguments fall on a spectrum from global to local, sweeping to selective. Some philosophers draw on cognitive science in an attempt to debunk all moral beliefs: we should withhold judgment not about some moral issues but all of them. Other philosophers attempt to debunk very specific moral beliefs, like, say, the doctrine of double effect. In what follows, I will illustrate how cognitive science can support debunking arguments in ethics, but I will also argue that these arguments can incorporate plausible empirical premises only if they are very selective (see Kumar and May forthcoming for more detail).

Arguably, nearly all of our moral beliefs are based on intuitions about concrete cases. We believe that killing is wrong, in general, because particular instances of killing feel intuitively wrong. Walter Sinnott-Armstrong (2008) argues, however, that moral intuitions are distorted by cognitive biases. Sinnott-Armstrong cites evidence on framing effects: people who are presented with a moral dilemma make different judgments based on how the dilemma is framed. For example, intuitions are biased by the order in which dilemmas presented, or by whether the dilemma highlights either lives lost or lives saved (e.g., Petrinovich and O'Neill 1996). It is entirely uncontroversial that framing effects are an epistemically disreputable process. One should accept this premise no matter whether one subscribes to utilitarianism, deontology, or some other ethical theory. And so, if moral intuitions are based on framing effects we have reason to abandon any moral beliefs nourished by intuition—which might include all of them.

This argument is ambitious, indeed too ambitious for its own good. As Kumar and Joshua May (forthcoming) point out, framing effects account for only a small fraction of the influences on moral intuition (Young and Tsoi 2013). If a more complete set of causes were offered, it would not be clear that all intuitive moral beliefs are based on disreputable processes. So, the empirical burden of this debunking argument has not been met, at least not if the aim is to debunk all intuitive moral beliefs. (To be fair, Sinnott-Armstrong's principal aim is not to debunk all moral beliefs, but to criticize an intuitionist moral epistemology.)

Joshua Greene (2008, 2014) develops a debunking argument with slightly narrower scope. His target is characteristically deontological moral intuitions, i.e., those associated with deontology. Greene argues that characteristically deontological intuitions are produced by heuristics and biases within a fast, automatic, unconscious psychological system. Furthermore, he argues that these heuristics and biases often lead us astray. Greene himself has conducted studies which suggest that these intuitions can be influenced by morally irrelevant factors, like whether harm is inflicted by direct force rather than by remotely pressing a button. According to Greene, deontology is based on intuitions; these intuitions are unjustified; so too, then, is deontology. What's more, deontology conflicts with utilitarianism, and since Greene thinks we have reasons to distrust deontology, utilitarianism gains an edge.

Greene's debunking argument does not beg the question against deontology. All ethical theorists, deontologists included, should agree that whether harm is inflicted nearby or remotely is morally irrelevant; a belief forming process that is sensitive to this difference does not confer justification. Nonetheless, Greene's debunking argument against deontology fails because his empirical premises do not stand up to scrutiny. It may be true that characteristically deontological intuitions are influenced by certain morally irrelevant factors. But these intuitions are also sensitive to a wide range of other factors that seem to be morally relevant: the intensity of the harm, whether it was caused intentionally, whether there are mitigating circumstances, etc. (see Kumar and May forthcoming) Some of these factors may be morally irrelevant from the perspective of utilitarianism, but Greene cannot rely on a normative premise to this effect without introducing assumptions that are more controversial than the moral beliefs he targets (cf. Berker 2009). Furthermore, new evidence in moral learning theory suggests that the fast, automatic, unconscious system underlying moral intuition can be flexibly attuned to changing material and social conditions (Railton 2014; Kumar forthcoming-a). Once we understand more fully the causes of the relevant intuitions, then, it no longer seems as if they are epistemically disreputable.

