

Psychopathy and internalism

Victor Kumar

Philosophy Department, University of Toronto, Toronto, Canada

ABSTRACT

Do psychopaths make moral judgments but lack motivation? Or are psychopaths' judgments are not genuinely moral? Both sides of this debate seem to assume either externalist or internalist criteria for the presence of moral judgment. However, if moral judgment is a natural kind, we can arrive at a theory-neutral criterion for moral judgment. A leading naturalistic criterion suggests that psychopaths have an impaired capacity for moral judgment; the capacity is neither fully present nor fully absent. Psychopaths are therefore not counterexamples to internalism. Nonetheless, internalism is empirically problematic because it is unable to explain psychopaths' moral deficits.

ARTICLE HISTORY Received 6 April 2015; Accepted 10 March 2016

KEYWORDS Psychopaths; moral judgment; motivation; internalism; natural kind; moral/conventional distinction; moral concepts

Psychopathy is a disorder characterized by egocentric personality, diminished sympathy, lack of guilt and remorse, impulsivity, and anti-social behavior (Hare 1993). Psychopaths exhibit a range of affective and motivational deficits that disrupt aspects of their moral agency. But do they retain the capacity for moral judgment? If they do have the capacity, then psychopaths appear to be real-life counterexamples to the internalist theory that binds moral judgment to motivation. Research on psychopathy, it seems, has the potential to empirically disconfirm a philosophical theory about the nature of moral judgment.

However, neither critics nor defenders of internalism have addressed the worry that their shared approach inevitably begs the question. I will argue that research on psychopathy *can* be used to evaluate internalism given a naturalistic approach to philosophical inquiry. After laying out this naturalistic approach, I will outline a general theory of moral concepts that is based on a rich and extensive body of empirical research on the 'moral/conventional distinction' (Kumar 2015). Studies that probe psychopaths' grasp of moral concepts suggest that they do not clearly possess, nor do they clearly lack, the capacity for moral

CONTACT Victor Kumar  victor.c.kumar@gmail.com

© 2016 Canadian Journal of Philosophy

judgment: the capacity is impaired without being fully present or fully absent. It turns out, then, that psychopaths are not straightforward counterexamples to internalism – not clear cases of people who make moral judgments but lack corresponding motivation. Nonetheless, I will argue that internalist accounts of moral judgment have difficulty accounting for psychopaths' peculiar combination of moral deficits. Thus, in psychopaths we find empirical evidence that challenges a philosophical theory about the nature of moral judgment.

1. Circularity

Motivational internalism is the view that necessarily, if a person makes a moral judgment he or she is motivated to act as the judgment prescribes. If Sonia judges that she morally ought not to eat meat that guarantees that she is motivated, at least to some degree, not to eat meat. Whether motivation is necessary for moral judgment is interesting in itself, but the question also has broader significance. As Michael Smith (1994) argues, internalism, cognitivism, and Humeanism form a jointly inconsistent triad. If moral judgments are necessarily motivating, then either moral judgments are not beliefs or beliefs alone can necessitate motivation. Furthermore, internalism seems to conflict with certain forms of moral realism (e.g. Railton 1986; Brink 1989). Mental states that are tied necessarily to motivation seem not to represent mind-independent facts or properties (cf. Tresan 2006).

Internalism is traditionally interpreted as a conceptual thesis: the concept of moral judgment is the concept of a mental state that entails the presence of corresponding motivation. Externalist critics argue against the existence of a conceptually grounded, necessary link between moral judgment and motivation by appealing to the conceivability of 'amoralists' (Stocker 1979; Brink 1989, 45–60). Intuitively, externalists claim, we can imagine people who judge things right and wrong, but simply have no desire to be moral and therefore lack moral motivation.

Internalists respond by voicing a conflicting intuition, that the judgments of so-called amoralists are not genuinely moral (Smith 1994, 68–71). Often, internalists claim that the people externalists call to mind form moral judgments only in an 'inverted commas sense' (Hare 1952, 145–146). So-called amoralists judge that actions are 'morally wrong,' perhaps, no more than some atheists judge that religious artifacts are 'sacred.'

The debate seems to have reached a stalemate and it is tempting to conclude that intuitions on one or both sides are covertly driven by the theories that the case is designed to test (see Cholbi 2006; Levy 2007; Kauppinen 2008; Kennett and Fine 2008b; Prinz 2015). Externalists are able to conceive of amoralists because their theory of moral judgment is externalist. Internalists are unable to conceive of amoralists because their theory is internalist.

Research on psychopaths and their disordered moral agency promises to overcome this clash of intuitions. Psychopaths are empirical test cases for internalism. Because they have severe affective deficits and regularly and casually flout moral prohibitions when doing so is in their perceived self-interest, it is likely that psychopaths lack moral motivation. Whereas intuitions about amoralists fail to be univocal, empirical studies may decide in a more objective way whether psychopaths' moral judgments are genuine or ersatz. And so, it seems, the challenge to internalism from amoralists can be renewed by empirical research. Studies of psychopaths may provide externalists with real-life counterexamples to supplant their contested hypothetical cases.

The apparent problem with this naturalistic approach is that it cannot avoid begging the question against internalism. Internalists insist that moral judgment entails motivation in virtue of our concept of moral judgment, in which case anyone discovered to lack moral motivation simply does not count as making a moral judgment. To leave open the possibility that unmotivated psychopaths do make moral judgments is to deny the conceptual link, and thus to assume that internalism is false.

Conversely, externalists may argue there can be empirical evidence about whether or not psychopaths make moral judgments only because externalism is correct. Evidence even for the internalist hypothesis that psychopaths do not make moral judgments would be of little significance, according to externalists, because that discovery likewise requires leaving conceptually open the possibility of moral judgment without motivation (see Kauppinen 2008). Any empirical argument that psychopaths have or lack a capacity for moral judgment therefore presupposes externalism and cannot support it.

Smith (1994, 63–66) and Jackson (1998, 29–31) offer a precise characterization of the more general problem. Both authors distinguish conceptual claims about what it is to be F, on the one hand, from empirical or substantive claims about whether a given object is F, on the other. We must first know what it is to be F, they argue, before we can determine whether something is F. As Smith says:

Suppose we are interested in whether or not there are any witches. How are we to go about answering our question? First we must ask a *conceptual question*. What is our concept of a witch? ... Then, second, we must ask a *substantive question*. That is, having now fixed on what our concept of a witch is, we must ask whether there is anything in the world instantiating our concept of a witch. (Smith 1994, 64; emphasis in original)

Under these terms of inquiry, conceptual questions about moral judgment must be answered before any related empirical questions can be. Analysis of the concept of moral judgment and its putative link with motivation must precede investigation of whether psychopaths make moral judgments but lack motivation.

Now, some philosophers who pursue experimental approaches believe that conceptual questions are best answered using empirical methods. Experimental

philosophers suggest that if we want to understand a concept we should experimentally probe people's judgments. Thus, to understand whether moral judgment entails motivation, we might design surveys that assay the concept. For example, several researchers construct vignettes describing psychopaths and their motivational deficits; they then ask research participants to report whether they think the psychopaths' judgments are genuinely moral (see Nichols 2004a, ch. 3; Strandberg and Björklund 2013; Björnsson et al. 2014). Lacking philosophical commitments to internalism or externalism, participants' responses are unlikely to be theory driven. This sort of experimental approach does not presuppose that moral judgment is conceptually separable from motivation, since it aims to test that very idea. So, it does not beg the question against internalism.¹

An experimental approach, however, faces a general difficulty: distinguishing *competence* from *performance*. How do we know that participants' responses to surveys reflect their competence with the concept of moral judgment, as opposed to other psychological factors that distort their responses? Thus, it may be that people have a concept of moral judgment that is tied necessarily to motivation, but that extraneous features of cases lead them to make judgments about psychopaths that depart from their own conceptual competence. Arguably, the back-and-forth structure of philosophical dialectic – absent, of course, in surveys – minimizes such performance errors.

If philosophical study of psychopaths must confront a more basic conceptual question, and if this conceptual question cannot be answered empirically in the way that experimental philosophers favor, then we cannot use psychopaths to sidestep the original clash of intuitions about amoralists. Further a priori reflection upon amoralists may or may not ultimately break the stalemate. In any case, though, if the objection is sound philosophers must return to their armchairs and attend to more fundamental, conceptual questions about moral judgment. Empirical research on psychopathy cannot advance the internalism–externalism debate.

