Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment?

Author(s): Hanno Sauer

Source: *Ethical Theory and Moral Practice*, Vol. 15, No. 1, Sen's The Idea of Justice (February 2012), pp. 95-115
Published by: Springer
Stable URL: https://www.jstor.org/stable/41474509
Accessed: 15-05-2020 21:07 UTC

Ethic Theory Moral Prac (2012) 15:95–115
DOI 10.1007/s10677-011-9274-y

Psychopaths and Filthy Desks

Are Emotions Necessary and Sufficient for Moral Judgment?

Hanno Sauer

Accepted: 8 March 2011 / Published online: 24 March 2011

Abstract Philosophical and empirical moral psychologists claim that emotions are both necessary and sufficient for moral judgment. The aim of this paper is to assess the evidence in favor of both claims and to show how a moderate rationalist position about moral judgment can be defended nonetheless. The experimental evidence for both the necessity- and the sufficiency-thesis concerning the connection between emotional reactions and moral judgment is presented. I argue that a rationalist about moral judgment can be happy to accept the necessity-thesis. My argument draws on the idea that emotions play the same role for moral judgment that perceptions play for ordinary judgments about the external world. I develop a rationalist interpretation of the sufficiency-thesis and show that it can successfully account for the available empirical evidence. The general idea is that the rationalist can accept the claim that emotional reactions are sufficient for moral judgment just in case a subject's emotional reaction towards an action in question causes the judgment in a way that can be reflectively endorsed under conditions of full information and rationality. This idea is spelled out in some detail and it is argued that a moral agent is entitled to her endorsement if the way she arrives at her judgment reliably leads to correct moral beliefs, and that this reliability can be established if the subject's emotional reaction picks up on the morally relevant aspects of the situation.

Keywords Moral judgment · Moral emotions · Moral psychology · Experimental philosophy · Jesse Prinz · Jonathan Haidt

1 Introduction

I am walking along Amsterdam's *grachtengordel*, the city's world-famous belt of canals. I notice the splendid patrician mansions, stop by at one of the cozy "brown cafés", I am intrigued by the florilegium of small shops offering antique furniture, vintage clothing and

H. Sauer (✉)
Instituut voor Wijsbegeerte, Universiteit Leiden, Matthias de Vrieshof/Witte Singel 25, 2311 BZ Leiden, Netherlands
e-mail: h.c.sauer@hum.leidenuniv.nl

other curiosities until I accidentally pass by the *Torture Museum*; I take a quick look at the poster on the museum's front door, and the information I gather about the chairs, masks, cages, forks and blades that were used for torment and inquisition makes me think: "This is just wrong!" But—what am I thinking here? What state of mind am I in? How did I arrive at my judgment? And, not least, is it justified?

Very roughly speaking, philosophical metaethics offers two different answers to these questions: rationalism and sentimentalism. Rationalism claims that, in thinking "Torture is wrong!", I am thinking that torture is wrong; that I, and any other person, ought not to do it and that there is good reason for this; that it cannot be willed without contradiction to torture people or that it is simply immoral and, hence, irrational. It claims that I am in a cognitive state of mind, comparable to judgments about the most efficient means to cause the strongest pain in an unlucky suspect, yet thoroughly *sui generis*. And it claims that I have arrived at my judgment through careful weighing of reasons and conscious reflection, and that I have adopted the single one attitude towards torture that finally survived my incorruptible scrutiny. Sentimentalism, on the other hand, claims that I am in a state of emotional arousal, that I have arrived at my judgment unconsciously, through an emotionally triggered disgust response towards torture and the contagious distress that the images of excruciated bodies have caused in me; and that my judgment "Torture is wrong!" is supposed to spread the word and make other people feel the same kind of disapproval towards this atavistic practice.

Driven to the extreme in that way, both answers start to turn into caricatures. The truth is to be sought somewhere in the middle between the two. Rationalism is strongest in highlighting the distinctive role of reason, reflection and self-criticism. Without these, we wouldn't understand our practice of moral judgment anymore (and we wouldn't like it, either). Sentimentalism seems to be an attractive position when it comes to explaining the irreducibly emotional dimension associated with moral judgment: we care about our values, very deeply indeed, and we experience anger or guilt upon the violation of the moral norms we endorse. How could it not be true that the realm of moral values and norms is not the object of cool and sober contemplation, but of passionate engagement and emotional commitment? Sentimentalism tries to embrace this connection between emotional reactions and moral judgments. When pressed, however, it turns out that both positions are having a hard time explaining what their account of moral judgment really amounts to, and how the alleged association between emotion, or reason, respectively, and morality can be understood in a way that is neither trivial nor plain false.

Recently, empirical moral psychology has taken up that challenge, trying to leave the empty space of empirically frictionless conceptual analysis and to offer an empirical, scientifically credible account of the relation between emotional reactions and moral judgment. Experimental philosophers (Knobe and Nichols 2008; Appiah 2008) have paid a lot of attention to empirical findings about moral judgment and agency, claiming that philosophical meta-ethics must not ignore them and that metaethical questions can and have to be addressed in a descriptively adequate fashion. Normative ethics on the other hand, experimental philosophers claim, does not tell us anything informative and interesting about the real life of human beings out of flesh and blood if it is based on a stark opposition between "is" and "ought". If "ought" implies "can", then "cannot" implies "ought not". From this, it is only a small step to "if never happens, then probably cannot, then probably ought not". A moral theory that tells us what people ought to do, regardless of whether *anybody* has *ever* done it, or will ever do it, must eventually fail. Or so experimental philosophers argue.