Daniel Kelly (2011) also develops an ethical debunking argument and, like Sinnott-Armstrong and Greene, Kelly sets his sights on intuitive moral beliefs. However, Kelly's target is still narrower: all moral beliefs based on disgust. The consensus among cognitive scientists is that disgust is fundamentally an emotion designed to help us avoid the threat

of disease and infection (Rozin et al. 2008; Tybur et al. 2013). Only later, as Kelly explains, was disgust “co-opted” in moral cognition. Nonetheless, by now many moral beliefs are based on feelings of disgust. This includes conservative repugnance toward homosexuality but also liberal abhorrence of GMO food products. Yet other moral beliefs, besides these, seem to be tied to disgust. Cheating, deception, and exploitation, for example, often elicit moral disgust (see Kumar forthcoming-b). The problem, according to Kelly, is that disgust is an unreliable psychological mechanism. It evolved to “play it safe rather than sorry.” That is, it was evolutionarily advantageous for disgust to be oversensitive: better to be disgusted by innocuous objects than to fail to be disgusted by harmful objects. So, Kelly concludes, disgust is unreliable and should not be trusted in morality.

Kelly claims that disgust influences a range of moral beliefs. However, although available empirical evidence suggests that many types of moral violations elicit disgust (see Chapman and Anderson 2013), the evidence also suggests that, by itself, disgust has little influence upon moral beliefs. May (2014) argues that the effects of disgust on moral belief are insubstantial and infrequent. A recent meta-analysis by Justin Landy and Geoff Goodwin (forthcoming) confirms the insubstantial effect of disgust on moral belief. Perhaps if disgust did substantially influence certain moral beliefs we would have reason to abandon them (cf. Kumar forthcoming-b), but, as it stands, the empirical premise in Kelly’s debunking argument does not seem to be supported by available evidence.

Sinnott-Armstrong attempts to debunk intuitive moral beliefs by appeal to framing effects, Greene by appeal to simple heuristics and biases, Kelly by appeal to disgust. However, these arguments are unsound because they rest on dubious empirical premises. None of the aforementioned processes seem to exert a substantial influence on moral beliefs. The likely problem is that the target in each case is too broad. It is unlikely that cognitive science will provide a simple story about the influences on a broad and heterogeneous class of moral beliefs, such that the influences are epistemically defective. Perhaps, then, were we to focus on more narrow classes of moral beliefs, we would have some hope of finding disreputable influences.

Kumar and Richmond Campbell (2012) offer a schema for selective debunking arguments (see also Kumar and May forthcoming). The idea is to focus, narrowly, on a pair of divergent moral beliefs, and to investigate empirically why they diverge. Imagine two similar cases. We believe that action in one case is wrong, but that action in the other case is right. Imagine further that empirical research is deployed to study why we respond differently to the two cases—something for which an experimental approach is well suited (Kumar and May forthcoming). If it turns out that the *causally* relevant difference between the two cases is *morally* irrelevant, then it seems that our different beliefs about the cases are unjustified. We should abandon one or the other.

Let’s consider a simple illustration. Judges are tasked with making parole decisions about prisoners and one of the constraints they should follow is to treat like cases alike. If John deserves parole, and Bill’s case is not relevantly different then it is wrong not to of-

fer parole to Bill. A recent study by Shai Danziger et al. (2011) finds that some judges do not treat like cases alike, and that their parole decisions are, unsettlingly, influenced by whether or not they are hungry. Prisoners who face the judges before lunch are much less likely to be granted parole than prisoners who face the judges after lunch. Clearly, whether or not the judges are hungry is an epistemically disreputable basis for beliefs about merited parole. So, this empirical evidence suggests that one set of beliefs must be abandoned—though by itself it does not tell us which. Either the late morning prisoners deserve a higher rate of parole, or the early afternoon prisoners deserve a lower rate of parole (see Kumar and May forthcoming for further discussion).

No doubt there are other factors influencing the judges beliefs about parole, but because we have focused narrowly on the different beliefs they have before and after lunch, the empirical premises in this debunking argument are plausible. Selective debunking arguments with a similar structure have recently been developed that target our reluctance to donate to charity (Kumar and Campbell 2012), and the doctrine of double effect (May 2015).