2. Natural kind

The objection laid out in the last section rests on the traditional view that conceptual truths are epistemically prior to empirical truths – in this case, that in order to know as an empirical matter whether someone forms a moral judgment we must already know as a conceptual matter what it is to form a moral judgment. We must first settle on the satisfaction conditions for our concept of moral judgment, and only then can we investigate whether psychopaths meet those conditions.

But conceptual truths do not always have epistemic priority over empirical truths. We are often rightfully more confident about an object's membership within a category than we are about what defines the category – in other words,

rightfully more confident about the category's extension than its intension. For example, we know that seeing a red apple is an example of a perceptual process, even if we can't define what a perceptual process is. *Moreover*, if a philosophically interesting category is a natural kind, then understanding empirically its extension can help us to grasp the category itself. Thus, if moral judgment is a natural kind (Kumar 2015, Forthcoming), then not only can we answer empirical questions about moral judgment without first analyzing our concept, but answering those empirical questions can shed light on what moral judgment is.

As is well known, a natural kind can turn out to be much different from our prior conception of the kind (Kripke 1980; Putnam 1975; cf. Kumar 2014). To understand what underlies a natural kind we must shed reliance on our prior concept and instead study uncontroversial instances of the kind along with the causal role that scientific research accords to it. Many natural kinds, including some of those studied in psychology, are a species of functional kinds (cf. Kauppinen 2013). What we seek, then, are *the natural properties that account for the kind's presence and causal role*. This general approach is common to philosophical study of many putative natural kinds, such as moral properties (Boyd 1988; Sturgeon 1988; Brink 1989; see also Railton 1986), knowledge (Kornblith 2002; Kumar 2014), and a wide range of other psychological categories, such as belief (Fodor 1981, 1987), emotion (Griffiths 1997; Prinz 2004), desire (Schroeder 2004), intention (Holton 2009), and concepts (Weiskopf 2009).

In general, we have reason to believe that an object of study is a natural kind if it figures in scientific laws or generalizations. It turns out that moral judgment plays a rich causal/explanatory role in psychological generalizations concerning human thought and behavior. How we act and feel is deeply influenced by our moral attitudes, as empirical studies confirm. First of all, moral judgments shape pro-social, anti-social, and punitive behavior (Pillutla and Murnighan 1996; Fehr and Gächter 2000; Keser and van Winden 2000; Brandts and Schram 2001; Fischbacher, Gächter, and Fehr 2001; Turillo et al. 2002; Fehr and Fischbacher 2004a, 2004b). Moral judgments also regulate emotions like anger, guilt, and disgust (Rozin 1997, 1999; Rozin, Markwith, and Stoess 1997; Rozin and Singh 1999; Haidt 2003).

My aim in this essay is not to mount a broad defense of the approach that conceives of moral judgment as a natural kind (see Kumar 2015; Forthcoming), but instead to show that it makes philosophical study of psychopathy and internalism possible. Because there are many uncontroversial cases of moral judgment – mental states that we confidently classify as instances – empirical research enables philosophical inquiry into the underlying nature of moral judgment. A theory of moral judgment can be developed by investigating what is common to uncontroversial instances that enables moral judgment to play its distinctive causal role. Commonsense classification of moral judgments is needed for this sort of inquiry to begin, but empirical evidence can then provide a deeper understanding of the nature of moral judgment (cf. Kauppinen 2008).

By answering empirical questions about moral judgment, then, we can arrive at an answer to the 'conceptual' question about what moral judgment is. The resulting theory that answers this question will not be an analysis of the concept of moral judgment, of course, but an empirically grounded theory about the identity of the natural kind itself.

Internalism postulates a necessary connection between moral judgment and motivation, and this necessity is typically thought to be conceptually grounded. The approach that conceives of moral judgment as a natural kind suggests an alternative interpretation of internalism – as a synthetic, metaphysically necessary truth, as is characteristic of empirical identities (Kripke 1980; see Prinz 2015, 63–64). That moral judgment is motivating is thus on a par with claims like 'gold is an element' or, better, 'perception is encapsulated' and 'representations are structured.'

To be useful in the present context, a theory of moral judgment as a natural kind must be neither internalist nor externalist – it must be neutral with respect to the issue. We can then acquire *independent* empirical evidence about whether psychopaths form moral judgments by investigating whether they possess mental states that play the causal role that is characteristic of moral judgments.

To the objections voiced in the last section, then, we can respond as follows. Both the internalist and externalist claim that any argument from psychopathy – either criticizing or defending internalism – must presuppose externalism. However, psychopaths' mental states may be compared to moral judgment's causal role even if our concept of moral judgment does not already rule in amorality as a conceptual possibility. Without relying on conceptual assumptions, empirical research can indicate that psychopaths possess mental states that play the causal role that is characteristic of moral judgment. If moral judgment is a natural kind, we should infer from this finding that psychopaths do make moral judgments and that internalism misrepresents the nature of that kind. Alternatively, it might turn out that psychopaths' mental states do not play moral judgment's causal role, and thus that psychopaths are not amorality – not counterexamples to internalism. So, evidence that is free from externalist assumptions may undermine internalism, or it may absolve internalism. Backed by the view that moral judgment is a natural kind, research on psychopathy can be injected into the internalism/externalism debate without circularity.

3. Moral concepts

My next task is to outline an empirically grounded theory of moral judgment as a natural kind (see Kumar 2015 for more detail). If psychopaths satisfy criteria for the presence of moral judgment that issue from the theory, then we have reason to reject internalism as a synthetic, metaphysically necessary truth. If psychopaths do not satisfy the criteria, then internalism is safe from the alleged counterexample. In fact, research on psychopathy refuses to yield such clean

results, but, as we'll see later on, the results are nonetheless philosophically significant.

3.1. *Morality and convention*

We need a theory that is neutral with respect to internalism and externalism. A theory about the *attitude* that is constitutive of moral judgment is not helpful for this purpose. If we assume only that moral judgments are (ordinary) cognitive states, then it is obvious that psychopaths have the capacity for moral judgment; no one disputes that psychopaths form beliefs. If we assume instead that moral judgments are non-cognitive states or (non-ordinary) beliefs that are tied up with motivation, however, then we will have already decided that motivation is necessary for moral judgment and, consequently, psychopaths can no longer be used to test internalism.

A theory about the distinctive *conceptual content* of moral judgments is potentially more useful. Both internalists and externalists must make room for moral concepts as constituents of moral judgments. In general, a normative judgment categorizes an object of evaluation as right or wrong, obligatory or forbidden, good or bad, praiseworthy or blameworthy, etc. But in moral judgments, specifically, the concept of morality modifies these other normative concepts: an action is not simply wrong, but *morally* wrong (Kumar 2015). For externalists, moral concepts figure in ordinary beliefs. For internalists, moral concepts figure either in beliefs that are motivational, or in beliefs that are otherwise connected necessarily to motivation, or in non-cognitive mental states. Each of these attitudes must contain moral concepts as part of its content. So, a view of moral concepts is potentially consistent with internalism and externalism, and therefore can avoid settling the issue in advance. There is no guarantee that this is so. For example, Prinz (2007) develops a view of moral concepts in terms of moral emotions, which entails that moral judgments are motivational. However, as it happens, the view of moral concepts that I offer below is consistent with both internalism and externalism.

Moral concepts are probed in a rich and extensive body of research that examines the psychological distinction between morality and convention (see Turiel 1983; Smetana 1993; Tisak 1995; Nucci 2001 for review). In this research psychologists elicit moral judgments and other normative judgments from participants, and then analyze how these different types of judgments influence their responses to further experimental probes. Over and over, researchers find that participants consistently draw a distinction between violations of moral norms and violations of conventional norms.

In what's called the 'moral/conventional task' participants are presented with examples of moral violations and examples of conventional violations – examples about which we are confident participants make moral judgments and conventional normative judgments, respectively. Moral violations include such

actions as lying and assaulting someone without provocation. Conventional violations include such actions as chewing gum in class and wearing inappropriate clothing. After each example, participants are asked a series of questions of the following sort:

Is it wrong?

How seriously is it wrong (on a scale)?

Is it wrong in other places and times?

Is it wrong even if certain authority figures say that it is okay?

