

THE EMPIRICAL IDENTITY OF MORAL JUDGEMENT

BY VICTOR KUMAR

I argue that moral judgement is a natural kind on the grounds that it plays a causal/explanatory role in psychological generalizations. I then develop an empirically grounded theory of its identity as a natural kind. I argue that moral judgement is a hybrid state of moral belief and moral emotion. This hybrid theory supports the role of moral judgement in explanations of reasoning and action and also supports its role in a dual process model of moral cognition. Although it is composed of moral belief and moral emotion, moral judgement is unified, like other natural kinds, in virtue of underlying mechanisms that generate homeostasis among its components. Finally, this theory about the unity of moral judgement resolves a longstanding puzzle about moral motivation. The view explains how it is that moral judgements motivate all on their own but not all the time.

Keywords: moral judgement, natural kind, hybrid theory, dual process, homeostasis, internalism.

Philosophical discussion of moral judgement is generally attuned to evidence that is accessible from the armchair: intuitions about cases, the logic of moral discourse, moral phenomenology. Naturalized study of moral judgement, however, relaxes the grip of armchair evidence on philosophical discussion. Empirical research in cognitive science suggests that moral judgement is a *natural kind*. This research seems to shed light on the underlying nature of moral judgement.

A broad class of empirical work indicates that moral judgement is a *hybrid state* of moral belief and moral emotion. To judge that someone has wronged me, for example, is to believe that I have been wronged *and* feel resentment towards the wrongdoer. This hybrid theory is compelling as an empirically grounded theory about the identity of a natural psychological kind. Moral judgement *must be* a hybrid state of moral belief and moral emotion if it is to play the explanatory roles assigned to it in cognitive science.

A hybrid theory of moral judgement is empirically attractive, but it also has more familiar merits. Born from the union of cognitivism and non-cognitivism, a hybrid theory gets the best of both worlds. Yet, it also seems to get the worst. Specifically, it would appear that a hybrid theory of moral judgement, like

simpler views, is unable to account for the peculiar relationship between moral judgement and motivation. Puzzlingly, moral judgements seem to motivate all on their own, but not all the time. If moral judgement is a natural kind, however, a hybrid theory is poised to resolve this longstanding puzzle.

I. NATURAL KIND

Cornell Realism is the view that moral properties like the good and the right are natural kinds (Sturgeon 1988; Boyd 1988; Brink 1989; see also Railton 1986a,b; cf. Harman 1977). According to Cornell Realists, we have reason to believe that a category is a natural kind if it enters into scientific laws or explanatory generalizations that provide an objective account of the causal relationships found in the natural world (see also Bird and Tobin 2012). Many other philosophers argue, similarly, that certain mental state categories like belief, desire, emotion, etc. are natural psychological kinds on the grounds that these mental states enter into explanatory generalizations in the psychological sciences (Fodor 1981, 1987; Griffiths 1997; Kornblith 2002; Schroeder 2004; Prinz 2004; Holton 2009; Weiskopf 2009). I will begin the essay by arguing that moral judgement meets this general criterion for natural kindhood.

If an object of study is a natural kind, then we cannot understand its nature solely from the armchair. A theory about the identity of a natural kind must be empirically grounded. The theory is warranted if it isolates properties (1) that are present in paradigm instances of the kind and (2) that support the role of the kind in scientific laws or explanatory generalizations. If moral judgement is a natural kind, then, it is identical with properties that account for its presence and causal/explanatory role in psychological generalizations (see also Kumar 2015, forthcoming).

Ongoing research in empirical moral psychology primarily targets the cognitive and neurophysiological mechanisms that produce moral judgements (Haidt 2001; Greene and Haidt 2002; Nichols 2004; Greene 2008; Moll et al. 2008; Mikhail 2011; Campbell and Kumar 2012). This research offers explanations of moral judgements in terms of the processes that give rise to them. To establish that it is a natural kind, however, we must show not that moral judgement is explanandum, but that it is explanans—not what is explained, but what is doing the explaining.

Elsewhere in cognitive science, researchers use moral judgement to explain human reasoning and behaviour. First of all, moral judgements explain reasoning in which the judgements serve as premises in inference, although the ‘inferences’ may be largely unconscious. How we think about non-moral issues, it turns out, is deeply influenced by our moral judgements. Thus, to begin with, Joshua Knobe and others find that moral judgement explains reasoning about intentional action in the ‘Knobe effect’ or ‘side-effect effect’ (Knobe 2003a,b;

Nadelhoffer 2005; Knobe and Burra 2006; Leslie et al. 2006; Cushman and Mele 2008; see Knobe 2010b for review). In the canonical demonstration of the Knobe effect, participants are presented with a case in which an agent foresees that his action will have either a harmful or beneficial outcome; in neither case does this influence his decision to act. Participants tend to report that the agent brings about the outcome intentionally when it is harmful, but that he does not bring about the outcome intentionally when it is beneficial. So, it seems, when participants make a negative moral judgement about someone's action, this leads them to infer that a foreseen outcome is intentional.

Knobe's interpretation of these results is that moral judgement explains reasoning about intentional action. This interpretation, however, has been challenged (cf. Alicke et al. 2011; Sytsma et al. 2011; Sripada 2010; Railton 2014; Robinson et al. 2015). For example, Chandra Sripada (2010) argues that participants draw an inference about the agent's 'deep self', and that this explains their conclusion about intentional action. At this stage of investigation, it is possible that an alternative explanation of this sort will turn out to be correct, and that moral judgement does not underlie the Knobe effect.

However, over the past decade or so, Knobe and others have carried out a series of striking experimental studies showing the moral judgement influences reasoning in a wide range of other domains. First of all, analogues of the Knobe effect arise for many other mental state ascriptions (Pettit and Knobe 2009), even for ascriptions of knowledge (Beebe and Buckwalter 2010). Moral judgements also seem to explain reasoning about happiness (Phillips et al. 2011), means and side-effects (Knobe 2010a), doing and allowing (Cushman et al. 2008), free action (Young and Phillips 2011), the 'true self' (Newman et al. 2014), and innate traits (Knobe and Samuels 2013). Even if an alternative explanation of the Knobe effect were extended so as to account for some of these findings, it is unlikely that it could account for all of them.

To further illustrate the broad influence of moral judgement on reasoning, let's look at a body of research suggesting that moral judgements explain aspects of causal reasoning (Alicke 2000; Knobe and Fraser 2008; Hitchcock and Knobe 2009; Cushman 2010; Knobe 2010a). In one study, for example, Mark Alicke (2000) had participants read about a student who gets into a car accident after speeding home from school. The student was speeding either to hide her parents' anniversary present before they got home or to recover a vial of cocaine that she had accidentally left on the coffee table. Participants are more likely to report that the student was the cause of the accident when she was returning home for the vial of cocaine. As Alicke argues, it seems that when participants make a negative moral judgement about the student they assign greater causal responsibility to her. Results like this suggest that moral judgements influence reasoning about the relative causal contribution of a person's action to its outcomes.