Debunking arguments in ethics can rely on cognitive science to supply empirical premises about the causes of moral belief. The principal challenge for debunking arguments is to identify credible empirical premises. These premises, I have suggested, will be more plausible if the arguments are highly selective—if they target narrow pairs of moral beliefs rather than large classes of beliefs. Of course, cognitive science cannot yield philosophical conclusions on its own. Philosophical work is needed to articulate plausible normative premises that do not beg important questions in first order normative ethics. Still, empirical work in cognitive science that probes the unconscious influences on moral beliefs might lead us to discover that some cherished moral beliefs are based on epistemically disreputable grounds. Any such moral beliefs are unjustified and we should abandon them in the absence of better grounds. This is how cognitive science has not just metaethical but also normative significance.

3. Non-Ideal Theory

So far, we've looked at two ways in which cognitive science can inform metaethical branches of moral philosophy. First, philosophers interested in moral judgment can understand its nature by attending to research on moral judgment and its causal/explanatory role. Second, philosophers interested in debunking moral beliefs can look to research on epistemically defective psychological processes, especially processes underlying certain narrow pairs of moral beliefs.

In the rest of this chapter we'll explore a way in which cognitive science informs *normative ethics*. As we have seen, if cognitive science is to furnish empirical premises that lead to normative conclusions, philosophy must do its part by supplying plausible normative premises. The critical normative premise that supports normative ethics of the sort developed here is a variant of *ought-implies-can*. Suppose you can fulfill a duty to a

friend by performing one of two actions. If one action is more feasible, few would deny that you have more reason to perform it. Or suppose you can cultivate a virtue by one of two means. If one is more feasible, again it seems obvious that you have reason to take that means over the other.

Some of the most striking research in cognitive science over the past few decades examines the implicit biases that underlie certain forms of discriminatory behavior (see Brownstein 2015 for review). This work is of obvious ethical importance in light of the severe negative consequences that implicit biases have on women, people of color, and other vulnerable and marginalized groups. Of particular importance is research that investigates how to reduce the influence of implicit biases (Lai et al. 2014). This research offers a guide about how best to fulfill our duty to eliminate or reduce unjust discrimination. Obviously, we should pursue the means that cognitive science suggests are most effective.

This example makes clear that research in cognitive science on psychological feasibility has the potential to inform ethical reasoning. But my concern in this chapter is with ethical *theory*, and so far it is not at all clear that research on things like implicit biases has any theoretical significance in ethics. The following picture seems to capture the division of labor between philosophy and science. A priori philosophy supplies theoretical claims about what we owe to others and what sorts of character traits are virtuous. Science, it seems, only offers guidance in the *application* of these theoretical claims—no guidance in how to articulate theoretical claims in the first place.

I believe that cognitive science can also inform theory construction in normative ethics. But how? To make headway, we must lay out in detail a form of *non-ideal normative ethics*, beginning with the recent history of non-ideal theory in political philosophy.

A distinction between ideal and non-ideal theory finds its original formulation in the work of John Rawls (1971), who employs the distinction in relation to theories of justice. The aim of ideal theory is to understand the principles that structure an ideally just state, abstracting away from current injustices that our societies face. It doesn't matter whether people as they currently are would comply with these principles. For Rawls, we want to know what principles would structure an ideal society *if* people were to comply with them. The aim of non-ideal theory, by contrast, is to examine the current state of our societies, their most pressing injustices, and the series of steps that will lead incrementally to an ideally just state. As is evident, ideal theory is prior to non-ideal theory within Rawls' framework. We first must understand what an ideally just state looks like, and only then can we take the steps needed to appropriately address current injustices.