Participants consistently exhibit a similar pattern of responses to these questions. Moral violations are judged to be more serious, general, and authority-independent. Conventional violations are judged to be less serious, not general, and authority-dependent. Thus, participants are likely to say that assaulting someone is quite seriously wrong, that it is wrong in other places in times, and that it remains wrong even if authorities say that it is okay. These findings generalize across several populations, extending to individuals from many different cultures (Nucci, Turiel, and Encarnacion-Gawrych 1983; Snarey 1985; Hollos, Leis, and Turiel 1986; Nucci 1986; Song, Smetana, and Kim 1987; Yau and Smetana 2003), as well as to children as young as three years old (Nucci and Turiel 1978; Smetana 1981; Nucci 1986; Smetana and Braeges 1990; Smetana, Kelly, and Twentyman 1984).

What we seek now is an account of moral judgment that is derived from empirical research on the causal role of moral judgment. And research on the moral/conventional distinction reveals something general about its causal role. Moral judgment is causally linked to judgments about how serious the relevant normative issue is, about how it generalizes to other places and times, and about its validity even when authorities disagree. The best explanation for this causal role is that moral judgments encode a concept of morality as serious, general, and authority-independent (Kumar 2015). Thus, what distinguishes morality from other normative domains are formal characteristics (see Hare 1952), rather than any substantive connection to, say, harm or injustice. Those who moralize mere matters of etiquette elevate them beyond their proper significance, scope, and authority.²

I propose, then, that moral judgment is a natural kind and that it is individuated in part through the concept of morality that it encodes. To believe that an action is morally wrong, say, is to conceptualize the wrong as serious, general, and authority-independent. Notice that the moral/conventional task presents participants only with examples of norm violations. However, morality is much broader than that. It also concerns, at a minimum, obligation, virtue, and praise. That is, when I make a moral judgment, I may be judging that an action violates a moral norm, but I may instead be judging that an action is morally obligatory, that a person is morally virtuous, or that she is morally praiseworthy. A reformulation of the moral/conventional task, one that tests judgments of obligation,

virtue, praise, etc. would be valuable, and might either strengthen or weaken the view offered here.³

Nonetheless, the view I propose is that in distinctively moral judgments various normative categories are conceived as serious, general, and authority-independent. For example, what it is to be serious is for prohibitions, obligations, virtues, etc. to have a degree of priority that is not possessed by non-moral prohibitions, non-moral obligations, or non-moral virtues. Suppose we have a moral obligation to prevent harm and a non-moral obligation to obey local customs. When these obligations conflict, seriousness entails that harm takes priority over custom. Or consider the moral virtue of justice and the non-moral virtue of decorum. When these virtues conflict, seriousness entails that justice takes priority over decorum. Similarly, to conceptualize justice, but not decorum, as general and authority-independent is to understand justice, but not decorum, as a character trait that remains virtuous across many places and times and no matter what relevant authorities say.

The elements in the concept of morality constitute moral judgment as a natural kind because they are a homeostatic property cluster (cf. Kelly et al. 2007; Kelly and Stich 2007; Kumar 2015; see also Björnsson and McPherson 2014 for a similar but distinct view). Natural kinds in the special sciences tend to manifest as clusters of properties that are stable and mutually reinforcing, especially in the face of external perturbation (Boyd 1988, 1991). Core research on the moral/conventional distinction offers some evidence that the features cluster together. But other evidence more directly confirms this hypothesis. When participants are told that some unspecified norm violation has one of the features they are likely to infer that it has the others too (Smetana 1985).

This theory about the concept of morality encoded in moral judgments entails that there is no sharp distinction between moral judgments and other normative judgments. Although the three features form a cluster, homeostasis can be disrupted. Judgments that encode only two of the three features are atypical cases; or, alternatively, they may be borderline cases that do not fall clearly inside or clearly outside the natural category of moral judgment. As Boyd (1991, 1999) emphasizes, this kind of fuzziness is characteristic of natural kinds in the special sciences.⁴

3.2. Objections

Moral judgments encode certain violations, obligations, virtues, etc. as serious, general, and authority-independent. However, one may wonder whether distinctive conceptual content could suffice to carve our repertoire of psychological states at its joints. Does the view imply that any class of mental states with distinctive conceptual content counts as a natural kind? For example, I believe that I am in my office and this belief influences my behavior in various ways. Certainly 'office beliefs' are not a natural kind.

Moral judgments differ in an important way from many other mental states with distinctive conceptual content: they enter into psychological explanations at a higher level of generality. Moral judgments influence a wide range of pro-social and punitive behavior, and they also powerfully shape our emotional life. We feel guilt and shame, resentment and repugnance, awe and pride because of the moral judgments we form. The distinctive conceptual content of moral judgments defines moral judgment as a natural kind because it underwrites this general explanatory role. Judgments about serious, general, and authoritative prohibitions, obligations, and virtues evidently have a firm and broad grip on our behavior and our feelings.

Despite the impressive amount of empirical evidence supporting its existence, some authors express skepticism that the moral/conventional distinction is robust and applies across the moral domain (Kelly and Stich 2007; Kelly et al. 2007; Shoemaker 2011). If research on the moral/conventional distinction cannot be used to characterize the causal role of moral judgment in a general way, then the present theory of moral concepts is untenable and many of the arguments in the rest of the essay are unsupported.

Daniel Kelly, Stephen Stich, and colleagues provide empirical fodder for this objection to the moral/conventional distinction (Kelly and Stich 2007; Kelly et al. 2007). The authors argue that the alleged connection between moral judgments and judgments of seriousness, generality, and authority-independence is based on a limited and biased range of cases. Once judgments about other types of moral violations are studied, the authors claim, the connection is disrupted. Thus, the three features are, allegedly, not bound together by homeostasis. Kelly et al. (2007) themselves gave participants novel examples of harmful norm violations and found that they were willing to regard some of the violations as serious but not authority-independent, or serious but not general. For example, participants read one case in which a navy officer whips his subordinate. On the whole, participants judged that the action is seriously wrong, but they also judged that it would not have been wrong 300 years ago, i.e. that the wrong does not generalize.

Kelly et al.'s argument presents a challenge to anyone who hopes to derive philosophical significance from research on the moral/conventional distinction. However, in their study the authors deliberately selected cases in which they hoped to pull the three features apart. These cases are dwarfed by a preponderance of evidence in support of the moral/conventional distinction. A large number of studies evince a pattern in the way that people think about morality and convention. But this pattern need not be exceptionless in order to obtain. Kelly et al. may have shown that, with effort, the pattern can be disrupted, but psychological generalizations of this sort cannot be conceived as immune to disruption in the first place. We have antecedent reason to expect atypical or borderline moral judgments that encode only some of the three features.

The concept of morality does not provide necessary and sufficient conditions, and so does not fit with the so-called 'classical' view of concepts (Laurence and Margolis 1999). Rather, morality seems to be captured more naturally by a cluster theory of concepts (Searle 1958), or prototype theory (Rosch 1978), or theory-theory (Carey 1985; Keil 1989). Because on this view seriousness, generality, and authority-independence are not absolutely essential to morality, the usual philosophical strategy of constructing counterexamples seems out of place. Nonetheless, the view *is* subject to counterexamples in the following way: if the view entails that what seems like a typical moral violation is *atypical*, then it nonetheless seems to get the classification wrong, and thus faces an apparent counterexample, as we'll see next.⁵

Some moral violations do not seem to be very serious. I think that littering is morally wrong, for example, but I am willing to tolerate a friend's littering if it is somewhat onerous for her to find a garbage can, or if objecting to her littering runs against local custom. Do we have here not just a moral violation, but a typical moral violation, and yet one that is not regarded as serious? I don't think so. It seems to me that littering is commonly regarded as an atypical moral violation. Perhaps it ought not to be, but what that means is that people should accord it greater seriousness than they actually do.

Another potential counterexample targets generality. Some actions are understood to be wrong here and now, but not in other places and times. Many people think it is wrong not to have a funeral ceremony for one's beloved parents, but they would not condemn someone from certain other cultures where funeral ceremonies are not customary. Furthermore, that it's wrong in our culture not to have a funeral ceremony seems to be a *typical* moral wrong. What we should say in response to this objection, however, is familiar to discussions of moral relativism. To conceptualize a wrong as moral is to think that there is some general prohibition related to it, but that it is one that can be applied to yield different verdicts in different places and times. So, many people think that, in general, one morally ought to honor one's dead parents, but they will likely admit that how one fulfills that obligation can depend on local resources and customs that influence how the general norm is applied.

Whether the view on offer can stand up to further counterexamples is an open question. However, as a final word, I would counsel against putting too much weight on intuitions about what *seems* like a typical moral violation. This conflicts with the naturalistic methodology suited to investigation of natural kinds. We must begin with a rough and ready classification of paradigm cases of moral judgment in order to empirically investigate what is uniquely common to them. However, commonsense classification about examples that lie in the periphery of a natural kind's extension should, in general, yield to empirical investigation of the properties that define central cases.

