

Daniel Eggers, Are emotions necessary and sufficient for moral judgement (and what would it tell us)? – *Final Draft*

The Version of Record of this manuscript has been published and is available in *Philosophical Explorations*, 2022, <http://www.tandfonline.com/> DOI: 10.1080/13869795.2022.2121848

Abstract: The 18th century debate between moral rationalists and moral sentimentalists has seen a striking renaissance in the past decades, not least because of research into the nature of moral judgement conducted by empirical scientists such as social and developmental psychologists and neuroscientists. A claim that is often made in the current discussion is that the evidence made available by such empirical investigations refutes rationalist conceptions of moral judgement and vindicates the views of Hume or other moral sentimentalists. For example, Jesse Prinz and Hanno Sauer have recently argued that the available data demonstrates that emotions are both necessary and sufficient for moral judgement and that the best or the only way to make sense of these findings is to conclude that moral judgements are constituted by emotions. The aim of this paper is to thoroughly examine this argument and the underlying empirical evidence and to show that there is currently no compelling evidence for the truth of either the necessity or the sufficiency thesis and that, even if both theses were true, they would fail to provide a sound basis for a plausible sentimentalist constitution claim.

Keywords: psychopaths; acquired sociopaths; frontotemporal dementia; moral dumbfounding; disgust; the moral/conventional distinction

1. Introduction: The sentimentalism/rationalism debate revisited

The 18th century debate between moral rationalists and moral sentimentalists is one of the great debates in the history of moral philosophy. Although there are different takes on what lies at the heart of this debate, it is widely believed that at least one important part of the controversy revolves around the psychological nature of moral judgement: While rationalists such as Clarke, Burnet, Balguy or Price assimilate moral thinking to visual perception and/or arithmetical judgement, thereby suggesting that moral judgements are just ordinary beliefs, writers such as Hutcheson, Hume and Smith suggest that moral judgements are constituted by what we would now refer to as non-cognitive states of mind: feelings, emotions or desires.

The inquiry into the nature of moral judgement that occupied the rationalists and sentimentalists of the 18th century has seen a striking renaissance in recent decades. To a considerable extent, this renaissance has taken place outside philosophy, namely in empirical disciplines such as social and developmental psychology and the emerging neurosciences. Yet, there has also been a resurgence of interest among metaethicists and moral philosophers, and even where approaches have been predominantly empirical, explicit references have been made to the classical philosophical debate. In turn, more and more philosophers who address the issue do so with an eye on the existing empirical evidence.

The general assumption of both modern moral scientists and empirically-minded moral philosophers is that empirical findings may help us to answer the underlying philosophical question and to actually decide the sentimentalism/rationalism debate. As a matter of fact, a number of writers has argued that the existing evidence is already sufficient for us to make this decision. According to them, the findings from social psychology and neuroscience refute the rationalist idea that moral judgements are the upshot of our rational or cognitive capacities and demonstrate that Hume and the other sentimentalists got things right: Moral judgements are, at base, emotional, which suggests that some version of moral sentimentalism must be true.

The view that the available data provide decisive support for moral sentimentalism may be most common among empirical scientists such as Jonathan Haidt (2001, 2013) who presents

his ‘social intuitionist model’ of moral judgement as a modern-day variant of Hume’s moral sentimentalism. However, it has also been defended by moral philosophers like Jesse Prinz (2014, 2015) and Hanno Sauer (2017). In Prinz’s and Sauer’s view, the empirical evidence suggests a sentimentalist constitution claim according to which moral judgements are constituted by emotions because it reveals that emotions are both *necessary* and *sufficient* for moral judgement.

While the philosophical significance of the existing empirical research has been the object of much discussion, Prinz’s and Sauer’s attempts to systematize it and turn it into a direct argument for moral sentimentalism have not thoroughly been addressed so far. The aim of this paper is to take a closer look at the argument and the way in which it appeals to the empirical evidence. After setting out the argument and adding some preliminary observations (section 2), I will first address the necessity thesis, by critically reviewing the relevant empirical findings and assessing their theoretical import (section 3). In like manner, I will subsequently review the sufficiency thesis, with a particular emphasis on the relatively weak interpretation of this thesis to which both Prinz and Sauer subscribe (section 4). I will then try to show that, even if the necessity thesis and the weak sufficiency thesis were true, their conjunction would not yield the conclusion that moral judgements are constituted by emotions or other non-cognitive states of mind (section 5). My ultimate conclusion, therefore, is that the available evidence is fully compatible with the rationalist view that moral judgements are constituted by beliefs (section 6).

2. The argument

The argument for moral sentimentalism I want to address works by way of two crucial assumptions. The first is that emotions are *necessary* for moral judgement; the second is that emotions are *sufficient* for moral judgement. A third premise, which is presupposed rather than

explicitly spelled out, is that, if emotions are both necessary and sufficient for moral judgement, moral judgements must be constituted by emotions.

The argument, then, takes the following form:

P1: Emotions are necessary for moral judgment (*necessity thesis*)

P2: Emotions are sufficient for moral judgement (*sufficiency thesis*)

P3: If emotions are necessary and sufficient for moral judgement, moral judgements are constituted by emotions.

C Moral judgements are constituted by emotions (*sentimentalist constitution claim*)

Jesse Prinz most clearly employs this argument in one of his more recent publications. Drawing on evidence provided by Haidt and others as well as on a number of his own empirical studies, Prinz claims that emotional or affective responses are necessary and sufficient for moral judgement and that, for this very reason, moral judgements themselves should be seen as consisting in some kind of emotional or affective response (Prinz 2014, 101f.; see also 2015, 71-73). This conclusion accords with Prinz's self-description as "a good old-fashioned sentimentalist" (2011, 216) who fully accepts Hume's "equation of moral judgments with feelings of praise and blame" (216).

The argument has also been employed by Hanno Sauer (2017, 169ff.; see also 2012) whose approach to the psychology of moral judgement is generally indebted to Prinz's work on the subject but yields a slightly different conception of moral judgement. In contrast to Prinz's 'good old-fashioned sentimentalism', Sauer defends a conception that tries to mediate between sentimentalism (or, as Sauer prefers to call it, "emotionism" [2017, 142])¹ and moral rationalism. According to his "sophisticated emotionism" (201), moral judgements are emotional in nature; however, they are not constituted by any old feeling or emotion but by

‘educated intuitions’, that is, by emotions “that can be reflectively endorsed by the judging person under ideal conditions of full information and rationality” (201).

Before I will address the question of whether the three premises of the argument are actually true, and whether the existing empirical data show them to be true, some preliminary observations seem in order. The first observation concerns the logical status of premises 1 and 2. Although both the claim that ‘emotions are necessary for moral judgement’ and the claim that ‘emotions are sufficient for moral judgement’ could be read as universal quantifications, they are obviously intended as existential quantifications, that is, as the claims that ‘there are certain emotions that are necessary for moral judgement’ and that ‘there are certain emotions that are sufficient for moral judgement’. Though this may seem like a rather trivial point to make, it will gain some importance when it comes to assessing the third premise of the argument.