A very different understanding of non-ideal theory is found in the work of Amartya Sen (2009), Elizabeth Anderson (2010), and a number of other political philosophers. This view rejects the need for ideal theory at all. It seems unlikely that a single brilliant philosopher, or even a community of brilliant philosophers working together, will be able to understand what a perfectly just society looks like. Moreover, non-ideal theorists argue

that we don't need to know what the ideal is in order to see what sorts of improvements are possible, what sorts of changes would enhance justice in our societies. The *theory* in this form of non-ideal theory consists in generalizations about progressive moral change. And the method for theory construction is broadly empiricist. What we should do is search for changes to current practices and to those in the recent past that seem to be progressive, and then draw general lessons about the nature of progressive change. The unit of analysis is not the ideally just state, nor the steps that lead to it, but types of changes that tend to lead to moral progress. For non-ideal theorists like Sen and Anderson, this is the most we can have and, fortunately, it is all that we need.

Non-ideal theory is typically pursued within political philosophy, but there seems to be a clear analogue in normative ethics. The aim of normative ethics, traditionally conceived, is to articulate standards on right action that provide a more-or-less unifying explanation of what it is that right action consists in. Utilitarianism and deontology are classical theories in “ideal ethics.” According to utilitarianism, an action is right because it maximizes aggregate happiness or well-being. According to Kantian deontology, an action is right because it follows from a rule that can be consistently universalized. Though competitors, each theory offers a universal account of what makes actions morally right.

Non-ideal ethics rejects not just the answers provided by utilitarianism and deontology, but also the very questions posed. It seems unlikely that a community of brilliant philosophers will be able to arrive at a grand unifying theory that explains what makes actions right in every conceivable circumstance. In fact, particular moral insights are hard won by communities of diverse people facing moral problems together, not in the abstract but on the ground (Anderson 2014, forthcoming). In addition, we don't need to have an ideal theory that specifies what one should do in every circumstance to make changes to our behavior, our character, and our institutions that are morally progressive. The aim of non-ideal ethics is not to provide a unifying explanation of right action. Rather, it is to draw generalizations about the sorts of changes that tend to lead to moral progress. These generalizations will be local to the sorts of material and social conditions that people face now and in the recent past, but for precisely that reason non-ideal theorists can more reasonably hope to arrive at justified conclusions, and to develop theories about progressive moral change that provide genuinely useful moral guidance.

Non-ideal ethics, unlike non-ideal political philosophy, is in very early stages of development. Extended criticisms and defenses of the project, as such, have yet to be articulated. Nonetheless, concrete contributions to non-ideal ethics have begun to emerge, and to get a clearer sense of the project as a whole it will help to examine some of these contributions. I will not attempt to unite various pieces of work in non-ideal ethics under a single banner, eliding, for example, important contributions in feminist philosophy and philosophy of race. My interest here is specifically with work in non-ideal ethics that draws on cognitive science. I will limit myself, in the rest of this section, to non-ideal approaches to virtue and to moral reasoning.

John Doris (1998, 2002) has mounted an empirical critique of traditional Aristotelian virtue ethics (see also Harman 1999). According to Doris, Aristotelian virtue ethics posits the existence of “robust” character traits, that is, traits undergirding behavioral dispositions that exhibit significant consistency across situational contexts. Bravery, for example, undergirds a disposition to face one’s foes in spite of fear, whether one’s foes are on the battlefield, in the boardroom, or in the press. Doris argues, however, that a wealth of research in social and personality psychology over several decades shows that people exhibit a remarkable lack of cross-situational consistency. To take just one example from a large body of research, psychologists find that whether or not people are willing to help another person in need often turns simply on whether they have recently found a coin in a telephone booth (Isen and Levin 1972). The “situationist” critique of Aristotelian virtue ethics is, perhaps, best understood as a critique of a certain type of *ideal* virtue ethics. Aristotelianism articulates an account of moral character that, Doris argues, is not in fact instantiated in human beings and may be out of reach.

Peter Railton (2011), drawing on a similar body of evidence in social and personality psychology, offers a more constructive contribution to virtue ethics. He suggests that if we are to develop a psychologically adequate account of virtue, we should begin by examining the sorts of character traits that are commonly instantiated in human beings. First of all, agreeing with Doris and other situationists, we should turn away from robust character traits and instead focus on traits that are narrowly attuned to situational contexts. For example, Railton suggests that “justice” is not in fact a real character trait, and that some people have a strong capacity for institutional justice but a weak capacity for interpersonal justice (2011: 314-5). However, once we distinguish between the traits that underlie a capacity for institutional and interpersonal justice, we may understand how someone who possesses the former can learn to cultivate the latter too (319-20).

