

Moral judgment as a natural kind

Victor Kumar

Published online: 5 February 2015
© Springer Science+Business Media Dordrecht 2015

Abstract In this essay I argue that moral judgment is a natural kind by developing an empirically grounded theory of the distinctive conceptual content of moral judgments. Psychological research on the moral/conventional distinction suggests that in moral judgments right and wrong, good and bad, praiseworthiness and blameworthiness, etc. are conceptualized as (1) serious, (2) general, (3) authority-independent, and (4) objective. After laying out the theory and the empirical evidence that supports it, I address recent empirical and conceptual objections. Finally, I suggest that the theory uniquely accounts for the possibility of genuine moral agreement and disagreement.

Keywords Moral judgment · Natural kind · Homeostasis · Moral/conventional distinction · Social morality · Moral objectivity · Moral disagreement

Moral judgments seem to be different from other normative judgments, even apart from their characteristic subject matter. Two people might both disapprove of an action, for example, although one judges it a moral violation and the other a breach of etiquette. Philosophers have traditionally attempted to define moral judgment through reflection alone. However, psychological research on the “moral/conventional distinction” offers a promising source of empirical evidence about the distinctive nature of moral judgment.

Several authors treat the ability to draw a distinction between morality and convention as a test for the presence of moral judgments (Blair 1995; Nichols 2004a; Prinz 2007; Levy 2007). None, however, develops the implied theory of

V. Kumar (✉)
Philosophy Department, University of Michigan, 2215, Angell Hall, 435 S State St.,
Ann Arbor, MI 48109, USA
e-mail: victor.c.kumar@gmail.com

moral judgment. Shaun Nichols (2004a) uses research on the moral/conventional distinction to develop a model of the psychological mechanisms that underpin moral evaluation. According to Nichols, internally represented rules guide judgments about moral and conventional violations, but moral judgments are also backed by emotion. Focused on psychological mechanisms, Nichols shies away from a definition of moral judgments linked to the moral/conventional distinction, saying that “the attempt to draw an analytic distinction between morality and convention is fraught with controversy” (2004a: 5).

I believe that psychological research on the moral/conventional distinction delineates moral judgment as a *natural kind*. In Sects. 1, 2 and 3 I explain why moral judgment might be a natural kind and survey relevant research. In Sect. 4 I draw on that research to develop a view about the concept of morality encoded in moral judgments, one that explains what moral judgments have in common and what sets them apart from other normative judgments. Daniel Kelly, Stephen Stich, and colleagues argue that a philosophical theory of moral judgment based on the moral/conventional distinction is empirically inadequate (Kelly et al. 2007; Kelly and Stich 2007; Nado et al. 2009). Walter Sinnott-Armstrong and Thalia Wheatley (2014) argue that any such theory is vulnerable to counterexamples. I address these objections in Sects. 5 and 6.

The principal focus of this essay is the nature of moral judgment. But this topic is also linked to the possibility of moral agreement and disagreement. If moral judgments are different from other normative judgments, then whatever differentiates them must be shared by people who agree or disagree about moral matters. Famously, some philosophical views of moral judgment fail because they entail that moral interlocutors consistently talk past one another. In Sect. 7 I show that the view on offer comes with an empirical explanation for the possibility of genuine moral agreement and disagreement.

1 Natural kind

The term ‘moral judgment’ is ambiguous in metaethics, connoting a number of different psychological and linguistic categories. I am interested here in moral judgment qua mental state. I will examine psychological research that investigates the difference between moral judgments and other normative mental states, and then draw philosophical conclusions about the distinctive nature of moral judgments. But first I must unpack the thesis that moral judgment is a natural kind and explain how that thesis impacts philosophical methodology.

In general, we have reason to believe that an object of study is a natural kind if it explains a broad class of phenomena by participating in scientific laws or generalizations (Bird and Tobin 2012). Thus, moral judgment is a natural psychological kind if it plays an explanatory role in psychological generalizations. In Kumar (2015a) I argue that moral judgment is a natural kind on the grounds that

it explains reasoning in several other domains and also explains cooperative, uncooperative, and punitive behavior.¹

If moral judgment is a natural kind we can develop a theory about its underlying nature in roughly the same way that scientists develop theories about other natural kinds. Philosophers have adapted this scientific methodology elsewhere in ethics (Boyd 1988; Brink 1989; Sturgeon 1988; cf. Railton 1986a, b), in epistemology (Kornblith 2002; Kumar 2014), and most extensively in the philosophy of mind (Fodor 1981, 1987; Griffiths 1997; Schroeder 2004; Prinz 2004; Holton 2009; Weiskopf 2009). Essentially, the idea is to investigate empirically the properties that are present in paradigm instances of the kind and that support the explanatory role of the kind in scientific laws or generalizations.

In Kumar (2015a) I use this methodology to argue that moral judgment is a hybrid state of moral belief and moral emotion (cf. Campbell 2007). This hybrid theory coheres with a “minimalist” dual process model of moral cognition that is independently plausible (Kumar 2015a: 10–12; see also Campbell and Kumar 2012: 276–84, 310–2). Conscious, deliberative processes generate moral beliefs. Unconscious, intuitive processes generate moral emotions. Critically, however, the two processes underlying moral cognition are integrated, and this explains why moral judgment is a unified kind (Kumar 2015a: 17–20). As Richard Boyd (1988, 1991) argues, a natural kind is a homeostatic property cluster, that is, a cluster of properties that are stable and mutually reinforcing, typically in virtue of an underlying mechanism and especially in response to external perturbation. A wealth of evidence indicates that moral belief and emotion are stably attuned to one another in moral cognition, and thus are in homeostasis (Kumar 2015a: 18–20).

Moral judgments have two ingredients: attitude and content. Thus, a central question in metaethics is whether moral judgment is a cognitive or non-cognitive attitude—or both, as I argue. The main topic of *this* essay is the internal, conceptual content of moral judgments. To see that this is a different issue, assume for the sake of argument that normative judgments, moral judgments included, are beliefs. Still, we might wonder, what is the distinctive conceptual content of moral beliefs, such that believing that it is morally wrong to cheat someone is a different kind of mental state than believing that it is prudentially wrong to do so?²

¹ For evidence that moral judgment explains reasoning in other domains see Knobe (2003a, b), Nadelhoffer (2005), Knobe and Burra (2006), Leslie et al. (2006), Cushman and Mele (2008), Knobe (2010a), Pettit and Knobe (2009), Beebe and Buckwalter (2010), Alicke (2000), Cushman et al. (2005), Knobe and Fraser (2008), Hitchcock and Knobe (2009), Knobe (2010a, b). For evidence that moral judgment explains cooperative, uncooperative, and punitive behavior see; Fischbacher et al. (2001), Keser and van Winden (2000), Brandts and Schram (2001), Fehr and Gächter (2000), Turillo et al. (2002), Fehr and Fischbacher (2004a, b), Pillutla and Murnighan (1996).

² Cognitivists tend to pay more attention to the conceptual content of moral judgment than non-cognitivists. The reason, perhaps, is that cognitivists tend to locate the normativity of moral judgment in the content of the judgment rather than the attitude. For cognitivists, after all, moral judgment is just an ordinary belief. Non-cognitivists, by contrast, tend to locate the normativity of moral judgment in attitude rather than content; the content is just the natural property toward which one has a non-cognitive state. However, non-cognitivists who accept this line of thought are misled. Even if moral judgment is a non-cognitive state, one must still identify the moral concepts through which non-cognitive attitudes apprehend their objects, concepts that distinguish moral attitudes from other non-cognitive attitudes of approval and disapproval.

In general, normative judgments have as their content that something is good or bad, right or wrong, praiseworthy or blameworthy, etc. In *moral* judgments, however, the concept of morality modifies these other normative concepts. Thus, one conceives of democratic equality as *morally* good, exploitation of workers as *morally* wrong, charitable giving as *morally* praiseworthy. The concept MORAL is what distinguishes moral judgments from other normative judgments. (I will follow standard practice of referring to concepts with small caps.)