So far, I have argued that moral judgements explain reasoning in a number of other domains. Still, critics may wonder whether it is moral judgement specifically, or instead some broader category of normative judgement, that underlies these effects. Are the judgements at play *moral* judgements? Inquiry in this essay is empirically grounded, but initial, pretheoretical classification is needed to get this sort of inquiry off the ground (cf. Kauppinen 2008, 2013). The sorts of mental states that are studied in Knobe, Alicke, and others' research seem, intuitively, to be moral judgements. The notion that someone is recklessly speeding typically elicits moral judgements about her behaviour, and not some other type of non-moral, normative judgement. Elsewhere, I offer an empirical theory that distinguishes moral judgements from other normative judgements and thus makes sense of our pretheoretical classification (Kumar 2015). However, all that is needed here is the initial, pretheoretical classification—not a deeper theoretical understanding of what supports it (cf. Clipsham 2014, who argues that normative matters bear on classification).

Nonetheless, since moral judgement is a species of normative judgement, perhaps it is the more general category that explains reasoning in other domains. Perhaps, for example, judging that an agent has violated a norm, moral or otherwise, leads people to infer that her actions are intentional and that she is the main cause of an outcome (see Robinson et al. 2015). Taken alone, then, it remains a possibility that the evidence reviewed thus far shows only that *normative* judgement is a natural kind. Moral judgement, however, has a more distinctive influence on behaviour; one that is not shared by normative judgement more generally.

Moral judgements seem to explain reasoning in other domains, as we have seen, but they also seem to explain pro-social and punitive behaviour. Whether or not we are disposed to act warmly or harshly toward others depends on the moral judgements we make about them. Empirical studies that support this observation are found in behavioural economics. In these studies moral judgements explain individuals' performance in games that involve real financial incentives (see Fehr and Fischbacher 2004a for review). For example, whether participants cooperate or defect in prisoners' dilemmas and public goods games seems to depend upon their moral judgements about the behaviour of other players (Fehr and Gächter 2000; Fischbacher, Gächter and Fehr 2001; Keser and van Winden 2000; Brandts and Schram 2001). When participants judge that another player has acted rightly they are likely to cooperate; when participants judge that another player has wronged them or others they are likely to defect. The effect obtains even in one-shot encounters where there is no expectation of further interaction, and thus where cooperation and defection cannot be explained by motives related to long-term self-interest.

Moral judgements also explain why people inflict punishments upon others (Fehr and Gächter 2000; Turillo et al. 2002; Fehr and Fischbacher 2004b). In the ultimatum game, for example, one player is instructed to offer part of a sum

of money to another player. The other player can either accept or reject the offer; if she rejects it neither player receives anything. Were players motivated solely by financial considerations, we would expect them to accept any offer. Often, however, players reject an offer when it is a small proportion of the sum and thus receive nothing. Participants, it seems, judge that the other player has acted unfairly and on the basis of their moral judgement inflict a punishment that runs against their own financial interest (Pillutla and Murnighan 1996).

Let's take stock. Several bodies of empirical research together indicate that moral judgement meets a general criterion for natural kindhood. Moral judgements explain reasoning about mental states, causation, and a great number of other seemingly non-moral matters, on the one hand, and they also explain cooperative, uncooperative, and punitive behaviour, on the other. However, if we wish not merely to establish that moral judgement is a natural kind, but also to develop an empirical theory about its identity, the next step is to isolate properties that are present in paradigm instances of moral judgement and that support its causal/explanatory role in psychological generalizations.

As I said, a pretheoretical classification of certain mental states as moral judgements is needed to gain an initial grasp of the subject matter. However, we should not rely too much on intuitions about cases if we are to achieve greater understanding. As is widely accepted, a natural kind can turn out to be different from the prior conception of it that is available from the armchair (Kripke 1980; Putnam 1975; cf. Boyd 1999; Kumar 2014). Empirical study of a natural kind may force us to revise our prior conception (though to what extent is an open question—see Kumar 2014). Thus, I will attempt to understand moral judgement itself—its empirically knowable identity. First, I will examine more closely how moral judgement explains reasoning and behaviour. After that, I will look at the psychological mechanisms that explain how moral judgements are produced. My attention, therefore, will be with psychological processes that are, causally speaking, both upstream and downstream of moral judgement. If moral judgement is a natural kind, as it seems to be, then it is identical to properties that are present in these processes and that support its role in our best scientific explanations of them.

II. HYBRID THEORY

Several authors have considered whether moral judgement is a natural kind and, if so, whether empirical research can provide an account of its underlying nature (Kelly et al. 2007; Kelly and Stich 2007; Sinnott-Armstrong and Wheatley 2014; Kumar 2015). However, much of the existing work on this subject focuses on the conceptual content of moral judgements. Elsewhere, for example, I argue that moral judgement is a distinctive category of normative judgement in which right and wrong, good and bad, etc. are conceptualized

as serious, general, authority-independent, and objective (Kumar 2015). This is what distinguishes moral judgements from other types of non-moral, normative judgements. Evidently, normative judgements with this content have a broad influence on reasoning and behaviour. So far, however, less effort has been expended applying the present naturalized approach to another metaethical question: What type of *attitude* is constitutive of moral judgement?

The disagreement between cognitivists and non-cognitivists is perhaps the most central and enduring philosophical debate about the nature of moral judgement. When the subject matter is psychological, rather than linguistic, what the two sides disagree about is whether moral judgement is fundamentally a cognitive, truth-evaluable attitude or a non-cognitive, motivational attitude. Both sides appeal to powerful considerations. On the one hand, it appears that moral judgements enter into logically valid inference, even in embedded contexts, and that they can embody moral knowledge. On the other hand, it appears that moral judgements have an internal link to motivation and that when someone makes a moral judgement she is not simply making a claim about what is true, but rather deciding what to do. Thus, moral judgements seem to be descriptive, and yet they also seem to be practical.

Hybrid theories offer a new perspective on this old debate by claiming that neither of these appearances is misleading: moral judgement consists in *both* a cognitive state and a distinct motivational state. Moral judgement is descriptive and also has a practical focus, but these aspects are realized in two separate components. What typically animates hybrid theorists are the familiar considerations mentioned above that pure cognitivists and pure non-cognitivists each have difficulty jointly accommodating (Copp 2001; Ridge 2006; Campbell 2007b). We'll attend to these considerations in the final section of the essay, but my immediate task is to argue that a hybrid theory is *empirically* attractive.

The body of research discussed in §1 suggests that moral judgements have the causal roles characteristic of cognitive states *and* motivational states. Moral judgements influence reasoning about mental states and causation, and they also influence pro-social and punitive behaviour. A hybrid theory makes sense of the two different types of psychological explanations in which moral judgements enters. Moral judgements are cognitive states because they are causally implicated in reasoning, and they are also motivational states because they cause action. Thus, a hybrid theory supports the dual causal/explanatory role of moral judgement in reasoning and behaviour.

Are other, competing theories of moral judgement also able to support its dual causal/explanatory role? Cognitivists may argue that moral beliefs can explain behaviour in virtue of a general background desire to act morally. Non-cognitivists may argue that motivational states can explain reasoning, especially in light of minimalist accounts of truth and belief. These are important alternatives; any complete defence of a hybrid theory must contend

with them. However, my aim in the present essay is to present the apparent advantages of a hybrid theory of moral judgement, construed as an empirically grounded theory about the identity of a natural kind. Whether we can conclude, further, that is better than all alternatives is a more difficult question that must be pursued elsewhere (Kumar, *forthcoming*). So, my focus here is on the empirical plausibility of a hybrid theory of moral judgement. And the first empirical reason that counts in its favour is that it offers a straightforward account of the dual causal/explanatory role of moral judgement.