Premise 3 also requires clarification in its own right. Since Sauer does not explicitly address the question of how exactly the constitution claim follows from the conjunction of necessity and sufficiency thesis, his treatment could be taken to suggest that the third premise expresses some kind of logical or conceptual truth. In contrast, Prinz suggests that the necessity and sufficiency of emotions for moral judgement supports a sentimentalist constitution claim only in the sense of an inference to the best explanation (see 2014, 104; and 2015, 71). These two possible interpretations of the third premise are something I will also revisit when discussing the premise in section 5.

Finally, there are two terminological points that need to be addressed. The first concerns the use of the term ‘emotion’. Much ink has been spilled over the question of how emotions differ from (other kinds of) feelings as well as from other kinds of non-cognitive mental states, and Prinz himself has importantly contributed to the relevant debates (2006). However, since these issues do not directly bear on how to assess the above argument for moral sentimentalism, I will leave them aside in what follows: If the available empirical data failed to prove that

emotions are necessary and sufficient for moral judgement, but revealed that *feelings* or *desires* are so, then some version of the argument for sentimentalism would still be successful. For this reason, I will interpret the term ‘emotion’ in a wider sense that covers all sorts of non-cognitive states of mind. Ecumenical approaches of this kind are not at all uncommon in the modern metaethical literature, where the term ‘desire’ is often taken to likewise refer to various affective and conative mental states. Given Hume’s broad use of the term ‘passions’, they may even be characteristic of classic moral sentimentalism.

The second terminological point concerns Prinz’s and Sauer’s appeal to the notions of necessity and sufficiency. One worry one might have with regard to the argument is that the talk of necessary and sufficient conditions cannot be applied to empirical contexts in the way Prinz and Sauer suggest because it belongs to the context of formal definitions and, more generally, to the realm of logical or conceptual relations. Empirically speaking, no single condition seems to be strictly necessary for a concrete event to come about since we can always conceive of a possible world in which the same event is produced in a different way. Moreover, no single condition or set of conditions seems to qualify as sufficient, either, given that causal explanations will typically focus on a particular part of a much more complex overall setting in which numerous background conditions contribute to the generation of the event as well. On the basis of such considerations, one might want to claim that Prinz and Sauer’s argument is fundamentally misguided. In fact, some philosophers would go even further and argue that, in virtue of considerations of the above kind, empirical evidence (psychological or otherwise) is generally incapable of contributing to the solution of metaethical or other philosophical questions since these are at base conceptual questions.

In my view, however, we should embrace a more lenient attitude. Talk of ‘necessary’ and ‘sufficient’ conditions may be confusing or deceptive in certain empirical contexts. However, if the terms are being used to describe psychological regularities to which there are no known exceptions – which is the way in which Prinz and Sauer use them –, they strike me

as ultimately innocuous. With regard to the more comprehensive idea that conceptual questions are generally immune to empirical evidence, a more moderate position seems to be in order as well. It is certainly true that empirical evidence itself will always be dependent upon prior conceptual commitments. However, even if the ultimate purpose is to answer a conceptual question or settle a conceptual dispute, it can be helpful to start from working definitions that omit controversial implications of the relevant concept and to argue for a more comprehensive notion on the basis of empirical data which pertains to the initial working concept. Again, I think that Prinz's and Sauer's approach can be interpreted along just those lines, which is why I believe that their argument can potentially contribute to the sentimentalism/rationalism debate even if this debate is ultimately to be understood as a debate about the concept of moral judgement.

3. Are emotions necessary for moral judgement?

The first more substantial question to be addressed, then, is: Are emotions really necessary for moral judgement, and do Prinz and Sauer provide us with good reasons for thinking so? There are three types of empirical evidence which Prinz and Sauer employ to support the necessity thesis:

- (i) evidence from studies with psychopaths, in particular the one provided by James Blair (1995; see also Blair et al. 1995),
- (ii) evidence from studies with patients suffering from frontotemporal dementia, such as the one forwarded by Mario Mendez et al. (2005), and
- (iii) evidence from studies with so-called 'acquired sociopaths', i. e. individuals who have suffered damages to the ventromedial cortex, which has been gathered over

the past decades by Antonio Damasio (1994) and his collaborators (see, in particular, Anderson et al. 1999).

According to Prinz, the evidence from psychopaths and individuals with frontotemporal dementia shows that impairments in emotional responsiveness lead to impairments in moral judgement, suggesting therefore “that, absent certain emotions, we lose the capacity to make moral judgments” (2014, 102). In a similar vein, Sauer claims that psychopaths and (some) subjects with acquired sociopathy cease to be morally competent subjects and lose the capacity for making full-blown moral judgements (2017, 172).

However, there are various reasons why things are not quite as easy. To begin with, it is not entirely obvious whether the emotional deficits of psychopaths and acquired sociopaths only affect their social *behavior* or also their ability to make moral judgements. A feature that undoubtedly characterizes psychopaths and acquired sociopaths is that they act contrary to widely accepted social norms. Although researchers examining the relevant behavior sometimes apply an unduly broad understanding of morality, it should be uncontroversial that at least some of these social norms qualify as moral norms and not just, for example, as etiquette norms. However, as far as the necessity thesis is concerned, what matters is not the behavior of psychopaths or acquired sociopaths, but the judgements they make. This distinction tends to get blurred in the empirical literature because Damasio and others fail to sufficiently distinguish between a person’s normative *judgements* and her *decisions* on how to act. Yet, unless we presuppose an implausibly strong version of motivational internalism according to which our normative judgements always provide us with overriding motivation to act accordingly, a person’s decisions and her subsequent actions will not tell us all we might want to know about the judgements she has made.

In order to determine whether psychopaths or acquired sociopaths provide evidence for the supposed emotional nature of moral judgement, therefore, one must examine their

judgements more directly. When it comes to the actual judgements of psychopaths or acquired sociopaths, however, the available evidence is by no means unambiguous. With regard to all of the three groups mentioned above, that is, psychopaths, acquired sociopaths and patients with frontotemporal dementia, the relevant studies describe cases of moral judgement – or at least cases of what suspiciously looks like moral judgement. Accordingly, some of the authors, like Mendez et al. (2005, 195f.), do *not* themselves subscribe to the view that their subjects displayed an inability to make moral judgements. Rather, they point to interesting ways in which their subjects' moral judgements substantially differ from the moral judgements of ordinary people, especially when it comes to responding to 'personal' moral dilemmas.