What I am after in this essay is a theory of moral judgment as a natural kind that accounts for its distinctive conceptual content. The target of the theory is the concept MORAL, but my methodology will be empirical. Instead of engaging in traditional conceptual analysis, I will examine psychological research that probes the structure of the concept. As we'll see, MORAL is complex, i.e., composed of simpler concepts.

A brief word of clarification is necessary before we continue. I will focus on the concept MORAL that distinguishes moral judgments from other normative judgments. But I will not be concerned with the *concept of moral judgment*. Armchair analysis of the concept of moral judgment is not especially useful since, of course, natural kinds can turn out to be quite different from our prior conception of them (see Kripke 1980; Putnam 1975; cf. Boyd 1999; Kumar 2014). Thus, we can empirically investigate whether moral judgment is a natural kind without first analyzing our concept of moral judgment.

2 Social morality

In Sects. 3 and 4 I will survey relevant psychological research and then draw philosophical conclusions about the underlying nature of moral judgment. Interpretation of empirical data, however, typically depends on background theoretical assumptions. My task in this section is to articulate and defend a key assumption that is needed to derive philosophical significance from the empirical data.

From one perspective, morality is different from certain other normative domains in virtue of being inherently *social*. Morality, in other words, is about regulating our interactions with others. Arguably, prudential norms, epistemic norms, and aesthetic norms are not inherently social, although they may happen to concern interactions with others. For example, prudence demands that we protect ourselves from others' malice; epistemic norms favor a social distribution of cognitive labor; aesthetic norms enjoin praise for large-scale cooperative art projects. Unlike morality, however, each of these normative domains is only sometimes and only contingently social.

I do not wish to argue that morality can *only* be understood as a social category. Rather, I am suggesting that one concept of morality is social. For reasons that will soon become clear, I wish to focus on that concept. To understand the concept at play, it is useful to take note of a distinction that is commonly drawn between morality and ethics (although I will not draw it in precisely the same way that some authors have, e.g., Williams 1985: 174–196). Morality is about “how to get along

with others.” Ethics, by contrast, is a broader normative domain, about “how to live one’s life”—which includes how to get along with others but quite a bit more besides that. For example, considerations about one’s own welfare or well-being are germane to ethics but not morality. The subject of this essay is the concept of morality, in this sense, not the broader concept of ethics.

A social concept of morality seems to rule out, by definition, moral duties to oneself—even though it allows that there are “ethical” duties to oneself. In general, I believe, the implication is correct. I have a moral reason not to deprive someone of food. Intuitively, however, I do not have a *moral* reason not to deprive myself of food, though I may have prudential reason not to do so. Self/other asymmetries like this suggest that there is a concept MORAL that is inherently social.³ Some self-regarding actions, however, are morally relevant because they are *covertly* social. A woman has moral reasons not to subordinate her plans and hopes to those of her husband, not because she has a moral duty to herself, but because she has a moral duty not to perpetuate or symbolize the social phenomenon of gender inequality. Morality is self-regarding, in this case and others, when “the personal is political.”

A social concept of morality is apt for the present study because research on the moral/conventional distinction taps into that very concept. Like morality, convention is social (see Smetana 1993: 114). Typical conventional normative judgments concern matters of interpersonal etiquette, e.g., how to greet someone, what kind of clothing it is appropriate to wear in public, how to conduct oneself at the dinner table, etc. Thus, as I will make clear in the next section, focusing on a social concept of morality allows us to interpret research on the moral/conventional distinction as exploring a nearby contrast class. Morality and convention are both social, but research suggests that morality is unlike convention in that morality is serious, general, authority-independent, and objective. In the next two sections I will make the case that the concept of morality is defined in terms of these simpler concepts.

Before proceeding, it is worth emphasizing that my topic is conceptual rather than ontological—the conceptual content of moral judgments rather than moral properties or facts. I wish to understand how we conceptualize morality, setting aside whether there are normative properties that answer to, and vindicate, our moral concepts. For example, I will argue that morality is conceptualized as objective, which may be true even if there are no objective moral properties. The

³ Michael Bukoski objects to my argument here as follows. He agrees, at least for the sake of argument, that one does not generally have moral duties to oneself, but argues that this is a substantive claim, rather than a definitional claim. For example, although I am morally obliged not to kill others, I lack a moral duty not to commit suicide just because the correct norms of morality permit suicide and not because there is some kind of incoherence in the idea of such a duty. The problem with this objection is that it is difficult to make sense of self/other asymmetries on many normative ethical theories. Suppose that we agree that suicide is not morally wrong, you as a substantive matter and I as a definitional matter. The problem is that on many ethical theories, it’s difficult to see why killing one’s (healthy) self is not morally wrong. It does not maximize utility, it causes harm, it violates Kantian duties to oneself, etc. So, because for many ethical theories there is no substantive basis for permitting suicide, the best explanation is that a prohibition on suicide is invalid on definitional grounds. Although some ethical theories entail self/other asymmetries, this won’t explain why many Utilitarians, Kantians and others hold, apparently without blatant inconsistency, that suicide is not morally wrong.

view is thus compatible with an error-theory according to which there are no genuine obligations or reasons, or anyway none that are distinctively moral.

3 Morality and convention

I will begin this section by reviewing core research on the moral/conventional distinction, highlighting the most important findings. I will then review complementary empirical work that supplies a more complete picture of the distinction between morality and convention. The evidence must be handled carefully, however, and it won't be until the following section that I develop a philosophical hypothesis that synthesizes the evidence.

The last few decades have witnessed an eruption of empirical research on moral judgment, but not all of it is directly relevant to the present study. The object of study is not, for example, the psychological mechanisms that produce moral judgments, but the type of mental state that those psychological mechanisms produce. Initiated by the psychologist Eliot Turiel and his collaborators, a program of research on the moral/conventional distinction studies just that (for review see Turiel 1983; Smetana 1993; Tisak 1995; Nucci 2001; see also Kohlberg 1963, 1981). Employing hypotheses inspired by the work of moral philosophers, researchers elicit from subjects moral judgments and conventional normative judgments in order to study the psychological basis of the commonsense distinction between morality and convention.

In what is called the "moral/conventional task" subjects are presented with prototypical moral violations and prototypical conventional violations. For example, violently attacking someone and stealing someone's possessions are typical moral violations, whereas wearing inappropriate clothing in public and chewing gum in class are typical conventional violations. Subjects are then asked a series of questions that probe the bases of the distinction:

- How seriously wrong is the action (on a scale)?
- Is it wrong in other places and times?
- Would it still be wrong if someone in authority said it was OK?
- Why is it wrong?

Individuals across a number of populations exhibit a similar pattern of responses in the moral/conventional task. Moral violations are judged to be more seriously wrong than conventional violations; moral wrongs are thought to generalize to other places and times whereas conventional wrongs do not; and moral wrongness, unlike conventional wrongness, is held to be independent of authority. Furthermore, when asked to justify their moral judgments, subjects typically cite the harm that moral violations cause, and sometimes violations of rights or injustice. By contrast, subjects typically justify conventional judgments by appeal not to harm, but to custom or the need for coordination.

These findings are quite general and extend across development. Children as young as 3 years old reliably distinguish between moral and conventional violations (Nucci and Turiel 1978; Smetana 1981; Tisak and Turiel 1984; Nucci 1986;

Smetana and Braeges 1990). This is true of children with high and low IQ scores, from high and low socioeconomic backgrounds, and from healthy and unhealthy homes (Smetana et al. 1984). It is even true of children who suffer from autism (Blair 1996). Interestingly, psychopaths and children with nascent psychopathic traits do not reliably draw the moral/conventional distinction (Blair 1995; Blair et al. 1995; Blair 1997; but see Aharoni et al. 2012). Arguably, however, individuals from these populations do not make genuine moral judgments (Prinz 2007: 42–47; Kennett and Fine 2008: 173–178; cf. Kumar 2015b).