In sum, typical moral judgments encode a homeostatically-clustered concept of morality as serious, general, and authority-independent. This concept of

morality helps to ground the distinctive causal role of moral judgments, and thus we can figure out whether psychopaths make moral judgments by investigating whether they possess mental states that play this causal role. Critically, the present view about the constitutive conceptual content of moral judgment does not make any assumptions about the relationship between moral judgment and motivation. The concept of morality may feature in the representational content of mental states that are cognitive and non-motivational, or cognitive and motivational, or non-cognitive. Thus, we are now poised to determine whether psychopaths are amoralists, and thereby contribute to empirical evaluation of internalism. To do this properly, however, it will help to understand previous attempts in this vein along with their shortcomings.

4. Previous attempts

The preceding discussion will serve as a lens through which to critically examine recent attempts to draw conclusions about psychopaths' capacity for moral judgment. Critics and defenders of internalism appeal to many different empirical studies and offer rival interpretations of the same studies. I will suggest, however, that their arguments consistently suffer from one of two fatal flaws. Either they provide evidence that is not independent of the conclusion, by employing criteria for moral judgment that are committed either to internalism or externalism. Or they appeal to evidence that is equivocal, focusing on the presence or absence of psychological factors that are not essential to the capacity for moral judgment.

Consider, first, externalist arguments that psychopaths *do* make moral judgments and thus are counterexamples to internalism. Partly on the basis of his own empirical theory of moral judgment, Nichols (2004a) argues that psychopaths make moral judgments because they recognize that harm is prohibited, even though they do not undergo the affective response that typically accompanies recognition of moral violations: 'one can have knowledge of harm norms, and voice one's disapproval of harming others, even if one has lost the affective response' (Nichols 2004a, 99).

Nichols' argument, however, does not provide independent reasons to reject internalism. A premise in the argument is that moral judgments are beliefs that a certain sort of norm has been violated. However, if he assumes that a belief about norm violation is sufficient for moral judgment, Nichols assumes an externalist theory of moral judgment. For all I have said, Nichols' externalist theory may be well supported. But even if it is, Nichols cannot use it as a premise in an argument that psychopaths are counterexamples to internalism, since it is not independent of the conclusion he draws. As I have made clear, any argument that psychopaths are (or are not) counterexamples to internalism must rely on a criterion for moral judgment that is neutral between internalism and externalism.⁶

Roskies (2003) offers the most detailed externalist treatment in the literature. Her subject is not psychopathy but 'acquired sociopathy,' a condition that shares with psychopathy a similar affective and behavioral profile. But whereas psychopathy is a disorder that emerges early in development, acquired sociopathy is caused by damage in adulthood to the ventromedial (VM) cortex. Roskies argues that VM patients form moral judgments but lack moral motivation, concluding that internalism is false.

However, support for Roskies' claim that VM patients make moral judgments is either equivocal or assumes externalism. Roskies says that VM patients' moral statements match our own (2003, 56–57), that they believe their moral statements are sincere (59, 60), and that their ability to engage in moral reasoning is unaffected (57, 60). But none of these factors is essential to moral judgment – not in the same way that its constitutive attitude or conceptual content are. VM patients might engage in reasoning that meets social expectations, but now fails to produce genuine moral attitudes with moral content. They might even believe themselves to be sincere, but lack the mental states that, unbeknownst to themselves, are conventionally expressed by moral statements.

Roskies also argues that VM patients make moral judgments on the grounds that they suffer no damage to neurological structures implicated in language and declarative knowledge (56, 60). But conservation of these structures shows that the ability to make moral judgments is preserved only if we assume that moral judgments are ordinary, non-motivating beliefs. Roskies might well have convincing reasons in support of an externalist theory of moral judgment on which, say, declarative knowledge is sufficient. But then it would be those arguments, not empirical evidence on VM patients, upon which the issue would turn. Because Roskies would have already settled the debate in favor of externalism, she could not use her theory to show on independent grounds that VM patients are counterexamples to internalism.

Let's turn now from the prosecution to the defense – from arguments criticizing internalism to arguments defending it. Kennett and Fine (2008a) offer several arguments that psychopaths do not make moral judgments and therefore are not counterexamples to internalism. Like Roskies, some of their arguments appeal to equivocal evidence. Psychopaths, the authors tell us, engage in abnormal moral reasoning (175–176), exhibit unusual speech patterns (176–177) and evince disunity of thought (177–178). But none of these deficits are essential to moral judgment.

To see this, let's reexamine the methodology defended above. Empirical criteria for the presence or absence of moral judgment must be derived from the general causal role of moral judgment. Kennett and Fine do cite evidence that indicates deviations in psychological processes associated with moral judgment. However, what is needed are deviations in psychological processes that are supported by the attitude or content that is constitutive of moral judgment. Otherwise, such deviations would seem to evince not the absence of moral

judgment but, rather, the absence of, or disorder in, associated psychological states or processes, e.g. an insensitivity to cognitive dissonance. Normal reasoning and speech patterns do not seem to be candidates for the general causal role grounded in moral judgment itself, not anyway without further argument that Kennett and Fine fail to provide. At any rate, then, the evidence that the authors cite is weak.

Kennett and Fine's other arguments depend on premises that are not independent of internalism. First of all, the authors argue that psychopaths have 'an inadequate understanding of moral concepts' because they 'make exceptions for themselves' (176). This *might* suggest a failure to conceptualize morality as general. But it might instead show that psychopaths are moved more by their own interests than by moral considerations, and this speaks only to their selfish motivations rather than their failure to grasp moral concepts. Motivated to commit immoral acts, they may rationalize excusing themselves from what they acknowledge are general prohibitions – at least, for all this evidence says. Of course, internalists are drawn to the idea that moral judgment is inconsistent with grossly immoral behavior that disregards those judgments. But the immoral motivations of psychopaths cannot be used to show that they do not make moral judgments, not without presupposing internalism.

Kennett and Fine also suggest that psychopaths do not understand moral language because they are insensitive 'to the emotional meaning of affective words' (177). However, to assume that moral words are affective or have emotional meanings, even only in part, is to assume internalism. Perhaps moral words do have emotional meanings, but then Kennett and Fine's defense of internalism hangs on that claim, and not on the moral deficits of psychopaths. Once again, an argument in the debate fails because it depends on a criterion for moral judgment that isn't neutral between internalism and externalism.

One of Kennett and Fine's arguments is more promising and given more thorough treatment by Prinz (2007, 42–47; cf. Kennett and Fine 2008a, 174–175). Prinz (2007, 2015) builds an empirical case for internalism by considering a broad and diverse body of evidence in cognitive science. We'll attend to his overall case for internalism later on in the essay, but one critical part of his argument rests on evidence from psychopathy. Like Kennett and Fine, Prinz defends internalism by arguing that psychopaths do not make genuine moral judgments. He relies mainly on studies by James Blair and colleagues that assay psychopaths' grasp of the moral/conventional distinction (Blair 1995, 1997; Blair et al. 1995). Blair's participants were adult psychopaths and children with nascent psychopathic traits, both of whom were given the moral/conventional task. As Prinz reports Blair's findings, neither group responded differently to moral violations than they did to conventional violations. Thus, Prinz argues, because they are blind to the distinction between morality and convention, psychopaths do not possess moral concepts. He concludes that psychopaths are incapable of making moral judgments and thus do not threaten internalism.

Prinz's defense of internalism does not rely on internalist assumptions. However, Prinz seems to misinterpret Blair's research. In an exchange with Levy (2007), Vargas and Nichols (2007) argue that Blair's empirical findings are sometimes exaggerated. Blair does not find that psychopaths have absolutely no grasp of the moral/conventional distinction. 'What Blair does find ... is a diminished sensitivity or capacity with respect to making the distinction' (Vargas and Nichols 2007, 157). In a sobering further thought, the authors continue:

It is worth bearing in mind that experiments on psychopathologies usually produce data that is less ordered than we might hope for. For example, it is not as though all autistic children fail the false belief task. Nor do psychopaths miss every case of the moral/conventional task. Rather, psychopaths tend to show relatively diminished response [compared with controls]. (Vargas and Nichols 2007, 157–158)

There are not simply two possible outcomes of the moral/conventional task: pass or fail. Rather, one can fail to varying degrees, and the psychopaths in Blair's studies fail to a significant degree without failing completely. That is, psychopaths exhibit a diminished sensitivity to the moral/conventional distinction. But Prinz misinterprets Blair's findings because he reasons as if psychopaths completely fail to draw the moral/conventional distinction, characterizing psychopaths as suffering from 'moral blindness' (Prinz 2007, 46) and concluding that they 'seem to comprehend morality, but they really don't' (43). Thus, even though the evidence he cites is apt, the argument Prinz formulates on its basis is unsound.