A second question to be asked, then, is: Do the studies of Blair, Mendez et al. or Damasio show that psychopaths, acquired sociopaths or patients with frontotemporal dementia cease to make any genuine moral judgements – or do they merely show that these individuals make different and perhaps normatively inappropriate moral judgements? The reason why the studies are often cited as evidence for the former view is that the subject's judgements, and in particular the judgements of psychopaths, deviate from 'ordinary' moral judgements in such a systematic fashion that some psychologists and philosophers consider it legitimate to no longer classify them as moral judgements at all.

Arguments to this effect are typically based on Elliot Turiel's moral/conventional distinction. According to a number of studies conducted by Turiel (1978a) and his collaborators (see, for example, Nucci and Turiel 1978), children from the age of 3 and normal adults evaluate violations of moral norms and conventional norms – such as etiquette norms – differently across a variety of dimensions.² First, they typically describe moral transgressions as more serious than conventional transgressions and judge the latter to be permissible more often than the former. Secondly, they treat moral norms as authority-independent, thinking that actions classified as moral transgressions would be wrong if they had not been forbidden by a person with authority such as a parent or a teacher, whereas actions classified as conventional

transgressions would not be wrong if there were no such prohibitions. Thirdly, they tend to view conventional rules relativistically in that they regard them as legitimately changeable from one setting to another, whereas they tend to view moral rules in universalist terms and as not being subject to human convention.

What makes psychopaths an interesting case is that they present an exception to this pattern. According to Blair (see 1995, 16f.), who confronted psychopaths and non-psychopaths with stories of moral and conventional transgressions and collected their judgements about permissibility, seriousness and authority-dependence, psychopaths do not make a significant moral/conventional distinction on any of these three criteria judgements while non-psychopaths make a significant distinction on all of them. According to one widely accepted interpretation, these findings reveal that psychopaths do not make any genuine moral judgements: Although they may be capable of making moral judgements in an ‘inverted commas’ sense, that is, of paying lip service to what others say about morality and ‘citing’ their judgements as if they were their own, they do not fully endorse the relevant judgements since they fail to systematically distinguish moral norms from social conventions.

The question raised above, therefore, leads over to a third question, namely what to make of Turiel’s and Blair’s evidence and whether to accept the specific interpretation that has been derived from it. According to Sauer, the view that the ability of making full-blown moral judgements hinges on the ability of making the moral/conventional distinction has been the object of “overwhelming agreement among philosophers and psychologists” (2017, 169). While Sauer may be right about this latter fact,³ I think that we have nevertheless good reasons to resist the view.

First, Blair himself has been unable to fully replicate his findings in a follow-up study: While in this second study, psychopaths made no significant moral/conventional distinction with regard to permissibility and authority judgements, they *did* make a significant distinction with regard to seriousness (see Blair et al. 1995, 745f.). Secondly, the interpretation that

psychopaths do not make any genuine moral judgements would be more plausible if psychopaths treated moral transgressions as conventional transgressions, thereby denying moral norms their true moral authority. As Blair explicitly emphasizes, however, the difference between psychopaths and non-psychopaths lies in the other direction: Psychopaths tend to *moralize* conventional transgressions, by treating the latter as authority-independent as well (1995, 17 and 20; see also Blair et al. 1995, 749). Finally, the view that by losing the ability of making the moral/conventional distinction one loses the ability of making full-blown moral judgements might actually be incoherent since the criterion judgements used in the experiments seem to qualify as first-order moral judgements in their own right.

Take the criterion judgements of non-psychopaths first. If a non-psychopath judges both hitting somebody in the face and interrupting another person to be impermissible, we may want to argue that only the first judgement is a proper moral judgement because in the latter case, the person is judging matters from a conventional rather than from a genuinely moral standpoint. However, if the person goes on to say that hitting somebody in the face is more serious than interrupting another person, it is natural to think that she can make this comparative judgement only if she evaluates the two forms of behavior from the very same standpoint and that this common standpoint is provided by the moral perspective – which means that the person’s judgement about seriousness qualifies as a genuine moral judgement.

When it comes to judgements about authority-dependence, things may be more controversial. Though some would think of judgements like ‘Hitting somebody in the face would be wrong even if the teacher permitted it’ as metaethical, second-order judgements, it has plausibly been argued – especially by modern-day sentimentalists such as Simon Blackburn and Allan Gibbard – that judgements about authority-dependence or objectivity are in fact first-order moral judgements as well. After all, when making such judgements, we merely specify the conditions under which a particular behavior would be morally wrong, and this seems to be a first-order moral affair.

Yet, if all or at least some of the criterion judgements of Blair's *non*-psychopaths qualify as genuine first-order moral judgements, the most natural interpretation would be that the criterion judgements of Blair's psychopaths are first-order moral judgements as well. Once we admit as much, however, we cannot go on to use the moral/conventional task as a means for distinguishing between moral and non-moral judgements: By doing so, we would effectively cite particular moral judgements of psychopaths as evidence for the fact that they do not make any moral judgements to begin with.

What would be required in order to avoid this incoherence is an explanation for why we should not apply the most natural interpretation. It is hard to see, however, what such an explanation could be. How can we justify interpreting the criterion judgements of psychopaths as *non*-moral judgements while interpreting the criterion judgements of non-psychopaths as *moral* judgements? How are we to show that, when psychopaths respond to the very same questions asked by Blair and his collaborators, they generally judge matters from a *non*-moral perspective and thus from a perspective crucially different from the one taken by non-psychopaths? The only strategy seems to be to *presuppose* what we are trying to prove, namely that, unlike non-psychopaths, psychopaths do not make any full-blown moral judgements and therefore *must* respond to the questions in non-moral terms. While this strategy manages to salvage the overall consistency of the interpretation, it would obviously render the argument circular.

It seems, then, that appealing to Turiel's moral/conventional distinction and the concurrent evidence from psychopathy does not provide any decisive support for the necessity thesis since such an appeal will turn out to be either incoherent or question-begging. In my view, therefore, the strongest *prima facie* support for the thesis derives from Anderson et al.'s research on early-onset acquired sociopathy. Anderson et al. found that patients who suffered damages to their ventromedial cortex at a very early age tend to evaluate social dilemmas from a purely egocentric perspective. Given that the intimate connection between morality and a

universal or impartial standpoint is even more widely accepted than Turiel's distinction between moral and conventional norm-violations, this evidence seems to allow us to conclude that early-onset patients are unable to make full-blown moral judgements even on a moderate and less demanding conception of what such judgements are all about.

Again, however, there are some problems with the relevant empirical evidence. First, as Anderson et al. themselves emphasize (see 1999, 1032), cases of damages to the ventromedial frontal cortex occurring before the maturation of the relevant neural and cognitive system, i. e. in infancy, are extremely rare. As a result, their study was conducted with only two patients. Given the small sample size, it cannot be ruled out that the deficits observed in the subjects were caused by some other factors and that further individuals with comparable damages would not display a similar kind of behavior.