The findings generalize not just across development but also cross-culturally. Researchers have found that people draw the moral/conventional distinction not just in North America, but also in a wide range of other places (Nucci et al. 1983; Snarney 1985; Nucci 1986; Hollos et al. 1986; Song et al. 1987; Yau and Smetana 2003). Whether the moral/conventional distinction is entirely universal is somewhat controversial. Some anthropological data does not confirm existence of the distinction (Shweder et al. 1987). What seems to explain this data, however, is that the sorts of violations that many cultures regard as conventional are treated by some cultures as moral. Thus, the distinction obtains in these cultures too, but it is obscured by cross-cultural variation in substantive moral commitments. The moral/conventional distinction may be universal. However, the most that we can say with great confidence is that it is near-universal.

Care is needed to sift the evidence. In the studies reviewed above, researchers probe subjects in ways that reveal systematic differences between their moral and conventional judgments. These differences have the power to reflect how people conceptualize morality. However, not just any difference matters. What matters are those differences that issue from the concept of morality, and not from anything outside of it. As Nichols (2004a: 5) says, the issue is fraught. But careful philosophical attention can serve to filter out synthetic claims from the concept of morality, as we'll see now.⁴

One finding reviewed above is that subjects tend to justify moral violations by appeal to harm, injustice, or rights, whereas conventional violations are not so justified. However, folk theories about how moral claims are justified do not seem to be part of the concept of morality. Someone who does not believe that harm, injustice, or rights justify moral claims—believes instead that, say, the nature of rational agency grounds morality or that God's will does—nonetheless seems to have a view about morality and not some other normative domain. Nicholas Southwood (2011) argues persuasively that morality and convention differ in virtue of whether they are grounded in social practices. However, it would seem that many

⁴ The distinction at play here is roughly between analytic and synthetic truths and while I assume that some version of the distinction obtains, I do not assume that it is sharp. There may be borderline cases of truths that are not clearly analytic or synthetic. Famously, naturalists tend to reject the analytic/synthetic distinction. However, following Harman (1967), Campbell (1998: 145–149), and others, I interpret Quine's most persuasive arguments in "Two Dogmas of Empiricism" (1951) as showing not there are no analytic truths, but that there are no analytic truths knowable a priori—in the sense that there are no truths about meaning knowable independent of experience, that cannot be revised on the basis of experience. (I do not reject weaker conceptions of the a priori.) Consistent with this interpretation, my Quinean approach in the essay is to empirically investigate analytic truths associated with the concept of morality.

people gain a facility with moral concepts before they have any theory about what grounds them. Justificatory grounds, whatever role they may play in marking important boundaries in moral philosophy, are not internal to the ordinary concept of morality.

Having reviewed core research on the moral/conventional distinction, let's turn now to complementary research by Nichols (2004b) and Goodwin and Darley (2008) (see also Nichols and Folds-Bennett 2003; Wainryb et al. 2004). The research suggests that adults and children tend to conceive of morality as objective (see Campbell and Kumar 2013 for further discussion). In Goodwin and Darley's study, for example, subjects were presented with statements about morality, social convention, and scientific fact. If they agreed with the statement subjects were then asked a question designed to test whether they think that the statement is objective: "Imagine that someone else disagrees with the statement. Must one of you be wrong?" It turns out that subjects tend to treat moral statements more like scientific statements than conventional statements. That is, they tend to think that in moral disagreement at least one of the two parties must be wrong. So, this research suggests that another feature distinguishes morality from convention—morality is objective, whereas convention is not.⁵

A newer study by Sarkissian et al. (2011) appears to conflict with the data from Nichols and Goodwin and Darley. The researchers found that subjects are less likely to think that moral disagreement entails error when the disagreeing party is from another culture or species. In one study subjects in both the U.S. and Singapore were told about someone who finds his child unattractive and kills him. Of course, subjects said that the action is morally wrong. Adapting the approach of Nichols and Goodwin and Darley, the researchers then told subjects that someone else disagrees and asked whether one of them must be wrong. In this study, however, subjects were told that the other party is a member either of their own culture, a remote warrior culture in the Amazon, or a strange alien species. Sarkissian et al. replicated previous findings when the other party was described as a member of the subjects' own culture. However, subjects were significantly less likely to say that the other party must be wrong when he or she was described as an Amazonian warrior, and even less likely when described as a strange alien. The authors conclude that people are objectivists in some respects and relativists in others. They believe that there is a fact of the matter about morality, apparently, but that it is a species and culture specific fact.

⁵ It is worth noting that in Goodwin and Darley's study subjects exhibited variation depending on the moral statement they were asked to evaluate. For example, subjects tended to think that disagreement entails error with respect to whether cheating is wrong and whether robbery is wrong, but not with respect to whether euthanasia and abortion are wrong. One possible explanation for this difference is that subjects think some moral norms are objective, whereas others are not. If so, then objectivity is not a defining feature of morality. However, since subjects were mainly objectivists in their responses, it is more likely that their other responses reflected not a perceived lack of objectivity, but doubt about the likelihood of rationally adjudicating disagreement about certain moral matters. This is the salient difference between the two classes of moral statements at issue. It is beyond dispute that cheating and robbery are wrong, but highly controversial whether euthanasia and abortion are. Subjects should be expected to recognize this difference about the way in which these actions are generally regarded, even if their own moral opinions are firm.

Sarkissian et al.'s research suggests that people do not think of morality as fully objective, nor fully general. However, the research does not even begin to suggest that there is no difference in the way that people conceptualize morality and convention. Morality can be objective without being absolute, general without being universal. More specifically, people seem to conceive of morality as objective for similar groups of people, but not objective for all humans or for all rational beings, whereas they do not conceive of conventions as objective even to this degree. Similarly, people generalize morality across a broad range of cultures, whereas they generalize conventions only across local communities. This more nuanced characterization of objectivity and generality will be important later on.

Summarizing the findings reviewed above, researchers find that nearly everyone tends to distinguish morality from convention in four respects: seriousness, generality, authority-independence, and objectivity. People also tend to provide different justifications for morality—in terms of harm, injustice, and rights—but justifications are not part of the concept of morality. These results support a view about the concept of morality that I will now lay out in detail.

4 The concept of morality

It is now time to draw philosophical conclusions about the nature of moral judgment. The conclusions must be tentative, and not just for the ordinary reason that psychology experiments can be misleading or poorly designed in hidden ways, but also because we do not know how complete the testing has been. The findings reviewed above are constrained by researchers' antecedent hypotheses. Drawing on ideas in moral philosophy, researchers had a hunch that morality is conceived as general or objective, say, and then went out and empirically confirmed their hunch. But there may be additional features that distinguish morality from other convention, though researchers have not yet tested for them. Consequently, the view I will present must be regarded as a working hypothesis, empirically revisable and potentially incomplete.⁶

In most experimental studies researchers present subjects with cases in which moral norms are violated, rather than cases in which moral norms are followed. Thus, researchers' conclusions are often put in terms of how subjects distinguish moral wrongs from conventional wrongs. I am in search of a more general account, one that covers right as well as wrong, but also other terms of normative appraisal. My approach is to understand normative judgments as judgments about what is right, wrong, obligatory, forbidden, good, bad, praiseworthy, blameworthy, etc. In moral judgments, specifically, the concept MORAL modifies these other normative concepts. Moral judgments categorize an object of evaluation not simply as wrong, but *morally* wrong. We are now in a position to say what defines MORAL. A moral wrong, for instance, is a wrong that is:

⁶ Copp (1995) and Campbell (2009) defend views that are similar but also specify further distinguishing features of morality. It would be worthwhile to experimentally test these views, using methods of the sort described above. In this way philosophical theorizing might inform and guide further psychological research.