A number of hybrid theories are already present in the literature. Some hybrid theories identify the motivational element of moral judgement with desire (Ridge 2006), others with a state of normative commitment (Copp 2001). As I say, these hybrid theories all offer a straightforward account of the role of moral judgement in reasoning and action. However, a great deal of empirical evidence suggests that emotions play a pervasive role in moral thought (for review see Nichols 2004; Prinz 2007; Campbell and Kumar 2012; Greene 2013; Railton 2014). This evidence suggests, tentatively, that in paradigm cases of moral judgement moral motivation might be grounded in emotion. That moral judgement consists partly in emotion is, at any rate, a hypothesis worthy of investigation.

In the rest of the essay, I will argue that moral judgement is a hybrid psychological state of moral belief and moral emotion—that is, a belief with moral content and a moral emotion like resentment, guilt, sympathy, outrage, repugnance, etc. (see Campbell 2007a,b; Campbell and Kumar 2012). Moral emotions are distinctive in the sense that they differ from various non-moral counterparts, e.g., in the way that resentment differs from anger. If a broadly cognitivist view of moral emotion is correct, this difference may be grounded in the moral concepts encoded in moral emotions (see again Kumar 2015). In any case, this view, like other hybrid theories, makes sense of the effect of moral judgement upon reasoning and action. But it also dovetails with what we know about the psychological mechanisms that produce moral judgements. To see this, it will help to begin with an interesting class of moral judgements that influence reasoning and action in the characteristic ways we have already observed, but in *opposite* directions.

III. BELIEF AND EMOTION

I have argued, so far, that moral judgement is a natural kind and that a hybrid theory supports its causal/explanatory role in reasoning and behaviour. One promising type of hybrid theory says that moral judgement consists in moral belief and moral emotion. When I judge that ‘stop-and-frisk’ policing is discriminatory and morally wrong, for example, my judgement consists in a belief that it is wrong and a moral emotion of resentment or outrage. In this

section, I will begin to make a case for this hybrid theory of moral judgement. I will build this case by appealing not just to the downstream influence of moral judgement on reasoning and behaviour, but also to the upstream psychological mechanisms that produce moral judgements.

In ‘conflict cases’, individuals reason and act as if they have contrary moral judgements. For example, one reasons as if one judges that an action is wrong, but acts as if one judges that it is right. Non-moral conflict cases are studied extensively by psychologists (see, e.g., Gendler 2008 for discussion). What conflict cases generally have in common is that they exhibit a ‘belief-behaviour mismatch’ (Kriegel 2012). One believes as if P, but acts as if not-P. For example, in one study participants explicitly affirm that certain actions are benign—eating fudge shaped like dog faeces, throwing darts at a picture of a loved one, etc.—but nonetheless betray negative implicit attitudes toward the actions and are reluctant to perform them (Rozin et al. 1986).

Psychologists have yet to narrow their focus on conflict cases that are distinctively moral—cases in which the attitudes in conflict are moral attitudes. But moral conflict cases are commonly found in philosophical discussion. For example, Jonathan Bennett (1974) finds a moral conflict case in *The Adventures of Huckleberry Finn*. Huck has helped Jim escape his life as a slave, and the two are paddling a raft down the Mississippi river. But then Huck reasons (speciously) that Jim is the property of Miss Watson, that in helping Jim escape he is acting as a thief, and therefore that he ought to hand him over to the authorities. This conclusion leads Huck to doubt his course of action and deliberate with himself about what to do. Later on, however, when the chance arrives to give Jim up, Huck cannot bring himself to do so. He feels too much sympathy for Jim and his plight, and thus, he acts to save Jim from slave catchers.

Richmond Campbell (2007b) introduces another conflict case. As he observes, conflict cases often arise during episodes of moral transition. Campbell invites us to imagine a moral transition that is likely common in North America. Many people enter adolescence believing that homosexuality is immoral, perhaps, because it is contrary to religious teachings. As they grow up, some begin to question their religious commitments. Many more are exposed to openly gay men and women in their personal life and in the media. Some of these people change their minds and come to believe that homosexuality is morally permissible. However, it is common for many to continue, at least for a while, to experience moral disgust or repugnance whenever they see gay people express their sexuality. The upshot is a tension between their moral attitudes. People commonly believe that homosexuality is morally permissible, but they are also morally disgusted by it (for empirical and philosophical discussion of moral disgust, or repugnance, see Rozin et al. 2008; Kelly 2011; Kumar forthcoming).

When moral reasoning and moral action diverge, it seems to be belief and emotion that are at the root of the conflict. Huck believes that it is wrong to help

Table 1. List of features characteristic of each process type in a minimalist dual process model of moral cognition.

Type 1 processes	Type 2 processes
Fast	Slow
Spontaneous	Controlled
Unconscious	Conscious
Emotional processing	Reasoned processing

Jim escape, reasoning on that basis. But he also experiences sympathy for Jim's well-being and is motivated to help him. Many North Americans believe that homosexuality is morally permissible and draw inferences on that basis. But they also experience moral disgust (or repugnance) towards gay people and are automatically motivated to avoid and exclude them. If we want to explain how individuals in conflict cases reason, we should appeal to their moral beliefs. If we want to explain how they act, we should appeal to their moral emotions. So, it seems, when we examine cases where reasoning and action dissociate, moral beliefs support the role of moral judgement in reasoning, while moral emotions support the role of moral judgement in action.

But how do moral belief and moral emotion dissociate? In general, conflict cases—moral and non-moral—call out for an explanation in terms of a 'dual process' model (see, e.g., Sloman 1996; Chaiken and Trope 1999; Stanovich and West 2000; Kahneman 2011). Elsewhere, Richmond Campbell and myself argue that there is empirical support for a 'minimalist' dual process model of moral cognition (Campbell and Kumar 2012). A minimalist model says that two types of processes generate moral judgements. Type 1 processes are fast, spontaneous, unconscious, and involve emotional processing; type 2 processes are slow, controlled, conscious, and involve reasoned processing. In short, some moral judgements arise as a flash of feeling, while others issue from conscious deliberation (See Table 1).

This characterization of a dual process model of moral cognition is so minimalist that it encompasses even those theorists who profess to be critical of dual process approaches (e.g., Nichols 2004; Railton 2014). To illustrate, notice that some moral psychologists defend views that entail a minimalist dual process model but reach well beyond it. Jonathan Haidt (2001, 2012) claims that type 1 processes produce nearly all of our moral judgements, while the function of type 2 processes is to generate post-hoc rationalizations of those judgements. Joshua Greene (2008, 2013) claims that type 1 processes are error-prone, while type 2 processes are rational. Greene also claims that deontology emerges from system 1, while consequentialism emerges from system 2. I do not believe that these ideas stand up to critical scrutiny (see, e.g., Kahane et al. 2012; Campbell and Kumar 2012; Kumar and Campbell 2012; Railton 2014),

and they form no part of the minimalist dual process model upon which I will rely.