One character trait that seems to be genuinely instantiated in human beings is self-control, roughly, a capacity to delay gratification by exercising will power and structuring one’s perception of the appetitive temptations in one’s environment. People vary in their capacities for self control, and studies suggest that those high in self control are more likely to achieve a number of distinct goods for themselves and for others (Mischel et al. 1992). Self control is not, however, a robust character trait. One learns to exercise it in some contexts but not in others. According to Railton, empirical work by Peter Gollwitzer (2009, cited in Railton 2011: 322) offers clues about how to develop situational self control. Open-ended plans tend to be ineffective. Participants in Gollwitzer’s studies are more likely to exert self-control by implementing “if-then” plans that are cued to particular situational contexts. If one is trying to lose weight, a plan simply not to eat too much food is of little help. It is more effective to form the plan that “*if* it is past 9:00 pm, *then I* will not eat.” Thus, Gollwitzer’s research seems to offer general lessons about how to cultivate self control in a psychologically feasible way. More broadly, one can now see

how psychological research on actual human character traits and their development could be used to develop a richer, non-ideal virtue ethics.

Let's turn now from character to moral reasoning. In non-ideal ethics the aim is not to arrive at a static criterion of right and wrong. Even if there were such a criterion, we cannot pretend to know what it is. We are even likely to be ignorant about what many of the right moral questions are, or how best to pose them. What's more likely to be within our grasp, however, are methods of improvement, ways of arriving at better guides for moral action, even if the very best guide forever escapes us. Railton (2011, 2014) and Kwame Appiah (2006) argue that unconscious processes of habituation are critical for moral improvement. Elizabeth Anderson (2014, forthcoming) argues that a certain type of social activism is an ineliminable tool for overcoming biases that afflict those in positions of power. She sees this type of activism as instrumental in the American abolitionist movement and suggests that it is likely to be useful in other, contemporary movements. However, another important tool for moral improvement is moral reasoning.

Rawls (1971) famously proposed that normatively ideal moral reasoning aims to arrive at wide reflective equilibrium (see also Daniels 1979). We begin by bringing together our considered judgments about particular cases, plausible moral principles, and other normative and empirical theories about human beings and society. We then search for conflicts between these sets of claims. To arrive at reflective equilibrium we revise or modify claims in each of the sets, preserving those that seem the most plausible, until we achieve consistency. Thus characterized, wide reflective equilibrium presents itself as a supremely rational process, and so offers a seemingly apt tool for moral improvement.

The problem, however, is that while wide reflective equilibrium might be an apt tool for certain kinds of creatures, it seems to be of limited use to human beings. First, bringing together *all* of our beliefs connected to morality is a herculean task, and figuring out which of these to revise is perhaps no less difficult. Second, except for certain single-minded moral philosophers like Bentham or Kant, nearly everyone is a *moral pluralist* (Ross 1930). We do not think one moral principle reigns supreme. We accept a number of different moral norms and values that sometimes come into conflict. Unless we pretend that some ordinal ranking of these norms and values is possible, wide equilibrium seems to be a chimerical goal. Third and finally, empirical evidence suggests that people are often likely to accept principles not because they find them independently plausible but, rather, because they serve to *rationalize* the moral beliefs they hold about concrete cases (see Haidt 2001). And so, any actual attempt to reach reflective equilibrium threatens to devolve into mere rationalization.

Richmond Campbell and Kumar (2012) invite us to consider another type of moral reasoning. "Consistency reasoning" consists, roughly, in treating like cases alike. It is typically a targeted process of social reasoning. For example, suppose that although I am an avid dog owner, I also eat meat. A vegetarian friend, however, presses me: what's the difference between factory farming and practices that I already consider abhorrent, like

dog fighting? Faced with this challenge, and provided that I am disposed to trust my interlocutor, I should either decline to condemn dog fighting or, more likely, change my opinion about factory farmed meat. Unless I can find some morally relevant difference between these two practices, I should treat like cases alike.