5. Psychopaths and moral concepts

It is time now to deliver a verdict on psychopaths' capacity for moral judgment. As Prinz foresees, Blair's research on their ability to draw the moral/conventional distinction suggests a deficit in psychopaths' grasp of moral concepts. However, Blair's evidence must be handled more carefully and, furthermore, newer and apparently conflicting evidence must be considered.

In one study Blair's participants were psychopath and non-psychopath criminals incarcerated in prison, both of whom were given the moral/conventional task (Blair 1995). Whereas the non-psychopaths reliably distinguished between moral and conventional violations, the psychopaths did not do so reliably. They did significantly better than chance but significantly worse than adults, worse even than young children. In another study, Blair finds that children with nascent psychopathic tendencies tend to treat both types of violations as fitting the conventional profile (Blair 1997).

Blair's findings are striking because they suggest an explanation of psychopaths' immoral behavior outside the lab. Psychopaths seem to treat morality as a set of inconvenient rules to be negotiated rather than duties to be respected, i.e. roughly the same way that anyone might treat conventional rules when they

conflict with important matters of self-interest. Psychopaths make promises for the sake of immediate ends only to disregard their promises when it suits them. They are motivated not to harm others only by external awards, rather than internal rewards. As Prinz puts it, psychopaths think of morality as 'a group of more or less arbitrary conventions that place demands on us only because they have been adopted by a social group' (2007, 44). The reason for this behavior, Blair's research suggests, is that psychopaths tend to view morality and convention similarly. Although Prinz claims that they have zero grasp of moral concepts, the evidence suggests, rather, that their grasp of moral concepts is impaired but not entirely absent. Psychopaths do not draw the moral/conventional distinction as reliably as non-psychopaths, but they are not entirely 'blind' to the distinction.

In §3 I noted that some atypical moral judgments encode only some of the three features that define the concept of morality. Psychopaths' moral judgments are also atypical, but not in the same way. Rather than encoding only some of the features, psychopaths have a limited grasp of all three features. The phenomenon runs parallel to intermediate stages of concept acquisition. Children often have a limited grasp of concepts such that they can identify only some members of their extensions and have difficulty applying the concepts to new cases. For example, at a certain age children have only a limited grasp of the concept of a promise. They understand that if you make a promise you should follow through on it, but they don't yet fully understand the difference between promising and merely stating one's intention (Mant and Perner 1988). Psychopaths' grasp of moral concepts is similarly immature. They do not have a full grasp of moral concepts, but nor do they completely fail to grasp them either. Rather, psychopaths have a more limited or tenuous grasp than non-psychopaths.

Psychopaths have some facility with the moral/conventional distinction. But they do not draw the distinction as reliably as others. Because psychopaths have an impaired grasp of moral concepts, their capacity to make moral judgments is likewise impaired. Psychopaths do not make 'full-fledged' moral judgments. Possessing an immature grasp of moral concepts, they instead make only 'proto' moral judgments. Relying on a theory of moral judgment as a natural kind, using that theory to identify the most relevant research, and after carefully interpreting that research, this is what the evidence suggests.

Sinnott-Armstrong (2014), however, argues that newer findings challenge conclusions in this vein. Aharoni, Sinnott-Armstrong, and Kiehl (2012) conducted a study on psychopaths that employed a modified version of the moral/conventional task, designed to test an alternative interpretation of Blair's data. The main distinguishing feature of their version of the task was that participants were given a list of violations and told that half of them are moral and the other half conventional. This eliminated any motivation psychopaths might have to classify all violations as moral in order to manage researchers' opinions of their character. Interestingly, Aharoni et al. found that psychopaths did not perform

significantly worse on this task than non-psychopath controls. That is, they reliably drew the moral/conventional distinction.

Aharoni et al., however, employed a relatively easy version of the moral/conventional task. Because psychopaths were told that half of the violations are moral and the other half conventional, they could use their answers to some questions to figure out answers to other questions. Thus, it seems, a partial grasp of moral concepts served them well enough on this easier task. The standard moral/conventional task is more difficult, and in Blair's studies psychopaths perform significantly worse than adult controls, worse too than young children. The hypothesis that psychopaths have a partial grasp of moral concepts and an ability to make only proto moral judgments explains why they performed well on Aharoni et al.'s relatively easy task, but systematically with only limited competence in Blair's relatively difficult task.

This interpretation of Blair and Aharoni et al.'s findings is acceptable only if it is better than the alternatives. Another interpretation begins with the following thought: while psychopaths do not care about morality themselves, they presumably appreciate, even if only dimly, that other people in society do. Perhaps, then, psychopaths' performance on different versions of the moral/conventional task reflects a moderately successful ability to identify what other people regard as moral violations, rather than a capacity to make proto moral judgments themselves.

This competing interpretation suggests, in effect, that psychopaths possess another type of mental state that plays roughly the same causal role as proto moral judgments, i.e. beliefs about others' moral attitudes. However, further data in Aharoni et al.'s own study cast doubt on this interpretation. The authors found that two sub-facets of psychopathy – affective and anti-social aspects of the disorder – reliably predict performance on their moral/conventional task. Participants who tested higher on these two sub-facets were less likely to categorize moral and conventional violations accurately. The view that psychopaths are employing a general ability to reason about others' opinions does not account for this finding. Rather, it seems that conditions internal to psychopathy account for variability in performance.

Studies of psychopaths' ability to draw the moral/conventional distinction, along with an empirically grounded theory of moral judgment as a natural kind, entails neither of the standard internalist or externalist verdicts on psychopaths. Had the empirical evidence shown that psychopaths do make moral judgments, it would follow that internalism is false. Or instead, had the evidence shown that psychopaths do not make moral judgments, nothing would follow; that is, while externalists have charged that psychopaths are counterexamples to internalism, that charge would be rebutted. But where does the debate stand given that psychopaths make proto moral judgments? Answering that question is less straightforward and the business of the next and final section.

6. Against internalism

Philosophical discussion of psychopathy and internalism tends to be narrow in one respect. Critics and defenders of internalism focus only on whether psychopaths are counterexamples to internalism. If moral judgment is a natural kind, however, we must demand more from a philosophical theory than mere consistency with the evidence. Specifically, theories of moral judgment should also be evaluated on *abductive* grounds, that is, with respect to how well they explain the evidence. Suppose that two theories of moral judgment are both consistent with the evidence from psychopathy, but one theory is part of a better explanation of the evidence. In that case it gains empirical support. I will argue in this section that although internalism is not refuted by the evidence from psychopathy, it seems to lack positive empirical support. This constitutes an empirical challenge to internalism.

Prinz has paid a great deal of attention to these issues, and he argues that two pieces of empirical evidence lend support to internalism. First of all, Prinz (2007) argues that internalism is empirically justified because it provides the best explanation of psychopaths' deficits. However, as part of a much broader empirical case for internalism, Prinz (2007, 2015) also argues that there is an empirically observed correlation between moral judgment and motivation, and that internalism provides the best explanation of this correlation (cf. Smith 1994). I will challenge both of these arguments, and my discussion will proceed in three stages. First, contrary to Prinz, internalism fails to provide an illuminating explanation of psychopaths' deficits. Second, the best explanation of psychopaths' deficits does not require internalism. Third, this explanation of psychopaths' deficits depends on a general theory that also accounts for the regular correlation between moral judgment and motivation, and thus provides a *prima facie* challenge to Prinz's broader argument. It would seem, then, that we have no good empirical reasons to accept internalism as a synthetic, necessary thesis.

6.1. Internalism does not explain psychopaths' deficits

Let's begin with Prinz's argument from psychopathy to internalism. He argues that psychopaths do not undermine internalism, as we have seen, but he also claims that 'they furnish internalists with a useful piece of supporting evidence' (2007: 44). Internalism, according to Prinz, is part of the best explanation of psychopaths' moral deficits:

The moral blindness of psychopaths issues from an emotional blindness. If this is right, psychopathy provides positive evidence for internalism ... If moral judgments are intrinsically motivating, it may be due to the fact that standard moral concepts are essentially emotion-laden. That is precisely what research on psychopathy seems to confirm. (Prinz 2007, 46)

As we have seen, however, Prinz's argument rests on a false premise: psychopaths make proto moral judgments, rather than, as he thinks, failing to make moral judgments at all. The question now is whether internalism helps to explain the finding that psychopaths make proto moral judgments. It appears not to.