For my purposes here, what is critical is that a minimalist dual process model explains the generation of hybrid moral judgements. Each process type gives rise to one element of moral judgement: type 1 processes generate moral emotions and type 2 processes generate moral beliefs. A minimalist model also explains conflict cases. Typically, belief and emotion complement one another in moral judgement, but when the process types underlying moral cognition are out of sync, one may find oneself with conflicting moral beliefs and moral emotions. Moral judgement is divided because feeling calls for one conclusion and deliberation calls for another. Thus, it seems, belief and emotion support the causal/explanatory role of moral judgement in a dual process model of moral cognition, including conflict cases. In the following sections, I will argue that closer scrutiny confirms this hypothesis.

Notice that dual process models and hybrid theories are not equivalent. The former is a scientific theory about the psychological mechanisms that produce moral judgements, while the latter is a philosophical theory about the nature of the mental state produced by those mechanisms. The one is about process, the other about product. Nonetheless, understanding the process enables us to better grasp the nature of the product. To participate in a dual process account of moral cognition, it seems, moral judgement must be a hybrid mental state.

Before moving on, it is worth asking whether this hybrid theory betrays some sort of confusion about the nature of judgements. Consider the following dilemma. Either moral emotion has propositional content, in which case it is a separate judgement. Or moral emotion does not have propositional content, in which case it does not contribute any judgemental component to the hybrid state. Either way, moral emotion is not constitutive of moral judgement. Whether moral emotion has propositional content is a difficult question, and seems to underlie the debate between cognitivists and non-cognitivists about emotion. I do not take a stand on it here. However, I do reject apparent implications of this view for the present topic. Pretheoretically, it seems plausible that if one mental state has different propositional content from another, then it is a separate judgement, and also plausible that judgements, *per se*, have propositional content. However, I think that these pretheoretical claims are revisable in light of empirical inquiry into the nature of psychological kinds. That moral emotion does, or does not, have propositional content does not bear on whether the nature of moral judgement *qua* psychological kind supports its causal/explanatory role in psychological generalizations.

Where has the empirical evidence led us so far? First of all, evidence suggests that moral judgement is a natural kind because it explains a wide range of reasoning and behaviour. Thus, we must rely much less on our prior conception of moral judgement and the armchair philosophical discussion that it anchors. Instead, we must empirically investigate the properties that are common to

instances of moral judgement and that support its causal/explanatory role. The view that moral judgement is a hybrid state of moral belief and moral emotion accounts for upstream influences described in a minimalist dual process model. The view also accounts for downstream effects of moral judgement on reasoning and action, i.e., those explored by experimental psychologists and behavioural economists. And so, it seems, a hybrid theory is an empirically justified theory about the identity of a natural psychological kind.

In the rest of the essay, I will flesh out and defend this hybrid theory. The next step is to look more closely at the dual process model of moral cognition that supports it. One general challenge that proponents of hybrid theories must address is to show that while the constituents of moral judgement are multiple, moral judgement is nonetheless unified. I will offer a naturalistic response to this challenge that rests on a richer understanding of moral cognition than is present in a minimalist dual process model. In the end, too, we'll see that a hybrid theory, once fully developed, seems to capture intuitions that have long divided internalists and externalists about moral judgement and motivation.

IV. INTUITION AND REASONING

To defend the present hybrid theory of moral judgement, and to make good on the claim that it is a unified natural kind, we must examine the psychological mechanisms underlying moral cognition in further detail. I have argued that a minimalist dual process model supports a hybrid theory of moral judgement, but so far I have said little about why we should accept that model. In this section, I will briefly review relevant evidence, defend my interpretation of this evidence, but also set the stage for a more expansive, *integrated* model of moral cognition that bears further philosophical fruit.

Dual process models are increasingly common in cognitive and social psychology. There are dual process models of statistical inference, decision-making, mental state attribution, and a host of other psychological processes. What all dual process models have in common is that they posit two types of processes underlying cognition in some domain. Type 1 processes are faster than type 2 processes, operate unconsciously rather than consciously, and are relatively spontaneous rather than controlled. In moral cognition, but not in certain other domains of thought, type 1 processes implicate emotion. But according to some dual process theorists, the two process types can be distinguished in yet further ways. For example, some theorists claim that type 1 processes are encapsulated, innate, domain-specific, and associative, while type 2 processes are unencapsulated, learned, domain-general, and rule-based. The dual process model of moral cognition to which I have so far appealed is minimalist. It does not entail whether the two process types can be distinguished in these further, more controversial ways.

Evidence for a minimalist dual process model of moral cognition comes from a number of different sources: moral ‘dumbfounding’ (Haidt et al. 2000; Haidt 2001; Cushman et al. 2006; Hauser et al. 2007); reaction time under ‘cognitive load’ (Greene et al. 2008); fMRI analysis (Greene et al. 2001, 2004); brain damage patients (Koenigs et al. 2007). Greene is one of the principal researchers behind these studies, and he argues that they support a more controversial version of the dual process model. According to Greene, type 1 processes generate ‘deontological’ judgements, while type 2 processes generate ‘consequentialist’ judgements. As Greene operationally defines these terms, ‘deontological’ judgements are those that fit naturally with deontology, for example, judging that it is wrong to push a large man into the path of an oncoming trolley in order to save five other people. ‘Consequentialist’ judgements are those that fit naturally with consequentialism, for example, judging that it is permissible to divert a trolley from a track on which it will run over five people to a track on which it will run over only one person. Thus, on Greene’s view, type 1 processes are triggered by things like violence and unfairness, while type 2 processes perform utility calculations.

In fact, however, the distinction between type 1 and type 2 processes cross-cuts the distinction between deontological and consequentialist judgements (Campbell and Kumar 2012: 279–81). On the one hand, type 2 processes are often engaged to apply explicit deontological principles. Many people have a negative, automatic reaction to bigoted speech, but reason from deontological principles concerning freedom of expression that censorship of bigoted speech is impermissible. On the other hand, intuition often generates characteristically consequentialist judgements. The judgement that drunk driving is wrong is fast, automatic, and emotionally charged, although it is based on the obviously harmful consequences of drunk driving.

Greene’s research exploits a pattern between type 1 processes and deontological judgements, and between type 2 processes and consequentialist judgements. The pattern is genuine, but it is local only to the particular moral dilemmas that Greene selects for his participants. A recent study by Guy Kahane et al. (2012) confirms this interpretation by employing a broader and more diverse class of moral dilemmas. The researchers find that the neurological correlates of type 1 processing are correlated not with deontological judgements but, rather, ‘intuitive’ judgements—both deontological and consequentialist. The neurological correlates of type 2 processing are correlated not with consequentialist judgements but, rather, ‘counterintuitive’ judgements—again, both deontological and consequentialist.

I have defended a minimalist dual process model of moral cognition, and in the course of doing so, I have rejected Greene’s controversial view that deontology emerges from type 1 processes, consequentialism from type 2 processes. However, I have interpreted the dual process model in yet another way that diverges from Greene. According to Greene, type 1 and type 2 processes

independently produce moral judgements. However, as I see it, type 1 processes generate moral emotions and type 2 processes generate moral beliefs. Are there any reasons to favour this interpretation over Greene's? There are, I believe, and they lie in the integration of intuition and reasoning.