- (1) serious
- (2) general
- (3) authority-independent
- (4) objective

Moral wrongs are serious, they generalize to other places and times, their force is not conditional on the dictates of authority figures, and they are objective. That morality is social explains why it is more meaningfully contrasted with convention rather than other normative domains that are not inherently social. The four features are given full meaning through this contrast. Convention is less serious, cannot be generalized to other places and times, depends on authority for its validity, and is not objective to the same extent.⁷

This account of MORAL finds an important precedent in Hare's (1952) work. Hare argued that moral imperatives are different from other normative imperatives because morality is universal and overriding. That is, morality applies in all relevantly similar conditions and overrides conflicting normative demands. Several parallels are evident. First, generality is, of course, akin to universality. Second, seriousness is tantamount to overridingness. More generally, the present account is, like Hare's view, formal rather substantive, and thus also has the ability to make sense of moral error. Some people moralize relatively trivial matters of convention that should properly be regarded as morally-neutral. Nonetheless, their judgments are recognizably moral because they elevate non-moral matters beyond their proper significance, scope, authority, and objectivity.

Much more detail must now be added to complete the present theory of MORAL. Although there is no analytic or necessary connection between them, the four features that define MORAL are stable and mutually reinforcing. Moral judgment, like other natural kinds, is a homeostatic property cluster (Kumar 2015a). The human cognitive system is organized in such a way that the four features have a nomological tendency to cluster together. Kelly and Stich independently suggest a view like this, though only to criticize it (Kelly et al. 2007; Kelly and Stich 2007).

Core research on the moral/conventional distinction indicates that the features typically co-occur in moral judgment, and thus provides some evidence that they are a homeostatic property cluster. Further evidence comes from two sources. First, the claim that the four features are a homeostatic cluster entails not just they frequently co-occur, but that there is a nomological tendency for them to do so. Thus, the view predicts that if people are told that a norm violation has some of the features, they will likely believe that it has others too. A study by Judith Smetana (1985) tests this prediction among young children. The children were told that an unspecified action, denoted by a nonsense word (e.g., "pigling"), was wrong not just at school but also at home. So, the children conceived of the action as generally wrong. When subsequently probed with a variation on the moral/conventional task, children also

⁷ Notice that some of these features are less precise than what one might expect from a typical philosophical analysis. This lack of precision is in fact a strength of the view. I have identified what is in the ordinary person's mind when they make a moral judgment. We shouldn't expect a definition that is any more precise, for that can be constructed only through explication of the ordinary concept of morality.

tended to say that the action's wrongness is serious and independent of authority. Thus, even from a very young age people are likely to infer from the presence of some features that the other features are present too.

A second piece of evidence for homeostasis comes from research on psychopaths. It is widely accepted that psychopaths are morally dysfunctional, even though disagreement remains about precisely how to explain their moral deficits (Blair 1995; Nichols 2004a; Maibom 2005). James Blair finds that psychopaths do not exhibit the standard pattern of responses in the moral/conventional task and thus, it seems, have an impaired ability to distinguish between morality and convention (Blair 1995; Blair et al. 1995; see also Blair 1997). We now have an explanation. Psychopaths are morally abnormal because they do not have a full grasp of MORAL (Kumar 2015b). They are less likely to distinguish moral and conventional violations with respect to not just one feature but several of them. Thus, the view that the features are in homeostasis has predictive and explanatory power.⁸

What is the mechanism underlying homeostasis? A plausible developmental hypothesis can be found in Nichols (2004a). Sometimes the mechanisms underlying homeostasis in a given natural kind are historical. Arguably, biological species are defined in terms of a cluster of properties that are in homeostasis by virtue of common evolutionary descent (Boyd 1991: 142–143). Similarly, Nichols argues that affective mechanisms play a critical developmental role in establishing the capacity to distinguish morality from convention (Nichols 2004a: 16–20, 25–29). Psychopaths, he argues, do not fully acquire the distinction because they have severely impaired affective mechanisms in development (19–20). Notably, individuals who exhibit psychopathic traits due to brain damage, but who experienced this damage as adults rather than during childhood, *do* reliably draw a distinction between morality and convention (Saver and Damasio 1991).

The present theory entails no sharp distinction between moral judgments and other types of normative judgments. Even though the four features are nomologically linked, we should expect that under some conditions homeostasis will be disrupted, and that individuals will make normative judgments that token only some of the four features. Nichols' work suggests that affect plays a critical role in binding the features together. So, for example, perhaps when moral judgment is emotionally detached and clinical, one becomes susceptible to the thought that a violation is, say, general but not serious.

Normative judgments that have most of the four features, but not all, are atypical cases of moral judgment; those that have even fewer are borderline cases that do fall clearly inside or clearly outside the natural category of moral judgment. This kind of fuzziness is similarly present among other natural kinds in the special sciences. Think about the phenotypic variation exhibited by members of a biological species,

⁸ Standard research on the moral/conventional distinction does not examine the feature of objectivity. So, although we know from Blair's work that psychopaths tend not to distinguish moral and conventional violations with respect to seriousness, generality, and authority-independence, we do not yet know whether they tend not to with respect to objectivity. See (Kumar 2015b) for further discussion, including objections that arise from Aharoni et al. (2012).

a precondition for Darwinian selection. As Boyd argues, organisms from a single species are members of the same natural kind because they share a cluster of morphological, physiological, and behavioral properties (Boyd 1991: 142–143). Atypical and even borderline members of a species arise due to normal phenotypic variation.

MORAL is a complex concept, but its constituents do not provide conditions that are individually necessary and jointly sufficient for application of the concept. Thus, MORAL does not mesh with the “classical” view of concepts that characterizes them in terms of necessary and sufficient conditions (see Laurence and Margolis 1999). Rather, MORAL is better characterized by “prototype” theory (Rosch 1978) or one or another “theory” theory (Carey 1985; Keil 1989). Further empirical work is needed to explore whether the predictions of either theory of concepts, when applied to MORAL, are borne out (see, e.g., Margolis and Laurence 2014 for the sort of evidence relevant to testing between them).

Let’s sum up. Empirical research supports a theory—I will call it “MCT”—according to which distinctively moral judgments encode a concept of morality that is composed of four features that are linked through homeostasis: serious, general, authority-independent, objective. The relationship between MORAL and the four features is analytic, in some sense, but it is one that I have attempted to establish empirically rather than a priori (see fn. 4). The object of MCT is a psychological concept, but my methodology is not that of traditional conceptual analysis. Experimental study of the mind seems able to reveal the structure of human concepts. The relationship that the four constituent features bear to one another is, however, nomological, rather than analytic, and this is what explains why moral judgment is a unified natural kind.

5 Empirical criticism

In this section and the next I will examine two lines of criticism, the first empirical and the second armchair. In both cases I will show that MCT is poised to withstand objections. But I will not always take a hard line. Some objections, while inconclusive, call for further research that would more decisively settle the issue.

According to MCT, the concept of morality is defined in part through a contrast with convention. Kelly, Stich, and colleagues argue, however, that the entire body of research on the moral/conventional distinction is misleading (Kelly et al. 2007; Kelly and Stich 2007; Nado et al. 2009; see also Stich 2006). My aim in this section is to respond to their attack, and thus disarm objections to MCT and to the view that moral judgment is a natural kind.