Empirically informed accounts of moral judgment and reasoning—like Jonathan Haidt's social intuitionist-model (Haidt 2001) or Joshua Greene's dual process-model (Greene et al.

2001)—promise to facilitate clear and conclusive answers in the long-standing debate between rationalism and sentimentalism about moral judgment. They use scientifically respectable methods and publicly verifiable procedures, particularly highly sophisticated social psychological and neuroscientific experiments. For obvious reasons, there has been a trend to draw broadly sentimentalist conclusions from the available evidence: the philosophical account of moral judgment that empirical moral psychology seems to favor is the view that emotional and intuitive "gut" reactions rather than genuine moral reasoning are crucial for normative judgment. As Daniel Jacobson puts it, "social intuitionism, considered as a thesis of moral psychology, best coheres with a *sentimentalist* metaethical theory, which holds that (many) evaluative concepts must be understood by way of human emotional response" (Jacobson 2008: 220). This trend has led to what might be called a psychological debunking of morality: the view that the importance of moral reasoning, similar to the experience of free will, can be shown to be a comforting, yet self-deceptive illusion. On that account, rationalism describes a practice of moral reasoning that, as the experimental evidence suggests, has no effect in the formation of moral judgment whatsoever, but serves to provide *post hoc* justifications for emotionally triggered intuitive judgments. In order to support this radical conclusion, however, one is committed to argue for a strong connection between emotional reactions and moral judgment; it does not suffice to show that moral judgment is usually *accompanied* by emotions, that it is sometimes, or even usually, *influenced* by emotions or that emotional reactions can *trigger* moral judgments. All this is trivial and uncontested. Call the claim that moral judgments are in some sense *essentially* connected to emotions "emotionism" about morality. What empirical moral psychology has to argue for, then, is a strong version of emotionism about morality: the thesis that having an emotional reaction of approval or disapproval towards morally relevant objects (typically actions) is both *necessary* and *sufficient* for moral judgment.

In the following paper, I address the question whether strong emotionism can be defended, that is, whether emotions really are necessary and sufficient for moral judgment. I shall answer this question affirmatively, but in a very specific way, namely—unlike Jesse Prinz, for example (Prinz 2006, 2007)—in a way that does justice to our most basic normative intuitions about the nature of moral judgment and is compatible with moderate rationalism about moral judgment. The structure of the paper is as follows: I will briefly present and assess the experimental evidence in favor of both the necessity-(1) and the sufficiency-thesis (2) concerning the connection between emotional reactions and moral judgment. I shall then develop a reading of both claims that I deem to be satisfactory for the moderate rationalist. I argue that a rationalist about moral judgment can be happy to accept the necessity-thesis (3). My argument draws on the idea that emotions play the same role for moral judgment that perceptions play for ordinary judgments about the external world. Provided an empirically adequate and normatively convincing interpretation is available, the same holds for the sufficiency-thesis. I develop such an interpretation and show that it can successfully account for the available empirical evidence (4). The general idea is that the rationalist can accept the claim that emotional reactions are sufficient for moral judgment just in case a subject's emotional reaction towards an action in question causes the judgment in a way that can be reflectively endorsed under conditions of full information and rationality (5). I spell out this idea in some detail and argue that a moral agent is entitled to her endorsement if the way she arrives at her judgment reliably leads to correct moral beliefs, and that this reliability can be established if the subject's emotional reaction picks up on the morally relevant aspects of the situation (6). One the face of it, this proposal seems committed to a strong form of moral realism. I show why it is not (7), and conclude

with some remarks about the general prospects of the account I have presented, and why one ought to prefer it to other accounts (8).

2 The Necessity-Thesis: Psychopathy and the Moral/Conventional Distinction

Moral judgments are related to norms, and norms are related to human action. Moral norms are prescriptive: they specify what *ought* to happen or what *ought* to be done. Examples for prescriptive norms are the rules of the road as well as norms of etiquette and, in general, rules that govern the interactive space between persons. But different prescriptive norms can be quite different in nature, depending in part on the source of their authority. Take, for example, the norm to shake a person's right hand upon meeting her. And now take the norm not to torture people out of boredom. Both norms are prescriptive. But intuitively it seems that the validity of the first norm depends upon a mere convention. We can say that *you ought to shake a person's right hand upon meeting her.* But if, instead of shaking somebody's right hand, a different norm, say, to shake a person's *left* hand, would be in place, it would not at all be wrong not to shake somebody's right hand. It is a mere convention. Emotionists about morality claim that in order to fully understand the fundamental difference between conventional and non-conventional norms, one must be able to experience certain emotional reactions towards their transgression. They claim that in order to grasp the specific moral authority of certain norms, one must be susceptible to experience guilt or outrage upon their violation. More generally, being susceptible to feel an emotional reaction towards certain types of norm-transgressions is seen as a psychologically necessary feature, an enabling condition for moral judgment. Call this the necessity-thesis.