The standard formulation of internalism states that necessarily, if someone makes a moral judgment he or she possesses corresponding motivation. This standard formulation implicitly treats moral judgment as an all-or-none category, rather than a graded category that instances can fall within to varying degrees. To fairly assess internalism, we must consider formulations that explicitly treat moral judgment as a *graded* category. Two formulations are salient, but neither seems to be adequate.

First of all, perhaps internalism should be cast explicitly as the view that full-fledged moral judgments necessitate motivation – call this view 'categorical internalism.' Moral judgments that are not full-fledged, including proto moral judgments found in psychopaths, do not necessitate motivation. Or, rather, categorical internalism is silent on whether they do, and thus the combination of proto moral judgments and flat moral motivation in psychopaths is not a counterexample to the view.

Categorical internalism, however, struggles to explain psychopaths' deficits. If motivation is central to full-fledged moral judgments, then how can proto moral judgments exist without motivation? That is, if moral judgment is a graded category, then how can it be present to some degree when its motivational basis is absent? No explanation is apparent. That is, it is not clear how categorical internalism could explain the phenomena: that once a proto moral judgment turns into a full-fledged moral judgment, motivation suddenly and as a matter of necessity joins it. The view is saddled with this consequence because moral judgment comes in degrees whereas, according to categorical internalism, the motivation that is constitutive of moral judgment is all-or-none.

Some philosophers might favor an alternative formulation of internalism – call it 'degree internalism' – according to which moral judgment necessitates a corresponding degree of motivation. That is, if one makes a moral judgment to some degree, then one is motivated to the same degree. We already have a sense of how moral judgment can come in degrees: one's grasp of moral concepts comes in degrees, as measured by more or less difficult tasks in which participants are asked to employ their moral concepts. The relevant sense in which motivation comes in degrees has to do with strength or intensity. Thus, degree internalism says that necessarily, to the extent one makes a moral judgment one will have motivation of corresponding strength.

Though it is difficult to find clear empirical evidence that bears on this issue, psychopaths seem to lack moral motivation altogether, or nearly so, rather than merely possessing it to a middling degree. As I noted near the beginning of the essay, psychopaths do not just act immorally; immoral behavior is consistent with the existence of moral motivation that is overridden by competing motives.

Rather, psychopaths violate moral norms so casually, and so frequently, that it seems they do not care about morality at all. Degree internalism therefore seems not to capture psychopaths.⁷

Suppose, however, that psychopaths have some degree of moral motivation. In that case Prinz's argument merits reconsideration. According to Prinz, recall, psychopaths suffer from 'moral blindness' and simply do not make moral judgments. I have argued, more precisely, that psychopaths make only proto moral judgments. However, it may seem as if Prinz's argument can be reformulated such that it provides an explanation. Perhaps psychopaths have some, albeit diminished emotional capacities, and thus are able to make proto moral judgments. This is consistent with degree internalism. Earlier, I noted Aharoni et al.'s finding that diminished reliability at drawing the moral/conventional distinction is associated with affective and anti-social aspects of psychopathy. Prinz might argue that these aspects also undermine the capacity for theory of mind, and thus explain why psychopaths do not reliably make inferences about how others draw the moral/conventional distinction. Thus, it seem as if degree internalism, in Prinz's hands, explains psychopaths' deficits.

This reformulation of Prinz's argument is plausible on its face, but it fails to accommodate other relevant evidence. Recall VM patients, who exhibit affective and motivational deficits that are very similar to those found in psychopaths, but that do not arise during development and instead are the result of brain damage that occurs in adulthood. It turns out that VM patients are able to draw the moral/conventional distinction reliably on the standard experimental task (Saver and Damasio 1991), that is, the relatively difficult task on which psychopaths perform poorly in Blair's studies. So, this evidence speaks against the possibility that occurrent emotional deficits explains psychopaths' deficits in moral judgment. For that view predicts that VM patients would also make only proto moral judgments, and would perform as poorly as psychopaths on the moral/conventional task.

6.2. *The best explanation of psychopaths' deficits*

The best explanation for psychopaths' impaired capacity for moral judgment does not entail any sort of necessary link between moral judgment and motivation. The explanation relies on a theory of moral judgment that accords emotion an important role, not in online moral judgment, but in development of the capacity for moral judgment.

Psychopaths' antisocial behavior seems to be rooted mainly in their severe affective deficits. One of their most noteworthy deficits is the absence of sympathy toward others. Normally, appreciation that another creature is in pain or experiencing another type of harm leads to sympathetic distress. Psychopaths are disposed toward criminal behavior, and often excel at it, in part because they do not undergo this reaction. Theoretical work by Nichols suggests that

psychopaths' affective deficits explain not just their anti-social behavior but also their limited grasp of moral concepts.

Nichols (2004a, 26–29) argues that proper functioning of an affective system is critical for acquiring moral concepts over the course of early development (cf. Blair 1995, 1997). To put this in terms of the theory of moral concepts developed above, it is through experiencing sympathetic distress that we arrive at a conception of morality as serious, general, and authority-independent. Thus, as Nichols argues, affective deficits in psychopaths explain why they do not fully acquire moral concepts. It explains too why in Aharoni et al.'s study 'reduced moral categorization accuracy was significantly predicted by affective and anti-social traits' (2012, 493). This piece of data fits with internalism, but it also fits with Nichols' account of moral concept acquisition.

Affective motivation is central to moral judgment, but not, it seems, in the way that some internalists think – central to historical development of the capacity, not its online exercise. Research on VM patients' grasp of the moral/conventional distinction supports Nichols' developmental account over internalism. Because VM patients had intact affective systems as children, Nichols' account predicts that they would acquire a normal facility with moral concepts, unlike psychopaths. Indeed, as noted above, VM patients perform like non-psychopaths on the moral/conventional task. Thus, evidence from VM patients supports Nichols' view.

Nichols' developmental account states that affect is critical for full acquisition of moral concepts. Internalism and externalism, notice, are both consistent with this account. Even if affect supports the development of moral concepts, affect or motivation might also play a constitutive role in moral judgment. However, we are now seeking more than mere consistency with the evidence. Nichols' developmental account provides an attractive explanation of psychopaths' limited grasp of moral concepts, one that depends on ideas that are independently supported and that have wide explanatory scope. Internalism, as we are understanding it, is a thesis about a natural kind. But then psychopaths give us no reason to accept internalism because it isn't needed to explain their deficits. A complete assessment of internalism must examine whether there is any other empirical evidence that speaks in its favor, and it is that third and final issue to which I now turn.

6.3. Against a broader empirical argument for internalism

Prinz (2015) argues for internalism on the grounds that there is a regular correlation between moral judgment and emotion-backed motivation. The best explanation for this correlation, he claims, is that motivation is necessary for moral judgment. Unfortunately for Prinz, the general theory that underlies Nichols' developmental account provides an alternative explanation for the data that is more economical, independently supported, and broader in scope. The reason

that moral judgment and motivation are regularly correlated is that morally significant events elicit not just moral judgments but also affective responses; the absence in development of these affective responses impairs psychopaths' grasp of moral concepts.

Moral judgments are typically accompanied by emotion, and this explains why there is a regular correlation between, e.g. judging that something is morally wrong and being motivated to avoid doing it. When an action is perceived as morally significant, we typically (but not always) form both a moral judgment about the action and a moral emotion directed toward the action that is motivating, including sympathetic distress in response to harm (see Haidt 2001; Nichols 2004a; Greene 2008; Railton 2014; Kumar *Forthcoming-a*).⁸ This is why when someone forms a moral judgment he or she is likely to possess the corresponding moral motivation – though not always, as in VM patients. The relationship between moral judgment and emotion is intimate, then, but it is contingent rather than necessary (see Kumar *Forthcoming-a* for further discussion). A lengthier treatment of this issue is necessary to evaluate the broader empirical case for internalism (see Kumar *Forthcoming-b*), but we seem to find here a *prima facie* challenge to Prinz's broader case.