Kelly and Stich’s target view is narrower than MCT, isolating only three of the four features—serious, general, and authority-independent. However, this view and MCT are similar in that both claim that the components of moral judgment make up a homeostatic cluster. Kelly and Stich’s main criticism is that the features do not cluster—that the moral/conventional pattern is merely an artifact of researchers’ narrow focus. When exposed to scrutiny the pattern is disrupted. If each feature is as likely as not to correlate with the rest, then the alleged homeostatic cluster is just a

hodgepodge. People make a range of different normative judgments, the authors argue, and some judgments fit the moral pattern closely but many others fit only in part. It would seem to follow that the theory of moral judgment derived from experimental research does not pick out a natural psychological kind.

Support for Kelly and Stich's argument comes from a pair of studies. Kelly et al. (2007) themselves find that some harmful transgressions are judged to be seriously wrong in the here and now, but permissible in other places and times, suggesting that the wrongs do not generalize. For example, subjects tend to think that it is seriously wrong for navy officers to whip their subordinates nowadays, but that it is permissible for them to have done so 300 years ago.⁹ In another study, Nichols (2004a) probes subjects' reactions to behavior that elicits disgust. Subjects in Nichols' study judge that the behavior is seriously wrong, fitting the profile for moral judgments. However, they also judge that the wrongness is not general or authority-independent, fitting the profile for conventional judgments. From these two studies Kelly and Stich conclude that the features are not a homeostatic cluster.

Kelly and Stich's argument suffers from three main flaws (see also Sousa 2009). First, it is plausible, as a descriptive claim, that people subscribe to different moral norms that sometimes conflict. And so, Kelly et al.'s subjects may believe that moral norms concerning harm do generalize, but that in some places and times other norms are overriding. For example, while it is *prima facie* wrong for navy officers to whip their subordinates, perhaps the average navy crewman 300 years ago required harsher discipline to be kept in line. Subjects might believe that officers were permitted or even obliged to override the norm prohibiting harm for the sake of order. Thus, the study does not show that norms at play are not conceived as general; the results are consistent with the norms being general but in some places and times being overridden.

Second, recalling discussion at the end of Sect. 3, MCT defines morality as general in contrast with convention. But morality need not apply universally to all beings in order to be more general than convention. If morality generalizes across many cultures, whereas convention applies only within local communities, there would yet be a difference between morality and convention with respect to their scope. So, although Kelly et al.'s findings may show that morality is not conceived as universal, they do not show that morality is conceived as no more general than convention.

Finally and most importantly, the moral/conventional pattern is not supposed to be exceptionless. MCT claims that the features usually cluster together and that there is an explanation for this in terms of the structure of our cognitive system. What Kelly and Stich might use to challenge this thesis is a certain kind of statistical evidence, for example, that the features do not cluster together significantly more often than not. But a few deliberately chosen cases in which the components come apart is quite a long way from disconfirming evidence. In fact, because so many

⁹ Kelly et al. also tested whether subjects would judge that certain actions would still be judged wrong if relevant authorities condoned them. However, their findings with respect to authority independence are much less powerful. With respect to many of the scenarios, subjects judged that wrongness of harmful actions *is* authority independent (Kelly et al. 2007: 127–128).

researchers have studied the moral/conventional distinction, and because their findings have been replicated so often across a broad range of populations, evidence available so far supports the claim that the features do form a mutually supporting cluster. The studies by Smetana on children and Blair on psychopaths reviewed in the last section provide further evidence. Kelly et al.'s findings show, at most, only that homeostasis can be disrupted and do not outweigh the preponderance of evidence that supports the existence of homeostasis.

I have responded to Kelly and Stich's main complaint that the features do not cluster, but the authors also advance a second line of criticism alleging that research on the moral/conventional distinction lacks ecological validity, i.e., that it does not study the full range of moral judgments that people make in ordinary life. First of all, core research on the moral/conventional distinction presents subjects with various types of moral and conventional violations. But, as Kelly and Stich observe, nearly all of the moral violations are those that fit a Western, liberal conception of right and wrong. Specifically, the violations tend to involve harm and unfairness. The authors suggest that among conservatives and non-Westerners morality includes quite a bit more besides this. For many people morality is also about loyalty to one's clan, honoring one's elders, and keeping oneself pure from defilement (Haidt and Joseph 2007). Of course, we need not accept that one genuinely ought to remain "pure" to acknowledge that many people see purity as a moral imperative. Thus, Kelly and Stich claim that research on the moral/conventional distinction is ethnocentric and unlikely to yield universal conclusions about moral judgment.

Kelly and Stich are right that further research is needed to determine whether other types of moral violations are also treated as serious, general, authority-independent, and objective. Still, the objection only takes us so far. Harm and unfairness are central to morality, even for conservatives and non-Westerners. Arguably, many of the most weighty moral issues that people deal with in their ordinary lives concern harm and unfairness. So, research on the moral/conventional distinction *does* seem to probe representative moral judgments. Nonetheless, it could do better. Specifically, a version of the moral/conventional task that probes subjects' reactions to, say, violations of loyalty, honor, and purity would be valuable. Defense of MCT depends not just on properly interpreting currently available empirical evidence, but identifying future research that would confirm or disconfirm the theory. If people continue to treat a broader class of moral violations as serious, general, authority-independent, and objective, then MCT gains more credibility. But if they don't, then MCT is in trouble.¹⁰

¹⁰ Kelly and Stich offer another, similar objection that does not in fact apply to MCT. Their target view states that transgressions that involve harm, injustice or rights-violations elicit moral judgments; transgressions that do not involve harm, injustice or rights-violations elicit conventional normative judgments. Kelly and Stich argue persuasively that both generalizations are false. However, MCT does not have any substantive implications about what sorts of violations are moral. MCT is a view about how we conceive of violations that are moral, whatever those violations happen to be. Kelly and Stich build harm/injustice/rights into their target theory of moral judgment. But which properties elicit moral judgments varies significantly across cultures. Whatever people regard as morally relevant, though, MCT entails that these factors elicit judgments that conceptualize morality in terms of the four features.

Kelly and Stich also present a separate but related objection. They point out that the moral/conventional task was originally used to test children's moral competence, and thus contains only "schoolyard" transgressions. "[T]he examples of harmful transgressions studied were all behaviors that would be familiar to youngsters, such as pulling hair or pushing someone off a swing" (Kelly et al. 2007: 121). For this reason, too, the authors claim, the research does not assay typical moral judgments. In a recent study by Bryce Huebner et al. (2010), however, researchers employed a wider range of transgressions appropriate to adults. The results confirmed the existence of the moral/convention distinction. Subjects tended to treat adult-appropriate moral violations, but not adult-appropriate conventional violations, as serious, general, and authority-independent.

In sum, Kelly and Stich offer two main empirical objections, but these objections do not warrant rejection of MCT. First, the experimental findings that they cite do not show that subjects understand morality and convention differently with respect to generality, objectivity, etc. So, the findings do not disrupt the moral/conventional pattern. Moreover, even if they do disrupt the pattern, MCT allows for exceptions on principled, naturalistic grounds. Second, Kelly and Stich argue that violations tested in the moral/conventional task are unrepresentative, both because they reflect only a narrow subset of moral norms and because they involve only schoolyard transgressions. The one challenge has been met head on and the other is less powerful than Kelly and Stich suppose. More research is needed that explores people's responses to a wider class of moral wrongs. Nonetheless, harm and unfairness are of central moral concern across all cultures, moral judgments about them representative.

6 Armchair criticism

MCT is resilient against empirical challenges, but it also faces philosophical challenges leveled from the armchair. Sinnott-Armstrong and Wheatley (2014) argue that a view based on the moral/conventional distinction is falsified by counterexamples, i.e., normative judgments that lack one of the features, but are intuitively still classified as moral judgments.¹¹ In this section I will consider these authors' cases along with another putative counterexample to MCT. My interest here is with philosophers' intuitions about cases, allowing for the sake of argument that the intuitions are trustworthy.