In the next section, I will move beyond minimalism and toward an 'integrated process' model of moral cognition. Type 1 and type 2 processes typically operate in coordination with one another. When type 1 processes generate a moral emotion, type 2 processes typically generate the corresponding moral belief; when type 2 processes generate a moral belief, type 1 processes typically generate the corresponding moral emotion. Contrary to Greene, then, one process gives rise to moral judgements only in the sense that one and not the other process takes this lead in this coordinated operation.

Greene seems to think that the processes underlying moral judgement are alien to one another—in much the same way that the two hemispheres of a brain are alien to one another when the connection between them is severed. This is likely because he focuses on an unusual and unrepresentative class of moral dilemmas that generate psychological conflict. In general, however, intuition and reasoning are integrated in moral cognition. Normally, neither process type works solely on its own. I will examine evidence for integration next. Moreover, I will argue that integration of intuition and reasoning provides the basis for an explanation of the unity of moral judgement.

V. INTEGRATION AND UNITY

I have argued on empirical grounds that moral judgement consists in moral belief and moral emotion. But does it follow from the evidence cited that 'moral judgement' is really two things rather than one? Perhaps the discovery that moral judgement is both moral belief and moral emotion is somewhat like the discovery that jade is two kinds of minerals rather than one. If this is right, then it seems to follow that moral judgement fractures into two natural kinds—one kind explaining reasoning and the other explaining action, one kind explained by conscious deliberation and the other by unconscious intuition (cf. Kriegel 2012).

In this section, I will argue that moral judgement is a unified natural kind. It is unified, I suggest, in the same way that complex natural kinds in the special sciences are typically unified: in virtue of a homeostatic relationship among its components. That is, the cognitive and affective constituents of moral judgement comprise a *homeostatic property cluster*. The underlying mechanism that sustains homeostasis is the integration of type 1 and type 2 processes.

Let's begin with some critical background. Richard Boyd (1988, 1991) argues that a natural kind is a homeostatic property cluster; that is, a cluster of properties that tend to support and reinforce one another, either directly or via

an underlying mechanism, and especially in the face of external perturbation. H_2O , for example, is a natural kind because it is composed of hydrogen and oxygen molecules in a stable state that resists change into different combinations of the same atoms, in virtue of underlying chemical bonds.

The homeostatic model of natural kinds is especially plausible in the biological and psychological sciences (less so in fundamental physics). Biological and psychological kinds often do not seem to have essential properties, nor do they admit of sharp boundaries. As Boyd (1991) argues, for example, in the evolution of one species into another there is often no clear transition point. Organisms arise that have some but not all of the properties that define either the new or the ancestral species. Essentialist theories of natural kinds have difficulty accommodating such cases. But on the homeostatic model, these organisms are simply the atypical cases that tend to arise when there is somewhat looser homeostasis than is found at the level of, say, molecules. They have some but not all of the properties in the homeostatic cluster that defines a given species, and thus creep towards its fuzzy extensional boundaries.

In moral judgement, I suggest, moral belief and moral emotion form a homeostatic property cluster, mutually supporting and reinforcing one another, especially in the face of external perturbation. Conflict cases provide preliminary motivation for this idea, since they tend to be unstable. We often resolve tensions between our moral beliefs and moral emotions by bringing them into conformity with one another. I will consider relevant empirical evidence below, but for the moment consider again the examples of conflict cases discussed above. Many people reflect on the tension between their belief that homosexuality is permissible and their seemingly irrational disgust or repugnance, and attempt to suppress their emotional response. The change happens in the other direction too, when people trust feelings over beliefs. Those who feel sympathy for another person often come to revise their ideologically driven belief that it is wrong to help him.

The underlying mechanism that regulates moral belief and moral emotion in homeostasis is the complex coordination of type 1 and type 2 processes. In moral cognition, each process generates one kind of mental state: type 1 produces moral emotions, type 2 produces moral beliefs. However, development into a mature moral agent involves harmonization between thinking and feeling. Thus, in moral development each process type becomes dependent upon the other and an associative link between moral beliefs and moral emotions is established. Once an output is generated the other process is engaged to produce the corresponding mental state. Feeling resentment toward someone, we reason that she has done something wrong. Or reasoning that someone has quietly acted in a brave manner, we immediately feel awe and admiration.

A great deal of empirical evidence supports the view that when type 1 processes give rise to an emotion, type 2 processes follow by giving rise to a corresponding belief, and vice versa. Of particular interest is when the two

systems produce conflicting outputs, and homeostasis is restored by modifying either the moral belief or the moral emotion. Homeostasis entails not just mutual reinforcement and support, recall, but also resistance to perturbation.

First of all, empirical findings indicate that emotions tend to modulate moral beliefs. Most persuasive are manipulation studies in which experimentally induced emotions influence moral beliefs. The psychological literature is replete with studies exhibiting this finding, but consider just a few examples. Simone Schnall et al. (2008) find that participants who sit at a filthy desk, and therefore experience disgust, express more negative moral beliefs than controls. Emotion induction can also influence moral belief when participants are otherwise not disposed to form the belief. Thalia Wheatley and Haidt (2005) find that participants who are hypnotically induced to feel disgust express more negative moral beliefs, even about apparently quite innocent actions.

Evidence suggests, too, that moral beliefs modulate moral emotions. One way this happens is relatively uncontroversial. Before we find ourselves with an emotional response, we typically appraise the situation. For example, we form the belief, perhaps unconsciously, that a criminal has committed intentional harm, and that leads us to feel anger towards him. In this example, the appraisal is a non-moral belief. However, moral beliefs can also provide the appraisal that activates affective processing (Haidt 2003). For example, if we reason that a politician has unfairly privileged his own supporters, that usually leads us to feel resentment or repugnance toward her.

Moral beliefs also exert an influence on conflicting moral emotions, thus restoring homeostasis after perturbation. Paul Rozin's (1997, 1999) research on moralization provides the clearest illustration of this effect, and it investigates how our attitudes and behaviour evolve once we believe things to be of moral concern rather than mere matters of preference. New moral beliefs about such things as smoking, eating meat, alcohol consumption, spitting, obesity, disease, etc. often recruit, over time, the emotion of disgust. Thus, the belief that smoking is morally wrong eventually activates moral disgust toward smokers (Rozin and Singh 1999). Those who become vegetarians for moral reasons are more likely than those who become vegetarians for health reasons to be disgusted by the sight and smell of meat (Rozin et al. 1997).

Although emotions often have an immediate effect on moral belief, the ability of moral belief to modulate moral emotion typically emerges only over a significant length of time (Campbell and Kumar 2012). The effect is therefore more difficult to study experimentally. Reasoning's effect on emotion most clearly unfolds at a historical scale, as societies undergo moral change. The moral emotional responses of our ancestors are much different than our own, and this is likely due in part to the gradual effect of reasoning on emotion (Bloom 2010; Campbell and Kumar 2012: 282–3; Kumar and Campbell 2016).