The second, and more serious problem is that Anderson et al.'s subjects displayed a variety of *cognitive* deficits as well. Thus they showed significant impairments with regard to "several cognitive tasks designed to assess their ability to plan and execute multi-step procedures, use contingencies to guide behavior, reason through social dilemmas and generate appropriate responses to social situations" (Anderson et al. 1999, 1033). Early-onset patients, then, do not seem to develop or acquire the kind of knowledge about social and moral rules that, despite their various impediments, we find in adult-onset patients. At the very least, they are unable to retrieve this knowledge in a given situation (see Anderson et al. 1999, 1034). According to Anderson et al., therefore, the fact that both early-onset patients displayed forms of anti-social behavior, such as stealing and violence against persons and property, is not indicative of a "dissociation between disrupted social behavior and preserved factual social knowledge" (1999, 1032), as in the case of adult-onset patients. To the contrary, there actually seems to be a certain continuity or conformity between what the early-onset patients *know* and what they *do* in so far as they perform actions commonly considered to be wrong without fully understanding that they are so being considered.

In view of their cognitive deficits, then, the evidence from early-onset acquired sociopaths does not provide any conclusive support for the necessity thesis, even if we were to concede that early-onset acquired sociopaths are indeed incapable of making full-blown moral judgements. We might just as well appeal to the relevant evidence in order to argue that particular *cognitive* capacities are necessary for moral judgement. The more general conclusion, then, is that neither of the three types of evidence cited by Prinz and Sauer unambiguously supports the view that we need certain emotional capacities in order to make moral judgements.

4. Are emotions sufficient for moral judgement?

The conclusion of the previous section was that the existing empirical evidence does not provide us with any compelling reasons for thinking that emotions are necessary for moral judgement. However, even if we were to arrive at a more charitable assessment, the necessity thesis could not by itself prove the emotional character of moral judgement. The reason is that emotions could be necessary for moral judgement in a variety of different ways. To take just one example, in order to apply some of our most firmly established moral norms, namely those relating to harm, we need to have an understanding of what will hurt other people and make them suffer. That individuals whose emotional abilities are seriously impaired, and perhaps are so from an early age, might violate those norms or apply them in a manner that systematically differs from how other individuals apply them, then, is something any reasonable rationalist will expect.

Without concurring evidence that emotions are also sufficient for moral judgement, therefore, the necessity thesis does not seem to put much pressure on rationalist conceptions of moral judgement. In order to support the sufficiency thesis, Prinz and Sauer refer to two main kinds of empirical evidence, namely to

- (i) Haidt's (2001, 2013; see also Haidt et al. 2000) studies on 'moral dumbfounding', and
- (ii) studies examining the influence of incidental disgust and other emotions on moral judgement, such as the ones conducted by Wheatley and Haidt (2005) and Schnall et al. (2008), or the ones conducted by Eskine et al. (2011), and by Prinz and Angelika Seidel (2013a, 2013b).

The discussion of the relevant findings and their theoretical import is somewhat clouded by the fact that Prinz and Sauer are slightly ambiguous about what the sufficiency thesis actually amounts to. In order to assess the true merits of the supporting evidence, therefore, we need to strictly distinguish between a stronger and a weaker version of the thesis.

The most natural interpretation of the claim that emotions are sufficient for moral judgement is that we cannot be in certain emotional states without making a moral judgement. Or, to put things differently: Emotions are sufficient for moral judgment if, on their own account, they provide a sufficient explanation for why a moral judgement was made. The only evidence that arguably qualifies as evidence for this stronger version of the sufficiency thesis is provided by Haidt's 'moral dumbfounding' studies – at least as long as we accept Prinz's interpretation of the relevant findings. According to Prinz (2014, 101), Haidt's evidence demonstrates the sufficiency of emotions for moral judgement by showing that people sometimes make moral judgements in the absence of supporting reasons.

In their studies, Haidt and his collaborators presented American undergraduate students with stories of deviant sexual behavior (like the *incest case*) and other forms of taboo violations they classified as harmless yet disgusting (like the *cannibalism case*). When pressing them on their initial moral judgements, they found that many participants clung to those judgements even if they were shown that there were no good reasons to support them, often resorting to saying things like 'I don't know, I can't explain it, I just know it's wrong' or 'I know it's wrong,

but I just can't come up with a reason why'. According to Haidt and his collaborators, these subjects were morally 'dumbfounded'.

Haidt's findings and the 'social intuitionist model' of moral judgement he derives from it have been the object of much discussion in the past and attracted a great deal of criticism, especially from moral philosophers (see, for example, Kennett 2011; Brink 2014; and Railton 2014). In what follows, I will focus on two points. With regard to the question of whether instances of 'moral dumbfounding' support the strong version of the sufficiency thesis, it first seems worth pointing out that the methodology employed by Haidt and his collaborators did not really treat the affective responses induced in the subjects as representing moral judgements in their own right. After having confronted the subjects with the *incest case* or the *cannibalism case*, the subjects were explicitly *asked* to make a moral judgement ('was it OK for Mark and Julie to make love?' – 'is there anything wrong with what Jennifer did?'), and only their verbal responses to these questions were counted as instantiations of moral judgement (or as linguistic reports of the underlying mental events). In one important sense, therefore, the affective intuitions of Haidt's participants were not viewed as being in themselves sufficient for a moral judgement to be made.

It might be, of course, that the prior existence of the mental events in question was simply taken for granted by Haidt and his collaborators, and it would indeed be implausible to assume that none of the participants performed any personal moral evaluation of the *incest* and *cannibalism* cases before explicitly being asked to do so. However, in order for Haidt's study to support the strong sufficiency thesis, there would have to be some experimental procedure tracking the existence of these prior evaluations and showing them to perfectly correlate with the participants' intuitive affective responses. In the absence of such a procedure, Haidt's evidence allows for the possibility that some participants did *not* morally evaluate the two cases before being asked to do so, in spite of their experiencing a negative affective response. Yet, if

it is possible for moral judgement and negative affective response to come apart in this way, the strong sufficiency thesis would not be true.

Secondly, it is not at all obvious that ‘moral dumbfounding’ really occurred in Haidt’s study, not even in cases in which the subjects ultimately conceded not to have any good reasons for their judgements. In order for a moral judgement to represent an example of true ‘moral dumbfounding’, it is not sufficient that the reasons the subjects present as grounds for their judgements are bad reasons. What is required is that the subjects *themselves* believe that they possess no rational justification for the judgements while nevertheless clinging to them. Yet, the transcripts of the interviews conducted by Haidt and his collaborators suggest that even the subjects who ultimately resorted to saying things like ‘I don’t know why, but it’s simply wrong’ thought to have good reasons to judge as they did and that they made their judgements on the basis of these putative reasons. Moreover, it is by no means clear that, after being challenged by the interviewer, all subjects entirely parted with the considerations that provided them with their reasons.