Before we begin, it is important to be clear about what type of theory MCT is. MCT defines the concept of morality in terms of a cluster of four features. Typical moral judgments encode all four features, but atypical moral judgments encode only three of the four. Thus, MCT allows for atypical cases. And so, the view cannot be refuted simply by pointing to cases of moral judgment in which one of the features are absent. These cases are not genuine counterexamples because the theory, as

¹¹ Another type of counterexample is normative judgments in which all four features are present, but that are intuitively classified as non-moral. These are more difficult to produce.

formulated, does not entail that they fall outside the moral domain. Objections must confront MCT on its own terms, and the view entails that there can be natural variation around central cases of moral judgments, variation of a type that is frequently exhibited by other natural kinds. Still, in the rest of this section I will show that atypical cases of moral judgment are more difficult to produce than critics suppose.

Sinnott-Armstrong and Wheatley argue that moral judgment is not a unified psychological category. They examine a number of views that seek to unify moral judgment and find each of them wanting. One view is based on the moral/conventional distinction. Sinnott-Armstrong and Wheatley's main objection is that people see some issues as authority-dependent, but nevertheless seem, intuitively, to treat them as moral issues. Imagine, they say, that a teacher tells a child to lie as part of an elaborate lesson to her students about trust. Intuitively, some people might judge that it is morally permissible for the child to lie. Thus, lying, though a moral issue, seems to be regarded as authority-dependent.

However, cases like this do not in fact show that people see morality as authority-dependent. Two elements of the conceptual content of moral judgment must be distinguished. A moral judgment about a duty, say, represents the duty as moral (having four features) and it also represents the object of the duty—what we have a duty to do. In Sinnott-Armstrong and Wheatley's case, it seems, people judge that there is a duty to follow the commands of those in authority, and that this duty overrides a prohibition against lying. But this duty can still be authority-independent since following the commands of those in authority is not itself authority-dependent. To see this note that people would likely judge that disobeying the teacher is (*prima facie*) wrong no matter what other authority figures say. Thus, obeying the teacher is authority-independent in the sense that matters. What is authority-dependent is what the norm commands, not whether it ought to be followed.

Although Sinnott-Armstrong and Wheatley do not discuss it, moral judgments concerning purity might also seem to be counterexamples to the claim that morality is inherently social—a claim needed to derive the four features distinguishing morality from convention—since purity norms seem to be sometimes only self-regarding. Some purity norms are clearly social, like those having to do with chastity. But others seem asocial, like those requiring that one keep oneself pure from defiled, inanimate objects. Once again, purity may be illegitimate, moral judgments about purity simply false, but what matters here is that people seem to regard purity as a moral issue. This is an apparent challenge to MCT's claim that morality is conceptualized as social.

My response to this argument will rely on previous discussion. In Sect. 2 I argued that other normative categories are self-regarding, e.g., “ethics”, but self-regarding normative categories are moral only insofar as they are covertly social. Moral purity is, I believe, covertly social, because it regulates social interaction. Someone who does not keep herself pure is not accorded full and equal status in the relevant community because she threatens contamination. Contamination is a social phenomenon. Purity norms, furthermore, engage reactive interpersonal attitudes, not resentment or indignation, typically, but moral disgust and repugnance (see

Rozin et al. 1999; Cannon et al. 2011). All of this suggests that moral judgments about purity are in fact social.¹²

MCT does not stand or fall with my responses to putative counterexamples. MCT articulates a cluster concept of morality and it entails that atypical moral judgments encode only three of the four features. Thus, the view is couched in terms that shield it from certain kinds of objections. However, it does make predictions that can be falsified. In particular, it makes the empirical prediction that the four features are in homeostasis. As I argued, evidence from Smetana and Blair supports this prediction and the evidence that Kelly and Stich cite does not undermine it. MCT also entails that judgments that lack several of the four features (instead of just one) are not moral judgments, that is, are neither typical nor atypical moral judgments. So, if clearer counterexamples could be produced, in which judgments that are intuitively classified as moral lack several of the four features, that would count against the view. In the absence of disconfirmation, I conclude that MCT is empirically and conceptually well supported. It has taken quite a bit of empirical investment to appreciate this, but the theory has philosophical payoffs. MCT accounts for the distinctive nature of moral judgment, and, as we'll now see, it also accounts for the possibility of moral agreement and disagreement.

7 Agreement and disagreement

G. E. Moore (1912: 89–93) famously argued that some theories of moral judgment do not account for genuine moral disagreement. According to one kind of subjectivist view, for example, moral judgments are reports of one's preferences.¹³ So, when I say that an action is morally wrong I say that I prefer not to perform the action. But then people who make seemingly opposing claims about the moral status of an action do not actually disagree. I say that violent protests against systemic injustice are right, you say that they are wrong. But, according to subjectivism, I say that I prefer violent protests and you say that you do not prefer them. There is no disagreement here, just as there is no disagreement if I say that I prefer I chocolate ice cream and you say that you prefer vanilla.

In moral disagreement two people must have attitudes that oppose one another. But to disagree, or even simply agree, they must also conceive of the issue at play as

¹² It may seem as if further counterexamples can be drawn from metaethical theories that reject one or another of the four features. Extreme relativists deny that morality is any more general than convention. Extreme subjectivists deny that morality is any more objective than convention. However, neither metaethical theory generates intuitive counterexamples—neither view yields concrete cases of judgments that are intuitively classified as moral. It may be true that, intuitively, some moral issues are not completely general: they generalize only across some cultures, rather than across the entire spectrum of human societies (see Sarkissian et al. 2011). But that still allows for a difference between morality and convention, which is thought to generalize only across local communities. Extreme forms of relativism and subjectivism are revisionary theories. Unlike, say, moral realism, these views revise our pre-theoretical concept of morality. Norms that lack all generality or objectivity are not intuitively classified as moral, and therefore are not candidate counterexamples.

¹³ Subjectivism is not the expressivist view that moral judgments *express* one's preferences, of course. See Stevenson's (1942) reply to Moore.

moral. Otherwise accord and discord are not distinctively moral. For example, imagine that one person is morally opposed to homosexual behavior and another thinks it merely unseemly. There is a sort of agreement in general attitude here, but no *moral* agreement. Conversely, two people might agree that they ought not to join the war effort, all things considered, perhaps because of overwhelming prudential considerations, and yet disagree about whether joining is *morally* the right thing to do. Genuinely moral agreement and disagreement require shared moral concepts.

As I explained in Sect. 3, the moral/conventional distinction is universal or at least close to universal. Individuals from a wide range of groups exhibit the standard pattern of responses in the moral/conventional task, including adults from many different cultures, as well as children from normal and abnormal populations. Thus, MCT explains why people not just from the same local group but from many different demographic groups genuinely agree and disagree when they voice moral opinions, and thus are not merely talking past one another. They have a shared concept of morality—as (prototypically) serious, general, authority-independent, and objective. Thus, when business owners in one country and workers in another country disagree about, say, the morality of factory conditions, they disagree—genuinely—about whether correct social standards that have the four features permit or forbid the conditions.¹⁴

Besides being empirically warranted, then, MCT has another major advantage, namely, that it accounts for moral agreement and disagreement. Moral interlocutors genuinely agree or disagree with one another because they share a concept of morality. A full exploration of this issue would require a separate inquiry devoted solely to it. Nonetheless, in the rest of the essay I will canvass alternative views and argue that they face serious difficulties. Either the views entail that moral interlocutors consistently talk past one another, or they do not enjoy the empirical support that any view on this issue must seek.

Consider first substantive views of moral concepts. According to these views, moral judgments are different from other normative judgments because moral judgments are about maximizing happiness, or they are about respect for persons, or they are about exemplifying virtue, etc. Most utilitarians, Kantians and virtue theorists do not accept a view like this, of course—only those that build substantive conditions into the very concept of morality. The problem with substantive views is that they cannot account for agreement and disagreement between people who disagree at the level of normative theory. A Utilitarian and a Kantian who seem to agree that solitary confinement is morally wrong are, then, not really agreeing because one claims that solitary confinement does not maximize happiness and the other that solitary confinement does not respect persons.