There is overwhelming agreement among philosophers and psychologists about the fact that for a person to be able to make moral judgments she must be able—among other things—to draw a distinction between moral and conventional norms (Nichols 2004: 3ff.; Nucci 1985). On Shaun Nichols' "sentimental rules" account, for example, the capacity for "core moral judgment" is introduced along the very lines of the moral/conventional distinction; moral judgment requires the capacity to understand a certain subclass of prescriptive social rules as non-conventional, transgressions of these norms as more serious, generalizably wrong (that is, in other countries or communities as well) and the validity of these rules as neither based on social acceptability nor dependent on authority. All these criteria spell out the non-conventional validity of moral norms.

Ever since the days of Phineas Gage, research on psychopathy and so-called acquired sociopathy has provided the best and most robust evidence for the thesis that a certain kind of emotional engagement is necessary for moral judgment and behavior (Saver and Damasio 1991; Damasio 1994). In psychopaths, we find two things combined: first, a highly impaired emotional make-up, and second, a reliably poor performance on tasks to draw the moral/conventional distinction (see the classical research conducted by Turiel 1983 and Blair 1995; for more recent discussions of the phenomenon, see Hare 1999 and Blair et al. 2005). Many empirical moral psychologists claim that the former is an enabling condition of the latter. In order to show that psychopaths are unable to draw the moral/conventional distinction, James Blair predicted that psychopathic patients will have difficulties not only to draw the distinction, but that this incapacity will also show on the justificatory level: their justification for why a certain norm transgression is "wrong" will be "less likely to make references to the pain or discomfort of victims that the non-psychopath controls" (Blair 1995: 13). Another prediction was that psychopaths were likely to treat

moral rules as conventional rules. Surprisingly, the latter prediction turned out to be false. Subjects were presented with several stories (a child hitting another child in the moral case, a child talking in class in the conventional case) they had to assess in light of the question whether the described transgressive behavior should be seen as a violation of a moral or a conventional norm. Blair found that psychopaths treated all transgressions as moral and the validity of the transgressed rules as authority-independent. But this doesn't, according to Blair, show that psychopaths fail to make the moral/conventional distinction because of an *increased* moral sensitivity, or an increased tendency to empathize with the victims of moral transgressions. One has to bear in mind that all test subjects (psychopaths as well as non-psychopaths) were inmates (most of them convicted murderers); they simply had a strong motivation to demonstrate that they improved their or even acquired new social and moral knowledge through the received treatment, and this made them overshoot the target. Psychopaths do fail to grasp this important feature of genuine moral judgment, only in an unexpected way.

The most natural explanation for this seems to be that psychopaths lack the capacity to "feel" the special character of violations that are wrong, regardless of what an authority thinks about them. This specifically moral sensitivity is developed at an early stage and is remarkably robust (see Nucci 1985, who found that children from the Amish community treat moral rules as independent even from God's authority). In combination with the evidence about the overly formal, non-harm- and non-welfare-based justifications psychopaths offer for their judgments and, in general, their shallow and undifferentiated emotional life (Blair et al. 2005), the evidence suggests an essential link between moral judgments and affective capacities. Moreover, recent research has shown that acquired sociopaths and patients whose social-moral emotions are impaired due to damage of the ventromedial prefrontal cortex (VMPFC) are less likely to respond to highly emotionally engaging moral dilemmas—such as the "smothering the baby"-case—in the same manner as normal subjects do (Koenigs et al. 2007; Kennett and Fine 2008: 173ff.). But it would be premature to conclude, as Jonathan Haidt does, that the "very existence of the psychopath illustrates Hume's statement that it is "'tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger"" (Haidt 2001: 824).

Adina Roskies (2003: 60ff.) has argued, in a slightly different context, that it is both empirically and conceptually implausible to assume that "acquired sociopaths" are unable to make genuine moral judgments. She argues that there is no evidence that, as their judgments remain adequate, the moral knowledge of patients with focal damage to the VMPFC is impaired as a result of their condition, that it would be implausible to argue that the content of their judgments has changed or that their injury has turned them into detached observers which merely report other people's moral beliefs in an "inverted comma" sense. All this seems to cast the necessity-thesis into doubt.

Let me note that Roskies' arguments do not undermine, but actually strengthen the rationalist's dialectical position. If she is correct, then emotions might not even be necessary for moral judgment, and the challenge the necessity-thesis seems to pose for the rationalist disappears. It might be that psychopaths and VMPFC patients (she only talks about the latter) do make genuine moral judgments, but simply fail to act accordingly (Cima et al. 2010). I doubt, however, that her argument can achieve this. Firstly, it only applies to cases of adult-onset brain damage. Emotions might still be *developmentally* necessary for moral judgment. Secondly, it remains true that psychopaths do not fully grasp the moral/conventional distinction. Her argument does not undermine that point. Given that the distinction is one essential element of moral judgment, psychopaths fail to understand one essential element of moral judgment, and hence fail to make full-blown moral judgments.