In a transcript cited in Haidt’s book *The righteous mind*, the subject initially explains her disapproval of the two siblings Mark and Julie’s having sex by pointing out that Julie might get pregnant and that there is a high risk of her child suffering from disabilities (“if the girl did get pregnant, the kids become deformed, most of the time, in cases like that” [Haidt 2013, 46]). The interviewer then reminds the subject that Mark and Julie used both a condom and birth control pills and concludes that “there’s no way they’re going to have a baby” (46). While the subject goes along with this in what follows and tries to come up with *new* reasons, his or her first response (“Well, I guess the safest sex is abstinence...” [46]) suggests that he or she might not fully buy the interviewer’s conclusion, but believes that the considerations about pregnancy and disability still have a point because, where close relatives are concerned, even an extremely low risk of becoming pregnant needs to be avoided. Under this reading, the reason why the subject gives in to the experimenter and tries to come up with an alternative justification is not

that he or she has been entirely convinced by the interviewer that the original justification does not work. It is rather that the experimenter has clearly displayed his unwillingness to accept the justification or appreciate its possible merits. Of course, we might want to agree with the interviewer that the risks involved are still negligible and do not warrant the subject's moral disapproval. Yet, even if we view the subject's reasons as bad reasons, it would still be the case that he or she clings to these reasons and is not 'dumbfounded' in the relevant sense.

On the other hand, we may also wonder whether all subjects who *were* in fact convinced by the interviewer's objections wholeheartedly maintained their initial judgements. One possibility which is not really accounted for by Haidt et al. is that subjects who have purposefully been led into error with subtly modified examples of incest or cannibalism will be embarrassed about their inability to spot the relevant modifications or feel tricked and exposed by the experimenter. Once this possibility is acknowledged, it seems that we must expect at least some participants to stick to their initial judgements out of sheer defiance or shame. Rather than taking their final declarations for granted, therefore, we need to ask whether they honestly continued to morally disapprove of Mark and Julie's or Jennifer's behavior or merely pretended to do so.

One possible strategy for answering this question would be to repeat the study and examine the very same participants' responses to the two scenarios at some later point in time, preferably with a new experimenter. As Haidt presents things, we should expect the supposedly dumbfounded participants to simply repeat their negative judgements about the *incest* and *cannibalism* cases. Yet, it seems quite plausible that, in virtue of the temporal and personal distance to the earlier experience of social pressure and shame, some participants will actually alter their judgements and take the points of the original experimenter into account.

In fact, there are now a number of further empirical studies which substantiate the view that Haidt and his collaborators overstate their case and which provide strong evidence that true 'moral dumbfounding' is quite a rare occurrence (see, for example, Royzman et al. 2015; and

Stanley et al. 2019). While this insight would not directly refute the strong sufficiency thesis, it would certainly raise the question of why affective episodes like the ones experienced by Haidt et al.'s subjects sometimes manifest themselves in moral judgements and sometimes fail to do so. One natural assumption is that even those episodes which appear sufficient for moral judgements to be made will turn out to be so only in conjunction with further, yet unrevealed factors, which would mean that the strong version of the sufficiency thesis rests on an uncertain foundation.

It is not very surprising, therefore, that Prinz and Sauer ultimately endorse the sufficiency thesis only in a weaker form, namely as the thesis that emotions are capable of *affecting* or *altering* moral judgements, that is, of influencing moral judgements *in conjunction with* other causal factors. As already indicated, Prinz and Sauer do not explicitly distinguish between the two versions of the thesis. Sauer even suggests that there is no relevant distinction, by naturally alternating between the claim that “[e]motions and feelings alone are [...] enough to account for changes in moral judgment” (2017, 196) and the claim that “emotional reactions [...] are sufficient to explain people’s moral judgments” (196). It should be obvious, however, that these two claims are not equivalent and that the former, weaker version of the sufficiency thesis could be true without the latter, stronger version being true as well.

In order to support the weaker claim and show that “emotion induction causally influences moral judgments” (Prinz 2014, 101), Prinz and Sauer refer to Wheatley and Haidt’s and Schnall et al.’s studies on disgust and moral judgement as well as to several follow-up studies in which Prinz and his co-researchers tried to replicate these and similar findings. According to Prinz, the studies reveal “that disgusting beverages make moral judgments harsher [...], and that irritating music increases negative moral judgment, and uplifting music increases positive moral judgments” (2014, 101).

As in the case of Haidt’s ‘moral dumbfounding’ studies, there are now studies in which the findings of the original studies could not be replicated and meta-studies which conclude that

the induction of disgust or other emotions does *not* influence moral judgements in the kind of systematic fashion that the authors of the original studies describe (see Landy and Goodwin 2015; and Ghelfi et al. 2020). In what follows, however, I will proceed on the more charitable assumption that *some* causal influence of emotions on moral judgement can indeed be verified and that the weak version of the sufficiency thesis can therefore be defended. The true problem with limiting the sufficiency thesis will become apparent in the following section, where I will address the question of what can actually be achieved by conjoining the necessity thesis and the weak sufficiency thesis in the way Prinz and Sauer suggest.

5. What would it tell us?

The third and final question that needs to be addressed is whether the conjunction of the necessity thesis and the sufficiency thesis would provide a good basis for a (plausible) sentimentalist constitution claim. In virtue of Prinz and Sauer's quite lenient interpretation of the sufficiency thesis, this question becomes even more astute.

The first point to be made here is that the claim that moral judgements are constituted by emotions is not a logical or conceptual implication of the conjunction of necessity thesis and weak sufficiency thesis. Obviously, emotions can both be necessary for moral judgements and capable of affecting or altering them without moral judgements being emotions themselves. If anything, then, the third premise and the resulting conclusion can only be defended in the sense of an inference to the best explanation.

However, even such an interpretation faces some serious problems. One problem derives from the fact that, on the evidence cited by Prinz and Sauer, different emotions affect moral judgements differently. According to the studies conducted by Prinz, "disgust increases the perceived wrongness of crimes against nature" (2015, 72) whereas "anger increases the perceived wrongness of crimes against persons" (72; see also Prinz and Seidel 2013a).

Moreover, while induced happiness increases positive but not negative moral judgements, induced anger has a reverse effect (Prinz 2015, 72; see also Prinz and Seidel 2013b). Yet, if different emotions affect different types of moral judgement differently, what kind of constitution claim are we to infer from this? The claim that moral judgements about crimes against nature are constituted by feelings of disgust – whereas moral judgements about crimes against persons are constituted by feelings of anger? And the claim that positive moral judgements are constituted by neither of these feelings, but by feelings of happiness?