A minimalist dual process model seems to account for conflict cases of moral judgement. But I have now made a case for not just minimalism but

also an integrated process model of moral cognition. Underlying coordination between two types of processing in moral cognition regulates moral belief and moral emotion in homeostasis. When one component changes, the other is likely to follow, sooner or later, because over the course of moral development each process type becomes attuned to the other. Moral judgement is thus a cluster of properties that is stable and mutually reinforcing. This is what grounds its unity as a natural psychological kind. It is also what resolves a longstanding puzzle about moral judgement and motivation.

VI. MORAL JUDGEMENT AND MOTIVATION

Moral judgement is a hybrid state of moral belief and moral emotion because these elements support the causal/explanatory role of moral judgement. And moral judgement is a unified mental state—it does not fracture into two entirely separate natural kinds—because intuition and reasoning are integrated in moral cognition. However, as is amply discussed in the literature, hybrid theories also satisfy more traditional philosophical criteria that seem to frustrate pure cognitivism and pure non-cognitivism (Copp 2001; Ridge 2006; Campbell 2007b; Schroeder 2009).

To begin with, hybrid theories straightforwardly account for the truth-evaluable purport of moral thought and discourse, which seems also to be a precondition for logically valid inference and for moral knowledge. As complex states of belief and motivation, hybrid moral judgements contain a truth-evaluable component. Hybrid theories and cognitivism therefore share a major advantage over non-cognitivism because both are able to more easily capture truth-evaluability, logically valid moral inference, and moral knowledge.

Like non-cognitivism, however, hybrid theories allow moral judgements to have a practical focus and they enable moral statements to have ‘directive’ force. When we engage in moral discourse we appear not simply to assert matters of fact. The point of telling someone that giving to charity is morally right is to express my own motivation to give, or enjoin my audience to give, or demand that they share my motivation, etc. Moral statements cannot serve in anything like these directive roles if their sole function is to express beliefs. However, moral statements function as directives if they express motivational states along with beliefs.

Despite these advantages, the complex relationship between moral judgement and motivation appears to frustrate hybrid theories no less than it frustrates pure cognitivism and pure non-cognitivism. The relationship seems internal, but it also seems contingent. On the one hand, moral judgements appear to have inherent motivational impetus. On the other hand, the conceivability of ‘amoralists’ suggests that moral judgement does not entail motivation (Brink 1989: 45–60). Someone can judge that an action is morally

right, it seems, but if she happens not to care about being moral, then she may not be motivated to perform the action. For hybrid theories, it would seem, the relationship between moral judgement and motivation must be internal rather than contingent, if motivation is in fact partly constitutive of moral judgement. Thus, it would seem, hybrid theories are no better able than non-cognitivism to capture intuitive data about moral motivation.

The relationship between moral judgement and motivation is, I believe, internal *and* contingent. We are now in a position to understand why. Prototypical moral judgements are composed of belief and emotion, unified because they are in homeostasis, and are therefore motivating all on their own. The link between the components of moral judgement is, however, not exceptionless—as is the case generally among complex, homeostatically unified kinds in the special sciences. As noted in the previous section, some biological organisms exhibit variation away from the species norm and thus are atypical examples of their kind. Similarly, cases in which a moral belief or moral emotion lacks its corresponding component, are atypical. Because there are atypical moral judgements that are only beliefs without their emotional counterparts, atypical moral judgements can fail to be motivating. Thus, the naturalistic account of the unity of moral judgement developed above makes sense of the otherwise puzzling relationship between moral judgement and motivation.

The upshot is a qualified rejection of internalism. At least two versions of internalism are discussed in the literature. The most common states that necessarily, if someone forms a moral judgement he or she possesses the corresponding motivation. This canonical version of internalism is false, as conflict cases and the more familiar case of the amoralist suggest (see also Kumar forthcoming). A second, less common version of internalism states that moral judgement is intrinsically motivating in the sense that motivation is an intrinsic property of moral judgement (Shafer-Landau 2003: 147–8). This version of internalism is true. Moral judgements are intrinsically motivating (as the second version of internalism says) because prototypical cases motivate on their own and without the help of a distinct motivational state, but they are not necessarily motivating (as the first version says) since moral judgement does not have necessary properties.

The crucial anti-essentialist idea here, suggested by Boyd, is that natural kinds can have intrinsic properties even if this isn't to be cashed out in terms of essential properties. Like other natural kinds, moral judgement is identical to a cluster of homeostatically related properties and exhibits natural variation in which only some of the properties are instantiated. Thus, we can hold on to the view that moral judgement motivates in and of itself, but also accept that moral judgement and motivation come apart. For moral judgements are intrinsically but not necessarily motivating.

The view I am now defending is sometimes called 'soft internalism': moral judgement has an internal link to motivation, but the link is defeasible

(Timmons 1999: 140–1; see also Shafer-Landau 2003: 148). On my view, moral judgement motivates defeasibly because its components are not just cognitive but also affective. However, the homeostatic link between them is sometimes broken, with the result that atypical moral judgements obtain that are only beliefs and therefore lack the affective, motivating component that is typically present in moral judgement.

Although I share with other philosophers a commitment to soft internalism, my naturalized approach forces an alternative interpretation of the view. Traditionally, the internal relationship between moral judgement and motivation is thought to hold as a conceptual matter: the concept of moral judgement is the concept of a mental state that is intrinsically motivational. On the present view, soft internalism is, rather, a synthetic claim—a truth about a natural kind and not a concept (Kumar forthcoming). That moral judgement is motivating is on par with claims like ‘mammals are warm-blooded’ or ‘representations are structured.’ Of course, claims like these must be evaluated empirically, and the empirical evidence that I have canvassed seems to vindicate soft internalism. Moral belief and moral emotion are both constitutive of moral judgement, but neither is essential.

VII. CONCLUSION

The aim of this essay has been to develop a view of moral judgement as an empirically grounded theory about the identity of a natural kind. Moral judgement is that which is present in paradigm cases and plays the causal/explanatory roles assigned to moral judgement in our best empirical theories of the mind. Identification of moral judgement with a hybrid state of moral belief and moral emotion captures moral judgement as a natural phenomenon and it also supports these causal/explanatory roles.

The science of moral cognition, evidently, can inform the philosophy of moral judgement. The key is to examine empirical work on the upstream and downstream psychological processes in which moral judgement participates. Type 1 processes are responsible for producing moral emotions, which are then responsible for guiding behaviour. Type 2 processes are responsible for producing moral beliefs, which are then responsible for influencing reasoning. But the two process types are integrated, yoking belief and emotion together. Moral belief and moral emotion are unified in moral judgement, then, because they are a homeostatic property cluster, subject to natural variation that gives rise to atypical cases. This is why moral judgement motivates all on its own, but not all the time. Thus, a naturalized approach offers insight into the nature of moral judgement and makes sense of its firm but fallible grip on motivation.

ACKNOWLEDGEMENT

For helpful feedback on this essay, I am grateful to many people, including Sara Aronowitz, Terry Horgan, Rachana Kamtekar, Uriah Kriegel, Theresa Lopez, Aaron Mead, Shaun Nichols, Peter Railton, Evan Tiffany, Mark Timmons, two anonymous referees for this journal, and especially Richmond Campbell. Previous versions of the essay were presented to audiences at the 2012 meeting of the Pacific division of the American Philosophical Association, the 2013 annual meeting of the Canadian Philosophical Association, the Moral Psychology Research Group, the University of Arizona, London School of Economics, the University of South Carolina, and the University of Michigan. Thanks to all of these audiences for valuable discussion.