One could argue that psychopaths make *abnormal* moral judgments. Abnormal moral judgments, however, are still moral judgments. Obviously, there is no clear cut-off point at which an abnormal practice turns into a completely different one. But if we take a look at the different morally important capacities psychopathic individuals lack, including

- morally inappropriate behavior,
- poor or absent moral emotions,
- failure to grasp the inferential structure of normative judgments (see Kennett and Fine (2008) for examples of so-called "retractor statements"),
- failure to grasp the moral/conventional distinction,

it is most plausible to deny them moral competence altogether. When it comes to deciding whether a judgment really is a genuinely moral one, one cannot look at singular instances but has to take into consideration a subject's overall status as a diachronic moral agent (Gerrans and Kennett 2010; Damm 2010). Moral judgments are those judgments that are made by morally competent subjects. Psychopaths and (some) acquired sociopaths do not pass this test. The phenomenon of psychopathy does suggest that emotional responsiveness is necessary for moral judgment (but that may still be only half the picture).

3 The Sufficiency-Thesis: Morality and Disgust

A strong form of emotionism has to argue for the necessity *and sufficiency* of emotional (dis)approval for moral judgment. How can the sufficiency-thesis be put to a test? What has to be shown here is that people's emotional reactions *alone* are sufficient to explain their moral attitudes. A change in their emotional make-up—that's the hypothesis—will result in a change of their moral judgments, either in their content or, at least, in their severity. Subject's moral judgments will vary against emotional changes much more than they will vary against moral reasoning. Call this the sufficiency-thesis.

There is a huge body of evidence in support of this claim. Valdesolo and DeSteno (2006) found that contextual variations that induce emotional changes can significantly influence moral judgments. In the footbridge-dilemma, subjects usually display a deontological hesitation to sacrifice a person's life in order to save five people from certain death (or, more precisely, to use that sacrifice as a means to the end of saving the five). But people are much more likely to judge it permissible to throw the fat man off the footbridge and prevent the trolley from killing the five people on the track after watching an episode of *Saturday Night Live* that cheered them up a little beforehand. Artificially induced mood changes seem to alter people's moral judgments about the permissibility of killing a person in a dilemmatic case. However, this result does not come as a surprise. We all know from experience that a slight change in one's mood can have a not at all slight effect on one's behavior.

But not only complex emotions like amusement have this kind of impact on our evaluations. The sufficiency-thesis can be defended on an evolutionarily more fundamental level as well. Schnall et al. (2008) found that an unpleasant odor—created, for instance, by the use of "fart spray" in subject's surroundings—can have a significant influence on people's moral judgments about marriage, sex, environmental issues, media or just about any other imaginable ethical topic. They observed the same effect when they put people behind a filthy desk and asked them to judge the permissibility of public policies and the like. In general, people tend to interpret their own bodily changes—from throat clenching to nausea—as a source of information about an issue at hand. They trust their emotional responses to a large extent, and take them to provide cues about the moral status of a described action or event.

But the most striking support for the sufficiency-thesis has been found using the method of post-hypnotic suggestion. Take the following experiment that was conducted with a group of highly hypnotizable test subjects. In a 2005 study, Wheatley and Haidt tested the effects of hypnosis-induced disgust on people's moral judgments. They confronted their subjects with several story vignettes (a congressman taking bribes, cousins performing incest, a man eating his dead pet dog) and asked subjects to assess them morally (Wheatley and Haidt 2005). Through posthypnotic suggestion, subjects had been primed to experience a quick flash of disgust upon hearing an arbitrary trigger word that occurred in the story (such as 'often'). What Wheatley and Haidt found was that in the disgust condition, subjects not only judged morally wrong actions (like a politician taking bribes from lobbyists) to be morally worse than they did in the neutral condition. They judged acts that were not at all wrong (and were judged accordingly in the neutral, no-disgust condition) to be morally blameworthy, too. When subjects were asked what exactly is wrong about, for instance, a student council representative trying to organize events that are interesting for both students and teachers, they started to confabulate, providing far-fetched ad hoc reasons ("It just seems like he's up to something") that bore no connection to the given information whatsoever. Emotions and feelings alone are, just as the sufficiency-thesis claims, enough to account for changes in moral judgment. Subjects' emotional reactions, ranging from changes in their mood and "natural" feelings of disgust in response to a filthy desk to completely extraneous disgust in response to a random trigger word are sufficient to explain people's moral judgments. Some researchers have claimed that feelings of disgust are the very essence of morality: electromyographical evidence about similarities in facial motor activity suggests that moral emotions originate in primitive, but highly adaptive response patterns to contamination and disease (Chapman et al. 2009) and cross-cultural research supports the idea that these mechanisms have then expanded from food-related to socio-moral matters in general (Haidt et al. 1997).

4 Perceptual Characteristics of Emotions

Emotionists draw on the observation that beyond the realm of abstract rights and duties, agents need to be emotionally involved in order to really grasp the complex infrastructure of everyday morality. In many cases, rational insight does not suffice, especially when it comes to motivation: there is something dubious and even callous about a person who visits his friend out of a sense of duty, and not because of his feelings of friendship. (Accordingly, emotionists often endorse a version of the "one thought too many"-argument; Williams 1981). Moral psychologists and experimental philosophers drive the claim that emotions are necessary for many important aspects of morality even further, arguing that the very capacity to make moral judgments and act on them fundamentally depends on a person's emotions. I have called this the necessity-thesis and discussed the evidence from psychopathy and acquired sociopathy that is cited to support it. Psychopaths and acquired sociopaths suffer from severe emotional impairments, which makes them insensitive to the practical "oomph" (Joyce 2007) that is typically connected to moral imperatives. However—does the rationalist have to worry about the claim that emotional reactions are necessary for—or lack thereof damaging to—moral judgment? Let's put things into perspective.