Though Prinz does not explicitly address the difficulties arising from the selective effects of different emotions on moral judgement, he admits that these effects “are not the result of a more general relationship between emotion and judgment” (2015, 72), thereby suggesting that the empirical evidence does not support *one* sentimentalist constitution claim, but rather a combination of such claims. Now the point here is not so much that it is impossible to come up with a version of sentimentalism that integrates the different pieces of evidence. Rather, it is that sentimentalists still have a lot of explaining to do and that their ultimate conjunction of constitution claims might lack overall plausibility. If the conception just sketched is the one we are to positively endorse, then we at least need an independent explanation for why we should categorize the different kinds of feeling under a common name and equally treat the relevant manifestations of disgust, anger and happiness as instantiations of moral judgement. We also need an explanation for why many *other* manifestations of disgust, anger and happiness do *not* deserve to be classified as moral judgements.

Simply equating moral judgements with emotions of disgust, anger or happiness is also problematic for another reason. By doing so, we would actually get into contradiction with the sufficiency thesis, since the latter posits a *causal* relation between these emotions and moral judgement – which relation could not obtain if the moral judgement actually *consisted* in the emotion. That the induced emotional response and the moral judgement are, after all, two

different things is also suggested by the way in which Prinz sums up the import of the relevant studies:

All this suggests that people use emotions as information when they decide whether something is right or wrong: when asked to make a moral evaluation, people introspect and report the intensity of their feelings (Prinz 2014, 101).

In order for the argument to work, then, one needs to defend a further claim to the effect that the second *relatum* of the causal relation is instantiated by an emotion as well. It should be obvious, however, that emotions can causally affect not only other emotions but also beliefs. Possible examples extend from the general phenomenon of wishful thinking to the specific ways in which fear may affect our beliefs about the probability of a terrorist attack or a plane crash.

Now one might think that the necessity thesis can provide some help in defending this further claim. As already emphasized, however, even if emotions were in fact necessary for moral judgement, they could be so in a variety of different ways, many of which will fail to support the idea that what is causally affected by incidental disgust, anger or happiness must itself be an emotion. A further difficulty arises from the fact that these latter emotions are quite different from the ones which typically provide the focus of the empirical studies that relate to the necessity thesis. The supposed inability of psychopaths or acquired sociopaths to make proper moral judgements is not usually explained by their inability to experience disgust, anger, happiness or some related kind of emotion, but rather by their inability to empathize and feel compassion or pity. The only way for conjoining the necessity thesis and the weak sufficiency thesis in a way that would support a sentimentalist constitution claim, therefore, seems to be to equate moral judgements with empathetic or sympathetic emotions and to claim that these latter

emotions are systematically affected by emotions of disgust, anger and happiness. Yet, it is by no means obvious that such a systematic link exists.

Moreover, if there is some intimate link between the experience of disgust, anger and happiness and the experience of empathetic emotions, then we must expect the former to also affect empathetic beliefs, which means that the link could be exploited by rationalist theories of moral judgement as well. Finally, it needs to be emphasized that the suggested conception would not strengthen the more specific cases of Prinz and Sauer: Sauer's 'sophisticated emotionism' does not assign any crucial role to empathetic or sympathetic feelings, and Prinz explicitly rejects the view that moral judgements are intimately related to such feelings in his paper *Against empathy* (2011).

6. Classic sentimentalism, sophisticated emotionism – or neither

The conclusion of the above examination of Prinz's and Sauer's argument, then, is that there is no compelling empirical evidence for either of the two fundamental claims on which the argument relies. Moreover, there are good reasons to doubt that, even if the necessity and the sufficiency thesis were true, they would provide a sound basis for a plausible sentimentalist constitution claim.

It may be for this reason that Prinz's actual descriptions of the nature of moral judgement are somewhat vague and often fall short of the constitution claim he officially wants to defend. The overall conclusion Prinz draws from his review of the existing empirical evidence is that emotions are "component parts" (2014, 102) or "components" (2015, 73) of moral judgement. While this idea may entitle Prinz to the colloquial statement that "to think that something is morally right or wrong is to have an emotional attitude toward it" (2015, 73), it does not yet amount to the claim that "[t]he judgment that something is wrong *consists in* a negative feeling toward it" (2014, 102). This latter way of putting things seems misleading even if one conceives

of emotions as “essential components” (2015, 73) of moral judgement since it fails to do justice to the fact that, according to Prinz’s own terminology, moral judgements must consist of other component parts as well.

Moreover, Prinz’s additional claims, according to which emotions play a “predicative role” (2015, 73) in moral judgements and are “used as information” (73) when moral judgements are made, all leave the possibility that moral judgements are constituted by genuine beliefs. In order to accept these claims, one need not even defend a hybrid conception of moral judgement. Even more straightforwardly rationalist conceptions seem to be in a position to do so. This especially holds for conceptions which conceive of the beliefs that constitute moral judgements as beliefs about the emotions or feelings that we or other persons would have in specific circumstances, such as in a state of full rationality or impartiality.

It has already been emphasized that Sauer’s conception of moral judgement is even less sentimentalist than Prinz’s. Though Sauer accepts the key idea that moral judgements are constituted by emotions or affective intuitions, he distinguishes genuine moral judgements from “*mere* preferences or likes or dislikes” (2017, 198; see also 203 and 220), arguing that moral judgements involve “at least some degree of normative abstraction” (198). Exercising our rational capacities, according to this view, is co-constitutive for our emotional responses to count as genuine moral judgements (see Sauer 2017, 193). According to Sauer’s formal characterization, ‘sophisticated emotionism’ and the related ‘educated intuitions account of moral judgement’ ultimately amount to the following claim: “It is necessary and sufficient for making a moral judgment to be disposed to an emotional attitude of (dis)approbation towards certain actions that causes the judgment in a way that can be reflectively endorsed by the judging person under ideal conditions of full information and rationality” (Sauer 2017, 201).

As in Prinz’s case, however, this characterization and Sauer’s further statements seems to run together quite different ideas. Moreover, there seem to be some internal problems with the way in which Sauer tries to reconcile the psychological constitution claim with the

conceptual point that genuine moral judgements are the upshot of our rational capacities. First, Sauer's definition of 'sophisticated emotionism' suggests that what actually causes the moral judgement is the "emotional attitude of (dis)approbation" that is itself a manifestation of an underlying disposition. This idea, namely that the moral judgement is *causally produced* by the relevant emotion, is suggested by other statements as well (see, for instance, Sauer 2017, 203). However, if this is Sauer's idea, the same difficulty arises which we already discussed with regard to Prinz: The emotion and the judgement that is being caused by it appear to be two different things and it is not clear why we need to conceive of the judgement as an emotion as well.

Secondly, it is not clear that the counterfactual aspect introduced by Sauer coheres well with the idea that moral judgements are constituted by feelings or emotions. One way of interpreting Sauer's claim that genuine moral judgements are caused "in a way that can be reflectively endorsed by the judging person under ideal conditions of full information and rationality" would be to straightforwardly identify moral judgements with the emotional responses or 'educated intuitions' we would have if we were fully rational and fully informed. The problem with such a counterfactual approach, however, is that Sauer's explicit intention is to describe the psychological nature of moral judgement, that is, to describe moral judgement in terms of empirical psychological fact. Yet, to say that genuine moral judgement is represented by those emotional responses we would have if we were quite different from how we actually are is equivalent to saying that, as a matter of psychological fact, there are no genuine moral judgements.