REFERENCES

- Alicke, M. (2000) 'Culpable Control and the Psychology of Blame', *Psychological Bulletin*, 26: 556–74.
- Rose, D. and Bloom, D. (2011) 'Causation, Norm Violation and Culpable Control', *Journal of Philosophy*, 108: 670–96.
- Beebe, J. and Buckwalter, W. (2010) 'The Epistemic Side-Effect Effect', *Mind and Language*, 25: 474–98.
- Bennett, J. (1974) 'The Conscience of Huckleberry Finn', *Philosophy*, 49: 123–34.
- Bird, A. and Tobin, E. (2012) 'Natural Kinds', in E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/win2012/entries/natural-kinds/>.
- Bloom, P. (2010) 'How do Morals Change?', *Nature*, 464: 490.
- Boyd, R. (1988) 'How to Be a Moral Realist', in G. Sayre-McCord (ed.) *Essays on Moral Realism*, 181–228. Cornell University Press.
- (1991) 'Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds', *Philosophical Studies*, 61: 127–48.
- (1999) 'Homeostasis, Species, and Higher Taxa', in R.A. Wilson (ed.) *Species: New Interdisciplinary Essays*, 141–85. Cambridge, MA: MIT Press.
- Brandts, J. and Schram, A. (2001) 'Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach', *Journal of Public Economics*, 79: 399–427.
- Brink, D. (1989) *Moral Realism and the Foundations of Ethics*. New York: CUP.
- Campbell, R. (2007a) 'The Problem of Moral Judgment', in S. Sherwin and P.K. Schotch (eds) *Engaged Philosophy: Essays in Honour of David Braybrooke*, 241–70. Toronto: Toronto University Press.
- (2007b) 'What is Moral Judgment', *Journal of Philosophy*, 104: 321–49.
- and Kumar, V. (2012) 'Moral Reasoning on the Ground', *Ethics*, 122: 273–312.
- Chaiken, S. Y. Trope, (eds) (1999) *Dual Process Theories in Social Psychology*. New York: Guilford.
- Clipsham, P. (2014) 'Does Empirical Moral Psychology Rest on a Mistake?', *Philosophical Studies*, 170: 215–33.
- Copp, D. (2001) 'Realist-Expressivism: A Neglected Option for Moral Realism', *Social Philosophy and Policy*, 18: 1–43.
- Cushman, F. (2010) (Unpublished data) *The Effect of Moral Judgment on Causal and Intentional Attribution: What We Say, or How We Think?*, Harvard University.
- and Mele, A. (2008) 'Intentional Action: Two-and-a-Half Folk Concepts?', in J. Knobe and S. Nichols (eds) *Experimental Philosophy*, 171–88. Oxford: OUP.
- Knobe, J. and Sinnott-Armstrong, W. (2008) 'Moral Appraisals Affect Doing/Allowing Judgments', *Cognition*, 108: 353–80.

- Young, L. and Hauser, M. (2006) 'The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm', *Psychological Science*, 1082–9.
- Fehr, E. and Gächter, S. (2000) 'Cooperation And Punishment In Public Goods Experiments', *American Economic Review*, 90: 980–94.
- and Fischbacher, U. (2004a) 'Social Norms And Human Cooperation', *Trends in Cognitive Sciences*, 8: 185–90.
- (2004b) 'Third Party Sanctions And Social Norms', *Evolution and Human Behavior*, 25: 63–87.
- Fischbacher, U., Gächter, S. and Fehr, E. (2001) 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment', *Economics Letters*, 71: 397–404.
- Fodor, J. (1981) *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Gendler, T. (2008) 'Alief in Action (and Reaction)', *Mind and Language*, 23: 552–85.
- 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. Cambridge, MA: MIT Press.
- (2013) *Moral Tribes: Emotion, Reason and the Gap Between Us and Them*. New York: Penguin.
- and Haidt, J. (2002) 'How (And Where) Does Moral Judgment Work', *Trends in Cognitive Science*, 6: 517–23.
- Greene, J. et al. (2001) 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science*, 293: 2105–08.
- (2004) 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment', *Neuron*, 44: 389–400.
- (2008) 'Cognitive Load Selectively Interferes with Utilitarian Moral Judgment', *Cognition*, 107: 1144–54.
- Griffiths, P. (1997) *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
- Haidt, J. (2001) 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, 108: 814–34.
- (2003) 'The Moral Emotions', in R. Davidson, K. Scherer, and H. Goldsmith (eds.) *Handbook of Affective Sciences*, 852–70. Oxford: OUP.
- (2012) *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon.
- Bjorklund, F. and Murphy, S. (2000) (Unpublished manuscript) 'Moral Dumbfounding: When Intuition Finds No Reason', University of Virginia, Charlottesville.
- Harman, G. (1977) *The Nature of Morality*. New York: OUP.
- Hauser, M. et al. (2007) 'A Dissociation Between Moral Judgments and Justifications', *Mind and Language*, 22: 1–21.
- Hitchcock, C. and Knobe, J. (2009) 'Cause and Norm', *Journal of Philosophy*, 106: 587–612.
- Holton, R. (2009) *Willing, Wanting, Waiting*. Oxford: OUP.
- Kahane, G. et al. (2012) 'The Neural Basis of Intuitive and Counterintuitive Moral Judgment', *Social Cognitive and Affective Neuroscience*, 7: 393–402.
- Kahneman, D. (2011) *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kauppinen, A. (2008) 'Moral Internalism and the Brain', *Social Theory and Practice*, 34: 1–24.
- (2013) 'Ethics and Empirical Psychology', in M. Christen et al. (eds) *Empirically Informed Ethics*, 279–305. Berlin: Springer.
- Kelly, D. (2011) *Yuck! The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.
- and Stich, S. (2007) 'Two Theories of the Cognitive Architecture Underlying Morality', in P. Carruthers, S. Laurence and S. Stich (eds) *The Innate Mind*, vol. 3, 348–66. New York: OUP.
- et al. (2007) 'Harm, Affect, and the Moral/Conventional Distinction', *Mind and Language*, 22: 117–31.
- Keser, C. and van Winden, F. (2000) 'Conditional Cooperation and Voluntary Contributions to Public Goods', *Scandinavian Journal of Economics*, 102: 23–39.
- Knobe, J. (2003a) 'Intentional Action and Side Effects in Ordinary Language', *Analysis*, 63: 190–3.
- (2003b) 'Intentional Action in Folk Psychology: An Experimental Investigation', *Philosophical Psychology*, 16: 309–24.