Is the necessity-thesis plausible? On a certain reading, the answer is "yes". Emotions complement moral reasoning and contribute to the phenomenal richness of moral experience; but that doesn't, as emotionists about morality argue, necessarily downgrade moral reasoning to an ineffective epiphenomenon. The right way to conceive of the significance of emotions for morality, I suggest, is along the lines of Kant's famous quote:

moral thoughts without emotional content are empty, moral emotions without reasoning are blind. Just like perceptual content is causally necessary to provide our judgments with objective material and prevent them from a "frictionless spinning in a void" (John McDowell), emotional reactions are causally necessary to provide our moral judgments with normative content. And just like perceptual appearances—think about the Müller-Lyer-Illusion—give defeasible reasons for belief, emotional responses give defeasible reasons for normative judgments. In the light of reasons to distrust our emotions, however, we have to rely on cognitively more elaborate principles of practical reasoning.

Rationalists should not succumb to the temptation to argue for "pure" reason in the realm of morality anyway. Rather, they should be happy to accept the emotional *impurity* of moral judgment and endorse it as an integral part of the moral life of human beings. Otherwise, rationalists will have to face what might be called the "pure reason-objection", an objection that has been raised by Michael Slote, Jesse Prinz and others: "Sure, if rationalism is true, we don't need the sentiments; we can rely on our rational cognition" (Slote 2004: 13). But can we even conceive of morality as having a "purely cognitive source" (Prinz 2006: 33)? And if morality is supposed to make efficacious demands on us—motivate us to act—can it be based on nothing but rational cognition? Strong rationalists about morality and externalists about motivation hold that it can. Although I cannot provide a conclusive argument for it here, I doubt it. At any rate, I shall try to meet the emotionist about morality half way and pursue the more parsimonious strategy here. We can grant that emotions do play a central role for moral judgment and behavior.

One can, I have suggested, think about the emotional underpinnings of moral judgments in a way that is analogous to the relation between perceptual content and judgments about the empirical world (Goldie 2004a, b, 2007). Advocates of genuine moral reasoning can adopt the view that emotional reactions are necessary to provide moral judgments with content, content that is related to us as feeling and acting human beings. In acknowledging the significance of emotional reactions for our moral thinking, we can even make sense of the fact that our emotional experiences seem to be beyond our rational control: in having emotions, we are passive. From that, however, it doesn't follow that emotions bear no rational connection to the web of our moral beliefs. Compare the case of perception again. Perception has an irreducibly passive element, too. But this only demarcates the line between objective thinking, which stands in front of the tribunal of experience, and fiction, which does not. Passivity is not a threat to the autonomy of a thinking subject, but a necessary condition for objectivity: the fact that our thoughts depend upon the way the world is, a way that is not at our disposal and independent of our arbitrariness. I take emotions and perceptions to share several characteristic features:

(i) Emotions have an *objective phenomenology*: they are experienced as being representations of something outside of them. This holds for simple feelings as well as more complex emotions. A pain in the funny bone is experienced as something that refers to a condition of the funny bone, not merely the experiencing subject. Guilt or remorse, on the other hand, are experienced as representing something blameworthy in the action deserving of guilt or remorse. Emotions are not experienced as combinations of facts and an emotional resonance in a feeling observer that is entirely disconnected from the events the emotion is directed towards. Bear in mind that this is a phenomenological, not a metaphysical point. An objective phenomenology need not reflect a corresponding relation between emotions and objective moral facts.
(ii) In having emotions, we are *passive*. It is often not up to us which emotions to have. To be subject to an emotion is something that can merely happen to a person. It typically does not require cognitive effort (like imagination or concentration) to have

them, persons don't have to find out inferentially whether they are in a state of emotional arousal or not and, usually, what kind of feeling or emotion they are subject to at a particular moment (pain, grief and so on).

(iii) Although emotions are often beyond our rational control, they can still be *rationally amenable*. This feature might also be called the reason-responsiveness of emotions. They are the proper object of evaluation and critical reflection, and we can—and often do—ask whether having a particular emotional response to a particular emotionally significant event makes sense. This is sometimes called the "Rational Assessment-Problem" (Prinz 2007: 60). The rational amenability of emotion does not entail, however, that emotions themselves are inferentially structured. Nevertheless, they stand in the space of reasons: it makes sense to give and ask for reasons for emotions.

(iv) Emotions can have an *informative nature*. In particular, they can "inform" a person about new values in unexpected ways, very much in the same way as perception can trigger new and unexpected beliefs: "Emotions are like perceptions in that they can arise independently of our considered convictions about the circumstances eliciting them, and may even conflict with those convictions." And, "when they persist despite those opinions they induce us to question the opinions. So even if I think beauty is only skin deep, and being smart or interesting or funny is what's important, I can be brought to think that one's appearance matters more than I previously acknowledged by finding myself ashamed of my flabby stomach at the beach" (D'Arms 2005: 9). The informative nature of emotional experience has also been borne out empirically, in terms of the so-called "affect as information"-paradigm (Schwartz and Clore 1983).