¹⁴ What about people who make atypical moral judgments? (Thanks to Bryan Chambliss for raising this question.) MCT entails vagueness not just about whether people make moral judgments but also about whether people are disagreeing. When one or more parties in a moral discussion are making an atypical moral judgment they are “more or less” disagreeing, just as two people who have distinct but overlapping concepts of “game” may more or less disagree about whether something counts as a game. In a few such cases, the disagreement is not genuine because one person is concerned especially with whether an action has one of the features that is not tokened in the other person’s judgment.

Of importance here is the distinction between formal and substantive matters—between the concept of morality and what things the concept applies to. Externalist theories of moral content, however, offer an account of the one in terms of the other (Boyd 1988; Brink 1989). For externalists, the meaning of a moral term is just the property that the term refers to; reference is fixed not by an internal concept but by a direct causal/historical relation between word and object. Thus, according to externalism, moral interlocutors agree or disagree with one another, no matter how divergent their substantive views, because both refer directly to the same properties. Suppose that the utilitarian represents the claim that solitary confinement is wrong as the claim that it decreases happiness, the Kantian that it fails to respect persons. But both do in fact agree, according to the externalist, because what their claims mean is that solitary confinement does not have the property to which their shared moral terms directly refer. For example, if ‘morally right’ refers directly to the property of maximizing happiness, then the moral judgments of Kantians, despite themselves, have the same content as the moral judgments of Utilitarians.

Over the past few decades, Horgan and Timmons (1991, 1992, 2000) have developed a sustained critique of externalism. Their main argument relies on “moral twin Earth,” an analogue of Hilary Putnam’s (1975) famous thought experiment. Earth and moral twin Earth are identical, except for one difference. On Earth moral terms refer directly to property A. On moral twin Earth the residents’ moral terms refer directly to property B. Thus, when Earthlings and twin Earthlings express seemingly opposing moral judgments, for example, about the morality of dog fighting, they do not in fact disagree. One says that dog fighting has property A, the other that dog fighting doesn’t have property B. Intuitively, Earthlings and twin Earthlings do disagree, but on an externalist semantics they do not. Thus, Horgan and Timmons conclude that moral terms do not refer directly.¹⁵ If they are right, then externalism is unable to account for moral agreement and disagreement.¹⁶

In light of all of this, the concept of morality seems to be formal rather than substantive. Michael Smith (1994) argues for a formal analysis that runs as follows: to say that I morally ought to do something is to say that I would do it if I were rational. Smith intends this view as an account of the internal conceptual content of moral judgments, rather than a synthetic thesis. Thus, when two people disagree about the morality of eating meat they disagree about whether one would eat meat if one were rational.

¹⁵ Some readers may balk at this argument, insisting that they have no intuitions about moral twin Earth and that what must be accounted for is agreement and disagreement in the real world. I am sympathetic to this reaction. However, the basic point does not require any such esoteric thought experiment. In one essay, Horgan and Timmons (2000: 150) deploy a similar though less clean objection using Hare’s (1952: 148–149) classic case of the missionaries and the cannibals. Suppose that each group uses moral terms that refer directly to different properties. Then, when the missionaries and cannibals disagree about the morality of eating human flesh, they are not really disagreeing. But, intuitively, they are. And once again, we have a reason to reject semantic externalism about moral terms.

¹⁶ Externalist theories of moral content have a certain broad affinity with my approach in this essay. Externalists often hold that moral properties are natural kind properties that can be investigated empirically. But the view that moral judgment is a natural kind does not entail that the objects of moral judgment, moral properties, are natural kinds.

The problem with Smith's view is likely to afflict other views constructed from the armchair. There is simply no empirical evidence available to support it. If a philosophical theory makes a claim about the concept of morality, and if people often agree and disagree about moral matters, then the theory makes an empirical prediction about human psychology: that people who agree or disagree share moral concepts. So, if people who make moral judgments are in fact representing objects of moral appraisal as Smith claims, then it is an empirical prediction of his view that people will be disposed to think and act in accord with these representations. Experimental probes analogous to the moral/conventional task should reveal the influence of those representations. To my knowledge no relevant studies have been conducted.

If philosophers commit themselves to views that have empirical consequences, they should be prepared to back them up with appropriate empirical evidence. Some philosophers might defend views like Smith's by appealing to their own intuitions. I am skeptical that the relevant intuitions among the philosophical community are univocal, but even if they are, what we want to explain is moral agreement and disagreement among the population at large, not just among professional philosophers. In this context armchair investigation alone simply will not do.

MCT does account for moral agreement and disagreement in the population at large. No matter how much our substantive moral commitments diverge, no matter whether moral terms refer directly, and no matter whether we draw a link between morality and rationality, we share a concept of morality as serious, general, authority-independent, and objective. MCT identifies the distinctive conceptual content of moral judgments. The theory provides an empirically compelling and philosophically defensible, working hypothesis about the nature of moral judgment as a natural kind.

Acknowledgments For helpful feedback I am grateful to an anonymous referee for *Philosophical Studies*, Alex Barber, Michael Bukoski, Richmond Campbell, Bryan Chambliss, John Doris, Terence Horgan, Ben Kearns, Ron Mallon, Peter Railton, Stephen Stich, Mark Timmons, and especially Shaun Nichols. Thanks also to audiences at an annual meeting of the Canadian Philosophical Association, the Moral Psychology Research Group, and the University of Arizona for stimulating discussion.

References

- Aharoni, E., Sinnott-Armstrong, W., & Kiehl, K. (2012). Can psychopathic offenders discern moral wrongs: A new look at the moral/conventional distinction. *Journal of Abnormal Psychology, 121*(2), 484–497.
- Alicke, M. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 26*, 556–574.
- Beebe, J., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind and Language, 25*(4), 474–498.
- Bird, A., & Tobin, E. (2012). Natural kinds. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/win2012/entries/natural-kinds/>
- Blair, J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*, 1–29.
- Blair, J. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders, 26*, 571–579.
- Blair, J. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences, 26*, 731–739.

- Blair, J., Jones, L., Clark, F., & Smith, M. (1995). Is the psychopath 'Morally Insane'? *Personality and Individual Differences*, 19(5), 741–752.
- Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Ithaca: Cornell University Press.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays*. Cambridge, MA: MIT Press.
- Brandts, J., & 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: Cambridge University Press.
- Campbell, R. (1998). *Illusions of paradox*. New York: Rowman and Littlefield.
- Campbell, R. (2007). What is moral judgment. *Journal of Philosophy*, 104(7), 321–349.
- Campbell, R. (2009). The Origin of Moral Reasons. In L. Johansson, J. Osterberg, & R. Sliwiniski (Eds.), *Logic, ethics and all that Jazz: Essays in honour of Jordan Howard Sobel*. Uppsala: Uppsala University Press.
- Campbell, R., & Kumar, V. (2012). Moral reasoning on the ground. *Ethics*, 122(2), 273–312.
- Campbell, R., & Kumar, V. (2013). Pragmatic naturalism and moral objectivity. *Analysis*, 74, 446–455.
- Cannon, P., Schnall, S., & White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science*, 2, 325–331.
- Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.
- Copp, D. (1995). *Morality, normativity and society*. Oxford: Oxford University Press.
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2005). Moral appraisals affect doing/allowing judgments. *Cognition*, 108, 353–380.
- Cushman, F., & Mele, A. (2008). Intentional action: two-and-a-half folk concepts? In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 171–188). Oxford: Oxford University Press.
- Fehr, E., & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8, 185–190.
- Fehr, E., & Fischbacher, U. (2004b). Third party sanctions and social norms. *Evolution and Human Behavior*, 25, 63–87.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90, 980–994.
- Fischbacher, U., Gächter, S., & 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.
- Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Goodwin, G., & Darley, J. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106, 1339–1366.
- Griffiths, P. (1997). *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.
- Haidt, J., & Joseph, C. (2007). "The moral mind: how 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 367–391). Oxford: Oxford University Press.
- Hare, R. M. (1952). *The language of morals*. Oxford: Oxford University Press.
- Harman, G. (1967). Quine on meaning and existence, I. *Review of Metaphysics*, 21, 124–151.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106, 587–612.
- Hollos, M., Leis, P., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, 17, 352–376.
- Holton, Richard. (2009). *Willing, wanting, waiting*. Oxford: Oxford University Press.
- Horgan, T., & Timmons, M. (1991). New wave moral realism meets moral twin earth. *Journal of Philosophical Research*, 16, 447–465.
- Horgan, T., & Timmons, M. (1992). Troubles on moral twin earth: Moral queerness revived. *Synthese*, 92, 221–260.
- Horgan, T., & Timmons, M. (2000). Coping out on moral twin earth. *Synthese*, 124, 139–152.