There is a correlation between moral judgment and (emotion-backed) motivation. However, externalists have at their disposal an independently justified explanation for this evidence. The explicit target of my argument is Prinz's sentimentalist version of internalism, on which moral judgments are constituted in part by moral emotions that are motivating. However, other versions of internalism also seem to be subject to my criticism, though only insofar as they also hope to gain support from the empirically observed correlation between moral judgment and motivation. One view that may avoid the criticism is so-called 'soft internalism,' the view that moral judgment is inherently but only defeasibly motivational – and therefore not bound of necessity to motivation, strictly speaking. For example, some authors claim that moral emotion is partly constitutive of 'typical' moral judgments, with the result that atypical moral judgments can obtain in the absence of motivation (Campbell 2007; Kumar *Forthcoming-a*).

At this point, it is worth considering Smith's brand of (1994) soft internalism, according to which moral judgment necessitates motivation provided that the agent who makes the judgment is rational. Smith takes this view to be justified a priori, but does empirical evidence count against it? First of all, Smith's view is controversial. In general, moral judgment seems to motivate just as reliably in irrational people as it does in rational people. However, in the present context Smith's internalist view may seem to escape my criticisms. If psychopaths are irrational, then they are not a counterexample to the view that moral judgment necessitates motivation only in rational people.

The problem with this response is that it does not seem as if psychopaths exhibit general deficits in rationality that are distinct from their moral deficits.

Psychopaths are well known for the ability to control and manipulate others. Furthermore, their scores on IQ tests are no lower than those in control populations (Blair et al. 1995). Thus, as Nichols (2004a) argues, there is no evidence to suggest that psychopaths exhibit general deficits of irrationality. The burden is on defenders of Smith's brand of internalism to furnish relevant evidence, but thus far no one has made a convincing case (cf. Maibom 2005). As with any topic in naturalistic philosophy, conclusions must be tentative and subject to revision in light of an evolving empirical literature. Still, it seems, there is no evidence for the sort of rational deficits among psychopaths that is needed to defend Smith's view.

7. Conclusion

Some philosophers argue that psychopaths are real-life counterexamples to motivational internalism. At first glance, this proposed empirical refutation of internalism cannot avoid begging the question. However, if moral judgment is a natural kind, empirical evidence *can* be used to evaluate internalism, in particular, by generating criteria for the presence of moral judgment that do not presuppose either internalism or externalism. With the aid of a theory of the moral concepts that are partially constitutive of moral judgment, research suggests that psychopaths make only proto moral judgments, and therefore are not counterexamples to internalism.

Nonetheless, internalism is empirically unsupported because it fails, relative to competitors, to provide the best explanation of relevant empirical evidence. Internalism does not explain why the necessary connection with motivation that it postulates is broken when moral judgment is not fully but still partially present in psychopaths. Furthermore, the evidence suggests that psychopaths' impaired capacity for moral judgment stems from their affective deficits during moral development, and not because affect or motivation is absent from occurrent moral judgments. This explanation of psychopaths' moral deficits does not entail internalism and rests on a theory that accounts for the regular correlation between moral judgment and motivation that, according to Prinz, also seemed to offer empirical support to internalism. Internalism, then, would seem not to provide the best explanation of this correlational evidence either.

We should conclude that internalism lacks empirical support. If moral judgment is a natural kind, then internalism is bereft of the evidence that is crucial to its vindication. And so, although psychopaths are not counterexamples, as externalists initially charged, they nonetheless pose an empirical challenge to internalism. We should seek an externalist theory of moral judgment that comports better with empirical evidence.

Notes

1. An experimental approach might only provide evidence *against* internalism, not evidence in support of it. The reason is that even if participants uniformly deny that psychopaths' make moral judgments, and even if well designed studies show that they deny it for the reason that psychopaths' lack moral motivation, this might be due only to a common belief that there is a tight, synthetic link, rather than due to a stronger, conceptual link. Furthermore, it is doubtful that experimental investigation of folk intuitions could be used to undermine a sophisticated version of internalism, like Smith's (1994), that postulates a necessary link between moral judgment and motivation only in a rational or normal agent. We should not place much confidence in the folk's ability to determine whether an agent is rational or normal. These points are owed to an anonymous reviewer.
2. Besides being serious, general, and authority-independent, morality is also conceived as objective (Kumar 2015). That morality is conceived as objective explains why participants think that in moral disagreement at least one of the parties must be wrong (see Nichols 2004b; Goodwin and Darley 2008). I will ignore this more complex account of moral concepts in this essay, since the studies of psychopaths discussed below depend only on the simpler account.
3. Thanks to an anonymous reviewer for pressing this issue.
4. In what follows I will rely on a theory of the conceptual content that is constitutive of moral judgments. Because possession of moral concepts is only a necessary condition on moral judgment, assessing whether psychopaths have moral concepts can serve only as a negative test.
5. Thanks to an anonymous reviewer for raising this issue.
6. To be fair, Nichols' primary target is moral rationalism, not motivational internalism. He argues that affective processes play a causal role in the process of moral judgment. Psychopaths have an impaired capacity for moral judgment and the explanation, according to Nichols, is that they have severe affective deficits. Thus, moral judgment is not based on reasoning alone (Nichols 2004b, 65–96; cf. Maibom 2005). More on this in the final section.
7. Degree internalism is implausible too on independent grounds. Although moral judgments and motivation do tend to correlate, it does not seem as if, in general, the degree to which someone makes a moral judgment correlates with degree of motivation. You and I both form full-fledged judgments that secret government surveillance is morally wrong, even though as an activist you are far more motivated than I to do something about it.
8. There is much controversy about whether emotions are among the causes of moral judgments, but all sides agree that there is a correlation between moral judgments and emotion. Even so-called 'moral rationalists' like Mikhail (2011) who hold that internally represented rules produce moral judgments accept that emotions are typically produced downstream of moral judgment.

Acknowledgements

For helpful feedback I am grateful to Michael Bukoski, Richmond Campbell, Michael Doan, Terry Horgan, Stephen Latta, Theresa Lopez, Alice MacLachlan, Shaun Nichols, Peter Railton, Nolan Ritcey, Mark Timmons, Jon Tresan, two anonymous referees for this journal, and audiences at the University of Arizona, the Southern Society for Philosophy and Psychology, and the Canadian Philosophical Association.

Notes on contributor

Victor Kumar is a postdoctoral research fellow in the philosophy department at the University of Toronto. His research is primarily in ethics, moral psychology, and cognitive science. His recent publications include "Moral Judgment as a Natural Kind" (Philosophical Studies 2015) and "Moral Reasoning on the Ground" with Richmond Campbell (Ethics 2012).