(v) Emotions are, just like perceptions, phenomenally *finegrained*. As feelings, they have a level of subjective detail that cannot be fully exhausted by conceptual means alone (Gunther 2003).

(vi) Just like perceptions, emotions transcend themselves towards an *intentional object*. People feel guilty *about something*, they resent others *for something*, and they are afraid *that something might happen*. (This special kind of intentionality has been described as a "feeling towards", cf. Goldie 2000; Döring 2007.)

(vii) Emotions and perceptions essentially involve a perspective that other mental states such as beliefs, for example, lack (de Sousa 1987: 149ff.; Deonna 2006). My belief that the table is round and your belief that the table is round are the same belief with the same propositional content. But my perception of the table and your perception of the table, as well as my and your experience of the concert last night, are irreducibly different due to their difference in perspective.

On the face of it, the oscillation of emotions between their passivity and rational amenability seems difficult to explain. That may be, but it only makes the emotion/perception analogy more plausible. In fact, this ambivalent relation to the space of reasons seems to be one of the key issues in the epistemology of perception (McDowell 1994) today. Peter Goldie (2007: 3) has suggested that in order to reconcile the elements of causal passivity and spontaneous activity of emotions, one can draw a distinction between two kinds of inferentiality: "[…] a belief can be both non-inferential in the phenomenological sense (no conscious reasoning by the subject) and yet inferential in the epistemic sense (justifiable by the subject)". Emotional experiences typically occur non-inferentially, but are susceptible to reasoning after they have emerged.

Moreover, the emotion/perception analogy is the best way to deal with the problem of recalcitrant emotions. If emotions are, as the James-Lange-Theory has it, mere feelings (i. e. the conscious awareness of bodily changes), then it remains puzzling how there can be

experiences of emotional recalcitrance in the first place. If emotions are, as cognitivists argue, primarily judgments of value, then the phenomenon of emotional recalcitrance consists in a conflict between unconsciously and consciously held evaluative judgments (Brady 2008). This is implausible. Rather, emotions that just won't go away are best described as analogous to optical illusions (Tappolet (2011)).

But aren't emotions, one might object, unlike perceptions in that they are not directed at the external world, but at an inner realm of subjective experiences that is only accessible through introspection? No, they aren't, because emotions are essentially connected to bodily changes: they are, among other things, feelings that detect such changes. But the body is part of the world. Perceiving what happens with and inside the body is perceiving a part of the world, the only difference being that this very part of the world is *me* (or any other perceiving subject, respectively).

There is one crucial difference between emotions and normal perceptions one has to bear in mind: perceptions tell us how the world is. Emotions, as motivational states, tell us how the world—according to the person who has them—ought to be. Thus emotions cannot be said to refer to facts of any kind, nor that they are subjective experiences of evaluative facts (as McDowell (1985) seems to suggest). They don't represent the world, but an agent's position in the world, and suggest a range of possible goals to pursue or states of affairs to avoid.

Let me sum up. In the case of propositional knowledge, epistemic agents struggle for knowledge about the world. Their immediate raw material is sensory perception. In the case of practical reason, moral agents try to figure out what to do, and their immediate raw material is emotion. A good reason for a belief about the empirical world is one that is based on sensory perceptions everybody could share; analogously, a good reason for action—or, third-personally, judgment about action—is one that is based on motivational states (desires, emotions or sentiments) everybody could share. On that view, emotions fit nicely into a normative perspective on moral judgment. Perceptions present the world to be in a certain way, a way that is subject to error and illusion. They nevertheless present a way the world appears to be—they are "is-appearances", as it were. Emotions figure in moral judgments in the same way as perceptions figure in perceptual judgments: they are, as it were, "ought-appearances" and defeasibly present a way the world ought to be according to someone's individual perspective. Consider the following, final illustration: if you witness an old lady being mugged by a young man you will respond to that event with distress, a feeling that is connected to a disposition to experience empathy upon the harm being done to other people. In this case, your empathy alerts you as the witness that something fishy is going on, something that calls for further attention or action. If it wasn't for your emotional reaction, you might well have overlooked the incident altogether. Our emotional reactions help us to stay *open* to new moral experiences and present to us the finegrainedness of the moral world; they are the material that forces us to recalibrate our moral convictions, the driving energy in a perpetual process of moral change and a force of stimulation, rather than deformation, for moral thinking (Arpaly 2003).[1]

[1] The most obvious and straigthforward way for the rationalist about moral judgment to incorporate the necessity-thesis seems to be to argue for a cognitivist position in the theory of emotion. If emotions just are cognitive states (evaluative judgments), then the fact that emotions are necessary for moral judgment does not have to intimidate the rationalist the slightest bit. For an overview over this position, see Deigh 1994. For a defense of cognitivism about emotions, see Solomon 1976; de Sousa 1987; Greenspan 1988; Helm 2001; Stocker and Hegeman 1996; Nussbaum 2001. I do not argue for emotional cognitivism in this paper, because I want to avoid the problems the cognitivist has with offering a psychologically realistic picture of the emotions without overintellectualizing them (a point that Peter Goldie (2000) insists on).