Campbell and Kumar point out that consistency reasoning is a common, socially embedded mode of moral change, not just in philosophical debates but also in the law and in everyday social engagement. Furthermore, Campbell and Kumar argue that consistency reasoning is psychologically feasible because it is implemented by two pre-existing psychological systems commonly described in so-called “dual process” models of moral cognition (2012: 291-6). The authors suggest too that consistency reasoning has likely been an engine of progressive moral change, for example, in attitudes toward homosexuality (2012: 287; see also Kumar and Campbell 2016). So, if we’re looking for a tool for moral improvement, one that can be implemented by human beings and that has a good track record, consistency reasoning seems to offer more hope than wide reflective equilibrium. What turns on is which type of moral reasoning is psychologically more feasible and which tends to lead more often to instances of progressive moral change.

Non-ideal ethics eschews traditional questions in moral philosophy about universal criteria for right action or about ideally virtuous agents. Rather, non-ideal ethics begins with our current behavior, character, and institutions, and then asks what sorts of recent and current changes are feasible and morally progressive. From this it draws generalizations about the nature of progressive moral change. The target of theoretical analysis is not ideal behavior, or ideal character, or ideal institutions, but incremental moral progress in each of these moral domains. The philosophical foundations of non-ideal ethics are just beginning to be unearthed. Still, the project has obvious appeal. A number of different social science disciplines are potentially relevant to non-ideal ethics, including anthropology and history. But cognitive science also seems to offer guidance about the psychological feasibility of various moral changes. This is how empirical research into the human mind can advance theory construction in normative ethics.

4. Conclusion

Skepticism about naturalistic ethics may stem from the belief in a gap between *is* and *ought*. In fact, however, the *is*-*ought* gap does not support skepticism. Some topics in ethical theory are descriptive rather than normative. Ethicists who seek answers to questions about the nature of moral judgment should examine research in social psychology. Furthermore, normative inquiry in ethical theory can also draw on cognitive science provided that normative premises are supplied. We can better assess our moral beliefs by examining their bases and subjecting them to normative scrutiny. Finally, we can also understand the nature of moral progress by studying the psychological feasibility of moral change, in light of a normative framework that privileges the non-ideal over the ideal.

This may not be all that cognitive science offers to ethics, but it is quite enough to merit enthusiasm about this new form of philosophical naturalism.

References:

- Anderson, E. 2010. *The Imperative of Integration* (Princeton University Press).
- Anderson, E. 2014. "The Quest for Free Labor: Pragmatism and Experiments in Emancipation," *The Amherst Lecture in Philosophy*, 9: 1-44.
- Anderson, E. Forthcoming. "The Social Epistemology of Morality: Learning from the Forgotten History of the Abolition of Slavery," in M. Brady & M. Fricker, Eds., *The Epistemic Life of Groups*.
- Appiah, K. 2006. *Cosmopolitanism: Ethics in a World of Strangers* (Norton).
- Ayer, A. J. 1952. *Language, Truth and Logic* (Dover Publications).
- Berker, S. 2009. "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs*, 37: 293-329.
- Boyd, R. 1991. "Realism, anti-foundationalism and the enthusiasm for natural kinds," *Philosophical Studies*, 61: 127-48.
- Brownstein, M. (2015). "Implicit Bias," in E. N. Zalta, Ed., *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = <http://plato.stanford.edu/archives/spr2015/entries/implicit-bias>
- Campbell, R. 2007. "What is Moral Judgment?" *Journal of Philosophy*, 104: 321-49.
- Campbell, R. & Kumar, V. 2012. "Moral Reasoning on the Ground," *Ethics*, 122: 273-312.
- Chapman, H. & Anderson, A. 2013. "Things Rank and Gross in Nature: a Review and Synthesis of Moral Disgust," *Psychological Bulletin*, 139: 300-27.
- Churchland, P. 1981. "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy*, 78: 67-90.
- Daniels, N. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy*, 76: 256-82.
- Danziger, S., Levav, J., & Avnaim-Pesso, L. 2011. "Extraneous Factors in Judicial Decisions." *Proceedings of the National Academy of Sciences* 108: 6889-92.
- Doris, J. 1998. "Persons, Situations, and Virtue Ethics," *Nous*, 32: 504-30.
- Doris, J. 2002. *Lack of Character* (Cambridge University Press).
- Fehr, E. & Fischbacher, U. 2004. "Social Norms and Human Cooperation," *Trends in Cognitive Science*, 8: 185-90.
- Fodor, J. 1981. *Representations: Philosophical Essays on the Foundations of Cognitive Science* (MIT Press).
- Fodor, J. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (MIT Press).
- Geach, P.T. 1965. "Assertion," *Philosophical Review*, 74: 449-65.