One strategy for avoiding this unwanted conclusion is to identify moral judgements with emotional responses we actually have, but to only count those emotional responses as genuine moral judgements which we also ratify from a superior cognitive position. There are some passages which suggest that this is closer to what Sauer has in mind. For example, Sauer sometimes refers to the 'educated intuitions' that constitute genuine moral judgements as

habitualized responses that are stable over time and withstand further rational reflection (see, for instance, Sauer 2017, 48 and 63f.).

However, not only is this constraint noticeably weaker than the constraint of full information and rationality introduced by Sauer's official definition of 'sophisticated emotionism, it also raises problem of its own. The obvious advantage is that it allows us to conclude that, psychologically speaking, we frequently make genuine moral judgements. A problem that remains, though, is that we cannot be *sure* that the particular judgements we make are genuine moral judgements – at least not when we are making them for the first time. It is only in retrospect that we find out whether or not we continued to endorse our initial judgements, which means that it is only in retrospect that we will find out whether or not we made a genuine moral judgement. Although Sauer seems willing to accept this latter conclusion, it seems at odds with our ordinary moral practice. This practice seems to be not only characterized by the presupposition that we do make genuine moral judgements and can identify those judgements in retrospect, but by the presupposition that we can consciously make new genuine moral judgements and that we (and our listeners) can instantly recognize whether we have done so. If we take Sauer's proposal seriously, we would have to give up this latter presupposition because the exact status of our judgements remains precarious.

It is natural to think that we can make up for this problem of Sauer's proposal by adding one further assumption, namely that, by the time an individual makes a judgement, she conceives of it as one she will also endorse from a superior cognitive position or as one that we will continue to hold over time. However, this further assumption is by no means innocuous. The reason is that the assumption turns what was intended as a sentimentalist constitution claim into a hybrid conception of moral judgement. On this conception, our moral judgements would be co-constituted by *both* an emotional response to the kind of behavior we are considering *and* a belief to the effect that we will continue to experience this emotional response and ratify it. Yet, if this is how we need to conceive of genuine moral judgement in order to make the

emotionist element work, the obvious question is why we should accept the emotionist element in the first place? Why not conceive of genuine moral judgements as beliefs about the emotional responses we would experience under certain conditions or experience over time? Such an account seems to provide us with all the advantages of Sauer's conception and none of its disadvantages – such as the insufficiently supported necessity claim. It can do justice to the fact that we sometimes make moral judgements without experiencing any occurrent emotional responses. Moreover, it would not even face problems with a more demanding counterfactual constraint in terms of full information and full rationality. According to the conception under consideration, actual moral judgements are not instantiated by counterfactual emotions, but by counterfactual beliefs, that is, by beliefs about counterfactuals. Unlike the emotions we would have if we were fully informed and fully rational, such beliefs are psychologically real even if their content is not. On closer inspection, then, the idea that individuals who make a moral judgement believe that they will emotionally respond to the evaluated behaviour in a particular way does not salvage Sauer's version of 'sophisticated emotionism' but ultimately undermines it.

7. Conclusion

Many claims to the effect that empirical evidence from social and developmental psychology and neuroscience provides a vindication of moral sentimentalism are based on what we may refer to as *shot-gun methodology*. When particular studies that strive to reveal a crucial link between moral judgements and emotions or feelings come under attack by moral philosophers, a common response is to concede the point, but refer to the many other studies that have investigated the contributions to moral judgement made by the affective side of human psychology, thereby suggesting that the sheer breadth of emotion-oriented empirical research

will ensure that some bullets will hit their target, i. e. that some version of moral sentimentalism must be true.

Despite its general shortcomings as a scientific methodology, the above view suffers from the fact that many authors work with a vague and uninteresting notion of sentimentalism which amounts to nothing more than the idea that affective or conative states of mind are somehow essential or crucial to morality. In accordance with that, studies intended to refute moral rationalism are often based on caricatures of the views moral rationalists actually hold and reveal links between moral judgement and the affective side of human psychology that no rationalist worth his or her salt would ever have denied.

What makes Prinz's and Sauer's approaches so interesting is that they promise us to get beyond the shot-gun methodology, by integrating a number of independent findings in such a systematic manner as to allow for a principled yet targeted refutation of one of the key ideas of moral rationalism, namely the idea that moral judgements are constituted by genuine beliefs. If what I have argued above is correct, however, Prinz and Sauer cannot make good on this promise. If we do not want to entrust ourselves to the shot-gun methodology, therefore, there seems to be nothing we can do than to compare specific rationalist and sentimentalist conceptions of moral judgement and to see which one of them ultimately achieves the best or most comprehensive reflective equilibrium with regard to the more reliable part of the available empirical evidence and our pre-empirical conceptions of what moral judgements are and how they work.

In my discussion of Sauer's 'sophisticated emotionism' in section 6, I have already hinted at what a plausible rationalist conception of moral judgement might look like. The crucial idea is to conceive of moral judgements as manifestations of beliefs *about* emotions or other non-cognitive states of mind, not as manifestations of non-cognitive states of mind themselves. There are several ways in which this general idea can be and has been developed in the past.

The variant I favour conceives of moral judgements as beliefs about the justifiability of overriding desires.⁴

While it is common to refer to such approaches as variants of ‘neosentimentalism’ (see d’Arms and Jacobson 2000; see also Kauppinen 2014), I consider this description to be somewhat misleading. As we have seen, both Prinz’s and Sauer’s discussions of sentimentalism focus on the question of which kind of mental states *constitute* moral judgements. This is very much in line with the classical 18th century debate. Moral sentimentalists such as Hutcheson, Hume or Smith all grant that reason makes an important causal contribution to moral judgement, just as moral rationalists like Clarke, Burnet or Price grant that moral judgements are accompanied by, and give rise to, specific moral feelings. The point of contention between the classic sentimentalists and the classic rationalists is not whether moral judgements crucially relate to cognitive or non-cognitive states of mind, but how they themselves are instantiated. The constitution claim sketched above, namely that moral judgements are constituted by beliefs about emotions or other non-cognitive states of mind, should therefore be classified, not as a cognitivist variant of sentimentalism, but as a variant of moral rationalism.