5 Justificatory Sufficiency

What about the sufficiency-thesis? Is it plausible as well? On a certain reading, the answer is "yes", too. But what exactly is it that the empirical evidence shows in support of this thesis? Take the experiments in which a filthy desk, fart spray or hypnotically induced disgust upon hearing an arbitrary word trigger an emotional response that prompts subjects to make a certain moral judgment. In these cases, subject's emotional reactions are in fact sufficient for their moral judgment. But they are sufficient in a very specific sense: they are sufficient to explain people's judgments, that is, they are *causally* sufficient. This is not, however, the kind of sufficiency one would expect from an analysis of the concept of moral judgment. Moreover, it is an uncontested fact, a truism. As Karen Jones puts it: "The folk already know that emotions influence moral judgement. It is part of the commonsense of moral epistemology that we must be on the lookout for the potentially distorting influence of emotions on moral judgements" (Jones 2006: 47). By naming the *causally* sufficient conditions for something to alter somebody's judgment, one misses the point of a conceptual decomposition of the notion.

The causal story that is told by empirical moral psychology doesn't rule out cases in which the judgments that are caused by changes in a subject's emotional make-up do not count as proper moral judgment (sometimes for that particular reason). How do we know that what people are doing really is to make a moral judgment? Consider the following example by Karen Jones: "Moral judgements are distinguished from judgements of mere liking or disliking by being answerable to reasons. We challenge, accept, and reject moral judgements on the basis of reasons. [...] If someone were to answer a challenge to their judgement that an act was morally right by citing the fact that it was done on a Tuesday, our first response would be bafflement and, if further questioning did not bring the cited consideration somewhere closer to the cluster of considerations we recognize as morally relevant, [...] we would conclude that the person lacked competence with moral concepts" (Jones 2006: 49). Accordingly, if a subject were to answer a challenge to her moral belief that an act was wrong saying that upon reading about it, she perceived an unpleasant scent from an unknown source, we would be surprised, to say the least, and reject this answer as illegitimate. Unlike mere expressions of disgust, moral judgments are held to certain standards of relevance and rationality. Irrelevant situational features such as scents or ambient noise just do not bear on the moral wrongness of an action, even on the most straightforward sentimentalist account. "Moral" judgments that are based on influences or considerations extraneous to the moral status of an action are deficient to a certain extent. Emotionists about morality have to account for this fact in a way that rules out such cases.

I have called the position that maintains that emotional reactions are essential for moral judgment "emotionism" about morality. At this point, I shall introduce a distinction between two different kinds of emotionism, which I will refer to as *simple* and *normative* emotionism. The account of moral judgment given by empirically supported emotionism about morality is best described as

Simple Emotionism
Emotions are causally necessary and sufficient for moral judgment.

Obviously, one has to understand the necessity and sufficiency of emotional reactions for moral judgment *dispositionally*. Any workable account of emotionism about moral judgment must allow for cases in which the associated emotional response happens to be absent. A subject can sincerely judge an action to be morally wrong even though at a particular moment, she doesn't feel any kind of resentment our outrage upon the deed. It

suffices if the appropriate response used to be associated with the judgment in the past, creating a disposition to have the feeling at any given moment in the future, a disposition that might not be realized in any instance.[2] If we also add the fact that typically, the objects of moral judgments are human actions, and that typically, moral emotions can be characterized as states of approval or disapproval, we get:

Simple Emotionism'
For a person to make a moral judgment, it is causally necessary and sufficient to have a disposition to feel an emotional reaction of (dis)approval towards a particular action.[3]

As we have seen, this account is not equipped to rule out deviant cases of manipulation, hypnosis or, in general, the influence of irrelevant factors on people's moral judgments, which, as we have also seen, undermines the moral competence of the people who are under the influence of these factors; it doesn't rule out cases where a person's emotional reaction is sufficient in a merely causal, but morally irrelevant and hence not justification-conveying way. What is needed is a condition that is sufficient for moral judgment in a way that doesn't threaten a moral judgment's standing in the space of reasons and, at the same time, a judge's moral competence as a whole. What has to be captured is the inevitable justificatory connection that a person's emotional reactions bear to her moral attitudes.

6 Reflective Endorsement and Reliability

When it comes to finding the essential components of this *justificatory sufficiency*, as it might be called, we can draw on the idea that, among other things, a proper standing in the space of reasons can be acquired by the following criterion: if a subject doesn't withdraw his judgment after a post-experimental debriefing about how he came to make the judgment he made (say through post-hypnotic suggestion), his judgment will not count as a genuinely moral one, provided that the debriefing revealed to him that he did not respond to the morally relevant features of the action in question. To put it in a more abstract fashion: genuine moral competence is characterized by the feature that if after a moral judgment has been made, a certain piece of information is added in the light of which the initial judgment is undermined, the judging subject will suspend his belief or adequately back it up with appropriate further grounding.