- Gibbard, A. 1990. *Wise Choices, Apt Feelings* (Harvard University Press).
- Gollwitzer, P. 2009. "Self-Control by If-Then Planning," presentation to the Research Seminar in Group Dynamics, Institute for Social Research, University of Michigan (March 16).
- Greene, J. 2008. "The Secret Joke of Kant's Soul," in W. Sinnott-Armstrong, Ed., *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders and Development*, vol. 3 (MIT Press).
- Greene, J. 2014. "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics." *Ethics* 124: 695-726.
- Griffiths, P. 1997. *What Emotions Really Are: The Problem of Psychological Categories* (University of Chicago Press).
- Harman, G. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society*, 109: 316-31.
- Haidt, J. 2001. "The Emotional Dog and its Rational Tail: a Social Intuitionist Approach to Moral Judgment," *Psychological Review*, 108: 814-834.
- Holton, R. 2009. *Willing, Wanting, Waiting* (Oxford University Press).
- Isen, A. & Levin, P. 1972. "Effect of Feeling Good on Helping: Cookies and Kindness," *Journal of Personality and Social Psychology*, 21: 364-88.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford University Press).
- Joyce, R. 2006. *The Evolution of Morality* (MIT Press).
- Kelly, D. 2011. *Yuck!: The Nature and Moral Significance of Disgust* (MIT Press).
- Knobe, J. 2010. "Person as Scientist, Person as Moralist," *Behavioral and Brain Sciences*, 33: 315-65.
- Kornblith, H. 2002. *Knowledge and Its Place in Nature* (Oxford: Oxford University Press).
- Kripke, S. 1980. *Naming and Necessity* (Cambridge: Harvard University Press).
- Kumar, V. 2014. "'Knowledge' as a Natural Kind Term," *Synthese*, 191: 439-57.
- Kumar, V. 2015. "Moral Judgment as a Natural Kind," *Philosophical Studies*, 172: 2887-2910.
- Kumar, V. 2016-a "The Empirical Identity of Moral Judgment," *The Philosophical Quarterly*, 66: 783-804.
- Kumar, V. 2016-b. "Psychopathy and Internalism," *Canadian Journal of Philosophy*, 46: 318-45.
- Kumar, V. Forthcoming-a. "Moral Vindications," *Cognition*.
- Kumar, V. Forthcoming-b. "Foul Behavior," *Philosophers' Imprint*.
- Kumar, V. & Campbell, R. 2012. "On the Normative Significance of Experimental Moral Psychology," *Philosophical Psychology*, 25: 311-30.