References

- Aharoni, E., W. Sinnott-Armstrong, and K. Kiehl. 2012. "Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction." *Journal of Abnormal Psychology* 121 (2): 484–497.
- Björnsson, G., J. Eriksson, C. Strandberg, R. F. Olinder, and F. Björklund. 2014. "Motivational Internalism and Folk Intuitions." *Philosophical Psychology* 28 (5): 715–734.
- Björnsson, G., and T. McPherson. 2014. "Moral Attitudes for Non-cognitivists: Solving the Specification Problem." *Mind* 123: 1–38.
- Blair, J. 1995. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57: 1–29.
- Blair, J. 1997. "Moral Reasoning and the Child with Psychopathic Tendencies." *Personality and Individual Differences* 22: 731–739.
- Blair, J., L. Jones, F. Clark, and M. Smith. 1995. "Is the Psychopath 'Morally Insane'?" *Personality and Individual Differences* 19 (5): 741–752.
- Boyd, R. 1988. "How to Be a Moral Realist," In *Essays on Moral Realism*, edited by G. Sayre-McCord, 181–228. Ithaca, NY: 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 *Species: New Interdisciplinary Essays*, edited by R. A. Wilson, 141–185. Cambridge, MA: MIT Press.
- Brandts, J., and A. Schram. 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*. Cambridge: Cambridge University Press.
- Campbell, R. 2007. "What is Moral Judgment?" *Journal of Philosophy* 104 (7): 321–349.
- Carey, S. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.
- Cholbi, M. 2006. "Belief Attribution and the Falsification of Motive Internalism." *Philosophical Psychology* 19 (5): 607–616.
- Fehr, E., and U. Fischbacher. 2004a. "Social Norms and Human Cooperation." *Trends in Cognitive Sciences* 8: 185–190.
- Fehr, E., and U. Fischbacher. 2004b. "Third-party Punishment and Social Norms." *Evolution and Human Behavior* 25: 63–87.
- Fehr, E., and S. Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980–994.
- Fischbacher, U., S. Gächter, and E. Fehr. 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., and J. Darley. 2008. "The Psychology of Meta-ethics: Exploring Objectivism." *Cognition* 106: 1339–1366.
- Greene, J. 2008. "The Secret Joke of Kant's Soul," In *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders and Development*, edited by W. Sinnott-Armstrong, vol. 3, 35–80. Cambridge, MA: MIT Press.
- Griffiths, P. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago, IL: University of Chicago Press.
- Haidt, J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108: 814–834.
- Haidt, J. 2003. "Moral Emotions," In *Handbook of Affective Sciences*, edited by R. Davidson, K. Scherer, and H. Goldsmith, 852–870. Oxford: Oxford University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hare, R. D. 1993. *Without Conscience: The Disturbing World of Psychopaths among Us*. New York: Guildford.
- Hollos, M., P. Leis, and E. Turiel. 1986. "Social Reasoning in IJO Children and Adolescents in Nigerian Communities." *Journal of Cross-Cultural Psychology* 17: 352–374.
- Holton, R. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Kauppinen, A. 2008. "Moral Internalism and the Brain." *Social Theory and Practice* 34: 1–24.
- Kauppinen, A. 2013. "Ethics and Empirical Psychology," In *Empirically Informed Ethics*, edited by M. Christen, 279–305. Springer.
- Keil, F. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
- Kelly, D., and S. Stich. 2007. "Two Theories of the Cognitive Architecture Underlying Morality," In *The Innate Mind*, edited by P. Carruthers, S. Laurence, and S. Stich, vol. 3, 348–366. New York: Oxford University Press.
- Kelly, D., S. Stich, K. Haley, S. Eng, and D. Fessler. 2007. "Harm, Affect, and the Moral/Conventional Distinction." *Mind and Language* 22 (2): 117–131.
- Kennett, J., and C. Fine. 2008b. "Could There Be an Empirical Test for Internalism?" In *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, vol. 3, 217–226. Cambridge, MA: MIT Press.
- Kennett, J., and C. Fine. 2008a. "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths,'" In *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, vol. 3, 173–190. Cambridge, MA: MIT Press.
- Keser, C., and F. van Winden. 2000. "Conditional Cooperation and Voluntary Contributions to Public Goods." *Scandinavian Journal of Economics* 102: 23–39.
- Kornblith, H. 2002. *Knowledge and its Place in Nature*. Oxford: Oxford University Press.
- Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Kumar, V. 2014. "'Knowledge' as a Natural Kind Term." *Synthese* 191: 439–457.
- Kumar, V. 2015. "Moral Judgment as a Natural Kind." *Philosophical Studies* 172 (2887): 2910.
- Kumar, V. Forthcoming-a. "The Empirical Identity of Moral Judgment," *The Philosophical Quarterly*.
- Kumar, V. Forthcoming-b. "The Ethical Significance of Cognitive Science," In *Current Controversies in Cognitive Science*, edited by S. Leslie and S. Cullen. Routledge.
- Laurence, S., and E. Margolis. 1999. "Concepts and Cognitive Science," In *Concepts: Core Readings*, edited by E. Margolis and S. Laurence, 3–81. Cambridge, MA: MIT Press.
- 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.
- Mant, C., and J. Perner. 1988. "The Child's Understanding of Commitment." *Developmental Psychology* 24 (3): 343–351.
- Mikhail, J. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge University Press.
- 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.
- Nucci, L. 1986. "Children's Conceptions of Morality, Societal Convention and Religious Prescription," In *Moral Dilemmas: Philosophical and Psychological Reconsiderations of the Development of Moral Reasoning*, edited by C. Harding, 138–174. Chicago, IL: Precedent Press.
- Nucci, L. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- Nucci, L., and E. Turiel. 1978. "Social Interactions and the Development of Social Concepts in Preschool Children." *Child Development* 49: 400–407.
- Nucci, L., E. Turiel, and G. Encarnacion-Gawrych. 1983. "Children'S Social Interactions and Social Concepts: Analyses of Morality and Convention in the Virgin Islands." *Journal of Cross-Cultural Psychology* 14: 469–487.
- Pillutla, M., and J. Murnighan. 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.
- Prinz, J. 2015. "An Empirical Case for Motivational Internalism," In *Motivational Internalism*, edited by G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund, 61–84. Oxford: Oxford University Press.
- Putnam, H. 1975. "The Meaning of 'Meaning,'" In *Language, Mind and Knowledge*, edited by K. Gunderson, 131–193. Minneapolis, MN: University of Minnesota Press.
- Railton, P. 1986. "Moral Realism." *The Philosophical Review* 95: 163–207.
- Railton, P. 2014. "The Affective Dog and Its Rational Tale." *Ethics* 124: 813–859.
- Railton, P. Forthcoming. "The Value of Truth and the Value of Belief," In *The Value of Truth*.
- Rosch, E. 1978. "Principles of Categorization," In *Cognition and Categorization*, edited by E. Rosch and B. Lloyd, 27–48. Hillsdale, NJ: Laurence Erlbaum Associates.
- Roskies, A. 2003. "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy.'" *Philosophical Psychology* 16 (1): 51–66.
- Rozin, P. 1997. "Moralization," In *Morality and Health*, edited by A. Brandt and P. Rozin, 379–401. New York: Routledge.
- Rozin, P. 1999. "The Process of Moralization." *Psychological Science* 10: 218–221.
- Rozin, P., and L. Singh. 1999. "The Moralization of Cigarette Smoking in the United States." *Journal of Consumer Psychology* 8: 321–337.
- Rozin, P., M. Markwith, and C. Stoess. 1997. "Moralization and Becoming a Vegetarian: The transformation of Preferences into Values and the Recruitment of Disgust." *Psychological Science* 8: 67–73.
- Saver, J. L., and A. R. Damasio. 1991. "Preserved Access and Processing of Social Knowledge in a Patient with Acquired Sociopathy Due to Ventromedial Frontal Damage." *Neuropsychologia* 29: 1241–1249.
- Schroeder, T. 2004. *Three Faces of Desire*. Oxford: Oxford University Press.
- Searle, J. 1958. "Proper Names." *Mind* 67: 166–173.

- Shoemaker, D. 2011. "Psychopathy, Responsibility, and the Moral/Conventional Distinction." *The Southern Journal of Philosophy* 49: 99–124.
- Sinnott-Armstrong, W. 2014. "Do Psychopaths Refute Internalism?" In *Being Amoral*, edited by T. Schramme, 187–207. Cambridge, MA: MIT Press.
- 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 *The Development of Social Cognition: The Child as Psychologist*, edited by M. Bennett, 111–141. New York: Guilford Press.
- Smetana, J., and J. Braeges. 1990. "The Development of Toddlers' Moral and Conventional Judgments." *Merrill-Palmer Quarterly* 36: 329–346.
- Smetana, J., M. Kelly, and C. Twentyman. 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., J. Smetana, and S. Kim. 1987. "Korean Children's Conceptions of Moral and Conventional Transgressions." *Developmental Psychology* 23: 577–582.
- Stocker, M. 1979. "Desiring the Bad: An Essay in Moral Psychology." *The Journal of Philosophy* 76 (12): 738–753.
- Strandberg, C., and F. Björklund. 2013. "Is Moral Internalism Supported by Folk Intuitions?" *Philosophical Psychology* 26 (3): 319–335.
- Sturgeon, N. 1988. "Moral Explanations." In *Essays on Moral Realism*, edited by G. Sayre-McCord, 229–255. Ithaca, NY: Cornell University Press.
- Timmons, M. 1999. *Morality without Foundations: A Defense of Ethical Contextualism*. New York: Oxford University Press.
- Tisak, M. 1995. "Domains of Social Reasoning and beyond," In *Annals of Child Development*, edited by R. Vasta, vol. 11, 95–130. London: Jessica Kingsley.
- Tresan, J. 2006. "De Dicto Internalist Cognitivism." *Nous* 40: 143–165.
- Turiel, E. 1983. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.
- Turillo, C., R. Folger, J. Lavelle, E. Umphress, and O. Gee. 2002. "Is Virtue Its Own Reward? Self Sacrificial Decisions for the Sake of Fairness." *Organizational Behavior and Human Decision Processes* 89: 839–865.
- Vargas, M., and S. Nichols. 2007. "Psychopaths and Moral Knowledge." *Philosophy, Psychiatry and Psychology* 14 (2): 157–162.
- Weiskopf, D. 2009. "The Plurality of Concepts." *Synthese* 169: 145–173.
- Yau, J., and J. Smetana. 2003. "Conceptions of Moral, Social-conventional, and Personal Events among Chinese Preschoolers in Hong Kong." *Child Development* 74: 1647–1658.