- Huebner, B., Lee, J., & Hauser, M. (2010). The moral/conventional distinction in mature moral competence. *Journal of Cognition and Culture*, *10*, 1–26.
- Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Kelly, D., & Stich, S. (2007). Two theories of the cognitive architecture underlying morality. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 348–366). New York: Oxford University Press.
- Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, T. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language*, *22*, 117–131.
- Kennett, J., & Fine, C. (2008). Internalism and the evidence from psychopaths and ‘acquired sociopaths’. In Walter Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 3, pp. 173–190). The neuroscience of morality: Emotion, brain disorders, and development Cambridge, MA: MIT Press.
- Keser, C., & 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*, 190–193.
- Knobe, J. (2003b). Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, *16*, 309–324.
- Knobe, J. (2010a). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*, 315–365.
- Knobe, J. (2010b). Action tree and moral judgment. *Topics in Cognitive Science*, *2*, 555–578.
- Knobe, J., & Burra, A. (2006). Intention and intentional action: A cross-cultural study. *Journal of Culture and Cognition*, *6*, 113–132.
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 2, pp. 441–448). Cambridge, MA: MIT Press.
- Kohlberg, L. (1963). The development of children’s orientation toward a moral order: I. sequence in the development of moral thought. *Vita Humana*, *6*, 11–33.
- Kohlberg, L. (1981). *Essays on moral development* (Vol. 1). The philosophy of moral development New York: Harper Row.
- Kornblith, Hilary. (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*, 439–457.
- Kumar, V. (2015a). The empirical identity of moral judgment. Under review.
- Kumar, V. (2015b). Psychopathy and internalism. Under review.
- Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). Cambridge, MA: MIT Press.
- Leslie, A., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science*, *17*, 421–427.
- Levy, N. (2007). The responsibility of the psychopath revisited. *Philosophy, Psychiatry, and Psychology*, *14*(1), 129–138.
- Maibom, H. (2005). Moral unreason: The case of psychopathy. *Mind and Language*, *20*(2), 237–257.
- Margolis, E. & Laurence, S. (2014). Concepts. In *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), E. N. Zalta (Ed.). <http://plato.stanford.edu/archives/spr2014/entries/concepts/>
- Moore, G. E. (1912). *Ethics*. New York: Henry Holt.
- Nadelhoffer, T. (2005). Skill, luck, control, and folk ascriptions of intentional actions. *Philosophical Psychology*, *18*, 343–354.
- Nado, J., Kelly, D., & Stich, S. (2009). Moral judgment. In J. Symons & P. Calvo (Eds.), *The Routledge companion to the philosophy of psychology* (pp. 621–633). New York: Routledge.
- Nichols, S. (2004a). *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press.
- Nichols, S. (2004b). After objectivity: An empirical study of moral judgment. *Philosophical Psychology*, *17*, 3–26.
- Nichols, S., & Folds-Bennett, T. (2003). Are children moral objectivists? Children’s judgments about moral and response dependent properties. *Cognition*, *90*, B23–B32.
- Nucci, L. (1986). “Children’s conceptions of morality, societal convention and religious prescription. In C. Harding (Ed.), *Moral Dilemmas: Philosophical and psychological reconsiderations of the development of moral reasoning*. Chicago: Precedent Press.
- Nucci, L. (2001). *Education in the moral domain*. Cambridge: Cambridge University Press.
- Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*, 400–407.

- Nucci, L., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts in the Virgin Islands. *Journal of Cross-Cultural Psychology, 14*, 469–487.
- Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind and Language, 24*, 586–604.
- Pillutla, M., & Murnighan, J. (1996). Unfairness, anger and spite. *Organizational Behavior and Human Decision Processes, 68*, 208–224.
- Prinz, J. (2004). *Gut reactions: A perceptual theory of emotions*. Oxford: Oxford University Press.
- Prinz, J. (2007). *The emotional construction of morals*. Oxford: Oxford University Press.
- Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, mind and knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.
- Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review, 60*, 20–43.
- Railton, P. (1986a). Moral realism. *Philosophical Review, 95*, 163–207.
- Railton, P. (1986b). Facts and values. *Philosophical Topics, 14*, 5–31.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574–586.
- Sarkissian, H., Park, J., Tien, D., Wright, J., & Knobe, J. (2011). Folk moral relativism. *Mind and Language, 26*, 482–505.
- Saver, J. L., & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*, 1241–1249.
- Shweder, R., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children*. Chicago: University of Chicago Press.
- Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology, 27*, 451–474.
- Smetana, J. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52*, 1333–1336.
- Smetana, J. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology, 21*(1), 18–29.
- Smetana, J. (1993). Understanding of social rules. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist*. New York: Guilford Press.
- Smetana, J., & Braeges, J. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly, 36*, 329–346.
- Smetana, J., Kelly, M., & Twentyman, C. (1984). Abused, neglected, and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development, 55*, 277–287.
- Smith, M. (1994). *The moral problem*. Malden, MA: Blackwell.
- Snarey, J. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin, 97*(2), 202–232.
- Song, M., Smetana, J., & Kim, S. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology, 23*, 577–582.
- Sousa, P. (2009). On testing the 'moral law'. *Mind and Language, 24*(2), 209–234.
- Southwood, N. (2011). The moral/conventional distinction. *Mind, 120*, 761–802.
- Stevenson, C. L. (1942). Moore's arguments against certain forms of naturalism. In P. Schlipp (Ed.), *The Philosophy of G. E. Moore*. Evanston: Northwestern University Press.
- Stich, S. (2006). Is morality an elegant machine or a Kluge? *Journal of Cognition and Culture, 6*, 181–189.
- Sturgeon, N. (1988). Moral explanations. In G. Sayre-McCord (Ed.), *Essays on moral realism*. Ithaca: Cornell University Press.
- Tisak, M. (1995). Domains of social reasoning and beyond. In R. Vasta (Ed.), *Annals of child development* (Vol. 11). London: Jessica Kingsley.
- Tisak, M., & Turiel, E. (1984). Children's conceptions of moral and prudential rules. *Child Development, 55*, 1030–1039.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
- Turillo, C., Folger, R., Lavelle, J., Umphress, E., & Gee, O. (2002). Is virtue its own reward? Self sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes, 89*, 839–865.

- Wainryb, C., Shaw, L., Langley, M., Cottam, K., & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development, 75*, 687–703.
- Weiskopf, D. (2009). The Plurality of Concepts. *Synthese, 169*, 145–173.
- Williams, B. (1985). *Ethics and the limits of philosophy*. London: Routledge.
- Schroeder, T. (2004). *Three faces of desire*. Oxford: Oxford University Press.
- Yau, J., & Smetana, J. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development, 74*, 1647–1658.