- Kumar, V. & Campbell, R. 2016. "Honor and Moral Revolution," *Ethical Theory and Moral Practice*, 19: 147-59.
- Kumar, V. & May, J. Forthcoming. "How to Debunk Moral Beliefs," in J. Suikkanen, Ed., *The New Methods of Ethics*.
- Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., Sartori, G., Dial, C. M., Sriram, N., Banaji, M. R., & Nosek, B. A. 2014. "Reducing Implicit Racial Preferences: A Comparative Investigation of 17 Interventions," *Journal of Experimental Psychology: General*, 143: 1765-85.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong* (Penguin).
- May, J. 2014. "Does Disgust Influence Moral Judgment?" *Australasian Journal of Philosophy* 92: 125-141.
- May, J. 2015. "The Death of Double Effect," under review.
- Mischel, W., Shoda, Y., Rodriguez, M. 1992. "Delay of Gratification in Children," in G. Lowenstein & J. Elster, Eds., *Choice Over Time* (Russell Sage Foundation).
- Murphy, S., Haidt, J. & Bjorklund, F. 2000. "Moral Dumbfounding: When Intuition Finds No Reason," Unpublished manuscript, Department of Philosophy, University of Virginia.
- Nichols, S. (2014). "Process Debunking and Ethics." *Ethics*, 124: 727-49.
- Petrinovich, L. & O'Neill, P. 1996. "Influence of Wording and Framing Effects on Moral Intuitions," *Ethology and Sociobiology*, 17: 145-71.
- Prinz, J. 2004. *Gut Reactions: A Perceptual Theory of Emotions* (Oxford University Press).
- Prinz, J. 2007. *The Emotional Construction of Morals* (Oxford University Press).
- Putnam, H. 1975. "The Meaning of 'Meaning,'" in K Gunderson, Ed., *Language, Mind and Knowledge* (Minneapolis: University of Minnesota Press), 131-193.
- Railton, P. 1986. "Moral Realism," *Philosophical Review*, 95: 163-207.
- Railton, P. 2011. "Two Cheers for Virtue: Or, Might Virtue be Habit Forming?" in M. Timmons, Ed., *Oxford Studies in Normative Ethics vol. 1* (Oxford University Press): 295-330.
- Railton, P. 2014. "The Affective Dog and its Rational Tale: Intuition and Attunement." *Ethics*, 124: 813-59.
- Rawls, J. 1971. *A Theory of Justice* (Harvard University Press).
- Rosenberg, A. 2011. *The Atheist's Guide to Reality: Enjoying Life without Illusions* (Norton).
- Ross, W. D. 1930. *The Right and the Good* (Oxford University Press).
- Rozin, P. 1999. "The Process of Moralization," *Psychological Science*, 10: 218-21.
- Rozin, P., Haidt, J., & McCauley, C. 2008. "Disgust," in M. Lewis, J. Haviland-Jones, * L. Barrett, Eds., *Handbook of Emotions* (Guilford Press).

- Ruse, M. 1986. *Taking Darwin Seriously* (Blackwell).
- Schroeder, M. 2008. *Being For* (Oxford University Press).
- Schroeder, T. 2004. *Three Faces of Desire* (Oxford University Press).
- Sen, A. 2009. *The Idea of Justice* (Harvard University Press).
- Shafer-Landau, R. 2003. *Moral Realism* (Oxford University Press).
- Sinnott-Armstrong, W. 2008. “Framing Moral Intuitions.” W. Sinnott-Armstrong (ed), *Moral Psychology, Vol. 2*, (MIT Press).
- Smith, M. 1994. *The Moral Problem* (Malden, MA: Blackwell).
- Stevenson, C. 1937. “The Emotive Meaning of Ethical Terms,” *Mind*, 46: 14-31.
- Stich, S. 1983. *From Folk Psychology to Cognitive Science* (MIT Press).
- Stich, S. 1996. *Deconstructing the Mind* (Oxford University Press).
- Street, S. 2006. “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127:109-166.
- Tybur, J., Lieberman, D., Kurzban, R., & DeScioli, P. 2013. “Disgust: Evolved Function and Structure,” *Psychological Review*, 120: 65-84.
- Weiskopf, D. 2009. “The Plurality of Concepts,” in *Synthese*, 169: 145-73.
- Wheatley, T. & Haidt, J. 2005. “Hypnotically Induced Disgust Makes Moral Judgments More Severe,” *Psychological Science*, 16: 780-4.