- (2010a) ‘Action Tree and Moral Judgment’, *Topics in Cognitive Science*, 2: 555–78.
- (2010b) ‘Person as Scientist, Person as Moralist’, *Behavioral and Brain Sciences*, 33: 315–65.
- and Burra, A. (2006) ‘Intention and Intentional Action: a Cross-Cultural Study’, *Journal of Culture and Cognition*, 6: 113–32.
- and Fraser, B. (2008) ‘Causal Judgment and Moral Judgment: Two Experiments’, in W. Sinnott-Armstrong (ed.) *Moral Psychology*, vol. 2, 441–8. Cambridge, MA: MIT Press.
- and Samuels, R. (2013) ‘Thinking Like a Scientist: Innateness as a Case Study’, *Cognition*, 126: 72–86.
- Koenigs, M. et al. (2007) ‘Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments’, *Nature*, 446: 908–11.
- Kornblith, H. (2002) *Knowledge and Its Place in Nature*. Oxford: OUP.
- Kriegel, U. (2012) ‘Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction’, *Australasian Journal of Philosophy*, 90: 469–86.
- Kripke, S. (1980) *Naming and Necessity*. Cambridge: Harvard University Press.
- Kumar, V. (2014) ‘Knowledge’ as a Natural Kind Term’, *Synthese*, 191: 439–57.
- (2015) ‘Moral Judgment as a Natural Kind’, *Philosophical Studies*, 172: 2887–910.
- (2016) (Under review) ‘Foul Behavior’.
- (Forthcoming) ‘Psychopathy and Internalism’, *Canadian Journal of Philosophy*.
- (Forthcoming) ‘The Ethical Significance of Cognitive Science’, in S.J. Leslie and S. Cullen (eds) *Current Controversies in Cognitive Science*. Routledge.
- and Campbell, R. (2012) ‘On the Normative Significance of Experimental Moral Psychology’, *Philosophical Psychology*, 25: 311–30.
- and Campbell, R. (2016) ‘Honor and Moral Revolution’, *Ethical Theory and Moral Practice*, 19: 147–59.
- Leslie, A., Knobe, J. and Cohen, A. (2006) ‘Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment’, *Psychological Science*, 17: 421–7.
- Mikhail, J. (2011) *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: CUP.
- Moll, J., de Oliveira-Souza, R. and Zahn, R. (2008) ‘The Neural Basis of Moral Cognition: Sentiments, Concepts and Values’, *Annals of the New York Academy of Sciences*, 1124: 161–80.
- Nadelhoffer, T. (2005) ‘Skill, Luck, Control, and Folk Ascriptions of Intentional Actions’, *Philosophical Psychology*, 18: 343–54.
- Newman, G., Bloom, P. and Knobe, J. (2014) ‘Value Judgments and the True Self’, *Personality and Social Psychology Bulletin*, 40: 203–16.
- Nichols, S. (2004) *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: OUP.
- Pettit, D. and Knobe, J. (2009) ‘The Pervasive Impact of Moral Judgment’, *Mind and Language*, 24: 586–604.
- Phillips, J., Misenheimer, L. and Knobe, J. (2011) ‘The Ordinary Concept of Happiness (and Others Like It)’, *Emotion Review*, 71: 929–37.
- Pillutla, M. and Murnighan, J. (1996) ‘Unfairness, Anger and Spite’, *Organizational Behavior and Human Decision Processes*, 68: 208–24.
- Prinz, J. (2004) *Gut Reactions: A Perceptual Theory of Emotions*. Oxford: OUP.
- (2007) *The Emotional Construction of Morals*. Oxford: OUP.
- Putnam, H. (1975) ‘The Meaning of “Meaning”’, in K. Gunderson (ed.) *Language, Mind and Knowledge*, 131–93. Minneapolis: University of Minnesota Press.
- Railton, P. (1986a) ‘Facts and Values’, *Philosophical Topics*, 14: 5–31.
- (1986b) ‘Moral Realism’, *Philosophical Review*, 95: 163–207.
- (2014) ‘The Affective Dog and its Rational Tale’, *Ethics*, 124: 813–59.
- Ridge, M. (2006) ‘Ecumenical Expressivism: Finessing Frege’, *Ethics*, 116: 302–36.
- Robinson, B., Stey, P. and Alfano, M. (2015) ‘Reversing the Side-Effect Effect: The Power of Salient Norms’, *Philosophical Studies*, 172: 177–206.
- Rozin, P. (1997) ‘Moralization’, in A. Brandt and P. Rozin (eds) *Morality and Health*, 379–401. Routledge.
- (1999) ‘The Process of Moralization’, *Psychological Science*, 10: 218–21.
- Haidt, J. and McCauley, C. (2008) ‘Disgust’, in M. Lewis, J. Haviland-Jones and L. Barrett (eds) *Handbook of Emotions*. Guilford Press: 757–76.

- and Singh, L. (1999) 'The Moralization of Cigarette Smoking in the United States', *Journal of Consumer Psychology*, 8: 321–37.
- Markwith, M. and Stoess, C. (1997) 'Moralization and Becoming a Vegetarian: The Transformation of Preferences Into Values and the Recruitment of Disgust', *Psychological Science*, 8: 67–73.
- Millman, L. and Nemeroff, C. (1986) 'Operation of the Laws of Sympathetic Magic in Disgust and Other Domains', *Journal of Personality and Social Psychology*, 50: 703–12.
- Schnall, S. et al. (2008) 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin*, 34: 1096–1109.
- Schroeder, M. (2009) 'Hybrid Expressivism: Virtues and Vices', *Ethics*, 119: 257–309.
- Schroeder, T. (2004) *Three Faces of Desire*. New York: OUP.
- Shafer-Landau, R. (2003) *Moral Realism: A Defense*. Oxford: OUP.
- Sinnott-Armstrong, W. and Wheatley, T. (2014) 'Are Moral Judgments Unified?', *Philosophical Psychology*, 27: 451–74.
- Sloman, S. (1996) 'The Empirical Case for Two Systems of Reasoning', *Psychological Bulletin*, 119: 3–22.
- Sripada, C. (2010) 'The Deep Self Model and Asymmetries in Folk Judgments About Intentional Action', *Philosophical Studies*, 152: 159–76.
- Stanovich, K. and West, R. (2000) 'Individual Differences in Reasoning: Implications for the Rationality Debate?', *Behavioral and Brain Sciences*, 23: 645–65.
- Sturgeon, N. (1988) 'Moral Explanations', in G. Sayre-McCord (ed.) *Essays on Moral Realism*, 229–55. Ithaca: Cornell University Press.
- Sytsma, J., Livengood, J. and Rose, D. (2011) 'Two Types of Typicality: Rethinking the Role of Statistical Typicality in Ordinary Causal Attributions', *Studies in History and Philosophy of Science*, 43: 814–20.
- Timmons, M. (1999) *Morality Without Foundations: A Defense of Ethical Contextualism*. New York: OUP.
- Turillo, C. et al. (2002) 'Is Virtue its Own Reward? Self Sacrificial Decisions for the Sake of Fairness', *Organizational Behavior and Human Decision Processes*, 89: 839–65.
- Weiskopf, D. (2009) 'The Plurality of Concepts', *Synthese*, 169: 145–73.
- Wheatley, T. and Haidt, J. (2005) 'Hypnotic Disgust makes Moral Judgments More Severe', *Psychological Science*, 16: 780–4.
- Young, L. and Phillips, J. (2011) 'The Paradox of Moral Focus', *Cognition*, 119: 166–78

University of Toronto, Canada