It goes without saying that rationalist conceptions of this kind face problems of their own and ask for a more comprehensive examination than can here be provided. The point to be emphasized in the present context, however, is that they seem in quite a good position to account for the empirical evidence reviewed above. It is highly plausible to assume that emotions or desires we actually have will systematically influence our views about which emotions or desires it would be appropriate, rational or justifiable to have. Something similar holds for the emotions and desires we empathetically or sympathetically ascribe to others. Individuals experiencing strong feelings of disgust when confronted with the *incest* or *cannibalism* cases, therefore, can be expected to form negative moral judgements about the behaviour of Mark, Julie or Jennifer even if these judgements are constituted by beliefs about the appropriate affective reactions to these cases and not by affective reactions themselves. Similarly, we can

expect individuals whose capacities for experiencing certain feelings or understanding and sympathizing with the feelings of others are seriously impaired to arrive at judgements that noticeably differ from the judgements of those whose capacities are not so impaired.

What the empirical evidence may show, then, is that emotions and other non-cognitive processes bear on the moral judgements we make because, as suggested by Prinz himself, we use them as information or heuristics for what it would be appropriate to do or feel. What it does not show – at least not in its current form – is that moral sentimentalism must be true.

Acknowledgements: Earlier versions of this paper were presented to audiences at the University of Konstanz and the University of Cologne. I would like to express my gratitude for the feedback received on these occasions. Also, I am greatly indebted to Philipp Hübl as well as to two anonymous referees for *Philosophical Explorations* for several helpful suggestions on how to improve the paper.

References:

Anderson, S. W., A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio. 1999. “Impairment of social and moral behavior related to early damages in human prefrontal cortex.” *Nature Neuroscience* 2: 1032–37.

Blair, R. J. R. 1995. “A cognitive developmental approach to morality: Investigating the psychopath.” *Cognition* 57: 1–29.

Blair, R. J. R., L. Jones, F. Clarke, and M. Smith. 1995. “Is the psychopath ‘morally insane’?” *Personality and Individual Differences* 19 (1995): 741–52.

Brink, D. O. 2014. “Principles and intuitions in ethics: Historical and contemporary perspectives.” *Ethics* 124: 665–94.

d’Arms, J., and D. Jacobson. 2000. “Sentiment and value.” *Ethics* 110: 722–748.

- Damasio, A. R. 1994. *Descartes' error: Emotion, reason, and human brain*. New York: Putnam & Sons.
- Eggers, D. 2019. "The language of desire. Expressivism, sentimentalism and the role of reason in moral judgement." *Habilitation thesis*, University of Cologne.
- Eskine, K. J., N. A. Kacirik, and J. J. Prinz. 2011. "A bad taste in the mouth: Gustatory disgust influences moral judgment." *Psychological Science* 22: 295–99.
- Ghelfi, E., C. D. Christopherson, H. L. Urry, R. L. Lenne, N. Legates, M. A. Fischer, F. M. A. Wagemans et al. 2020. "Reexamining the effect of gustatory disgust on moral judgment: A multilab direct replication of Eskine, Kacirik and Prinz (2011)." *Advances in Methods and Practices in Psychological Science* 3: 3–23.
- Haidt, J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment" *Psychological Review* 108: 814–34.
- Haidt, J. 2013. *The Righteous Mind. Why good people are divided by politics and religion*. London: Penguin.
- Haidt, J., F. Björklund, S. Murphy. 2000. "Moral dumbfounding: When intuition finds no reason." <https://polpsy.ca/wp-content/uploads/2019/05/haidt.bjorklund.pdf>
- Kauppinen, A. 2014. "Moral sentimentalism," *Stanford Encyclopedia of Philosophy*: 1–89.
- Kennett, J. 2011. "Imagining reasons." *The Southern Journal of Philosophy* 49: 181–92
- Landy, J. F., and G. P. Goodwin. 2015. "Does incidental disgust amplify moral judgment? A meta-analytic review of empirical evidence." *Perspectives on Psychological Science* 10: 518–36
- Mendez, M. F., E. Anderson, and J. S. Shapira. 2005 "An investigation of moral judgment in frontotemporal dementia." *Cognitive and Behavioral Neurology* 18: 194–97.
- Nichols, S. 2004. *Sentimental rules. On the natural foundations of moral judgment*. New York: Oxford University Press.
- Nucci, L., and E. Turiel. 1978. "Social interactions and the development of social concepts in

- preschool children,” *Child Development* 49: 400–407.
- Prinz, J. J. 2006. *Gut reactions: A perceptual theory of emotion*. New York: Oxford University Press.
- Prinz, J. J. 2007. *The emotional construction of morals*. New York: Oxford University Press.
- Prinz, J. J. 2011. “Against empathy.” *The Southern Journal of Philosophy* 49: 214–33.
- Prinz, J. J. 2014. “Where do morals come from? A plea for a culturalist approach,” in *Empirically informed ethics: Morality between facts and forms*, ed. by M. Christen et al. Cham: Springer: 99–116.
- Prinz, J. J. 2015. “An empirical case for motivational internalism.” in *Motivational internalism*, ed. by G. Björnsson et al. New York: Oxford University Press: 61–84.
- Prinz, J. J., and A. Seidel. 2013a. “Mad and glad: Musically induced emotions have divergent impact on morals.” *Motivation and Emotion* 37: 629–37
- Prinz, J. J., and A. Seidel. 2013b. “Sound morality: Irritating and icky noises amplify judgments in divergent moral domain,” *Cognition* 127: 1–5.
- Railton, R. 2014. “The affective dog and its rational tale: Intuition and attunement.” *Ethics* 124: 813–59.
- Royzman, E. B., K. Kim, and R. F. Leeman. 2015. “The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect.” *Judgment and Decision Making* 10: 296–313.
- Sauer, H. 2012. “Psychopaths and filthy desks: Are emotions necessary and sufficient for moral judgment?” *Ethical Theory and Moral Practice* 15: 95–115.
- Sauer, H. 2017. *Moral judgments as educated intuitions*. Cambridge, MA: The MIT Press.
- Schnall, S., J. Benton, and S. Harvey. 2008. “With a clean conscience: Cleanliness reduces the severity of moral judgments.” *Psychological Science* 19: 1219–22.
- Stanley, M. L., S. Yin, and W. Sinnott-Armstrong. 2019. “A reason-based explanation for moral dumbfounding.” *Judgment and Decision Making* 14: 120–29.
- Turiel, E. 1978a. “Social regulations and domains of social concepts.” in *New Directions in*

Child Development. Vol.1: Social Cognition, ed. by W. Damon. San Francisco: Jossey-Bass: 45–7.

Turiel, E. 1978b. “The development of concepts of social structure: Social-convention,” in *The development of social understanding*, ed. by J. Glick and A. Clarke-Stewart. New York: Gardner Press: 25–107.

Wheatley, T, and J. Haidt. 2005. “Hypnotic disgust makes moral judgments more severe.” *Psychological Science* 16: 780–84.

¹ Note, however, that this term can be found in Prinz’s earlier work as well (see, for example, Prinz 2007, 13).

² My following account is based on Turiel’s own summary of his findings in Turiel 1978b, 32–34.

³ For a philosophical approach which draws heavily on the distinction, see Nichols 2004.

⁴ This view is developed and defended in Eggers 2019.