In order to cash out this idea in more detail, one can use the notion of counterfactual, idealized conditions. A subject is to be counted as morally competent if, were a certain undermining piece of knowledge added to her set of beliefs, she would reconsider her moral judgment and refrain from it when necessary. Were a subject to be informed that in making her moral judgment, she responded to empty pizza-boxes on her table and a subtle smell of

[2] In what follows, I shall argue that emotions are sufficient for moral judgments only if they cause them in a justification-conveying way. On a dispositional view of the emotions, there seems to be a problem with *unmanifested* dispositions. If there can be such dispositions, and it seems that there can, one might ask how these can convey any justificatory force to the judgments which are based on them. In the theory of moral judgment, the motivation to use a dispositional concept of emotions is to allow for cases in which the respective emotion is not occurrent. The content of the judgment has been committed to memory. This is impossible, however, if the disposition to have a certain emotional reaction that judgment is based on has *never* been manifested. Thus, the problem of unmanifested dispositions will, though metaphysically possible, typically not be a problem in the case of *emotional* dispositions.

[3] This, in a nutshell, is Prinz' "constructive sentimentalism", cf. Prinz 2007.

rotten eggs, a competent moral judge would reconsider her judgment and either give it up or cite appropriate moral reasons that "repair", as it were, the initial status of the judgment. This is tantamount to saying that were that subject to possess full knowledge and flawless reasoning abilities already, she wouldn't judge the way she did in the first place. What counts for a moral judgment to be a genuine one is the way a moral subject reacts after being exposed to perfect information and equipped with perfect reasoning. Does the subject stick to her judgment and the way she arrived at it, and endorse both upon reflection? Or does she withdraw it, and reconsider her initial belief?

Note that on this view, there is no stark opposition between reasons and causes for judgments. The distinction that is used here is one between malign (undermining) and benign (justification-conveying) varieties of causation. My proposal is to understand justificatory sufficiency in terms of benign causal sufficiency. An emotional reaction is justificatory sufficient just in case it is causally sufficient in a way that can be reflectively endorsed under conditions of full information and rationality. Think about perception again. The acquisition of perceptual knowledge about the world is a causal process to a large extent. Nevertheless, we do not hesitate to describe this process as conveying the justificatory force that is necessary to render a perceptual belief that, say, the sun is shining, into knowledge.

If we add these qualifications to the emotionist account, we get:

Normative Emotionism
It is necessary and sufficient for making a moral judgment to have a disposition 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.

Remarkably, some philosophers who favor a strong version of emotionism subscribe to a set of qualifications similar to the one just presented. Jesse Prinz, for instance, cites the evidence from psychopathy and acquired sociopathy as well as filthy desk-style examples in favor of both the necessity- and the sufficiency-thesis. But he implicitly rejects it at the same time. After his discussion of the empirical findings about moral judgment, he writes: "The first problem [of one form of sentimentalism] has to do with error. If 'wrong' referred to whatever causes disapprobation in me, then I could not judge something to be wrong in error. To avoid this consequence, we must idealize. We should say that the word 'wrong' refers only to those things that irk me under conditions of full factual knowledge and reflection, and freedom from emotional biases that I myself would deem as unrelated to the matter at hand" (Prinz 2006: 35). Clearly, any reasonable agent has to deem picking up on an arbitrary trigger word as unrelated to a moral issue. The trigger word is arbitrary, after all.

At the center of the above account lies an idea adopted from Jeannette Kennett, the idea that "genuine moral judgments are those that are regulated or endorsed by reflection" (Kennett 2009: 78). What the empirical evidence shows, however, is that cases where a moral judgment is *regulated* by reflection from the outset are rare. It seems to be more promising to concentrate on counterfactual reflective endorsement. Competent moral judges are allowed to arrive at their judgments not by a causally effective conscious process of reflection, but by an emotionally triggered intuitive process that is reflectively acceptable. By and large, we can say that competent moral subjects reconsider their initial judgments after being informed about their causal genesis. If a judgment survives this scrutiny, one can say that the subject *reflectively endorses* it.

There are two possible objects of endorsement here: the moral judgment itself, and the method by which the subject arrived at it. My main focus lies on the latter, because I am not

talking about the conditions under which moral judgments are true (or otherwise correct), but the conditions under which judgments are moral judgments. If, by accident, a subject arrives at a correct judgment using a fluky method, she can still endorse the judgment and stick to it. But she will have to deem the way she arrived at it inappropriate because it was fluky.

It remains an open question under which circumstances a subject is rationally entitled to this endorsement, but we already have a suggestion on the table. What I want to argue for is the following simple but powerful idea: a subject can reflectively endorse her judgment if the method she used to arrive at it is reliable. Reliability is specified in counterfactual terms. A method is reliable with respect to X if and only if, under slightly different circumstances (or, perhaps more technical, in close possible worlds) where X obtains, the method would lead a subject to believe X. Conversely, under slightly different circumstances where X doesn't obtain, it would lead the subject to acknowledge that fact just as much. Reliability is a normative concept. It presupposes certain standards of correctness: reliability just means "likelihood to produce *correct* results". What else could it be that justifies my accepting the way my moral belief was brought about? Since as a sincerely morally judging subject, I am interested in the correctness of my judgments, I can endorse the way I arrive at them if it primarily serves that particular purpose.[4]

Two further important questions have to be addressed. First: what makes for a method's being reliable? How do we establish this property? And second: does the above discussion show that the empirical evidence merely supports *Simple Emotionism'*, but is ruled out by *Normative Emotionism*? Does it show that the way subjects arrive at their moral judgments in filthy desk-style scenarios really is unreliable?

As for the first question, I would like to return to an example from above. If I witness an old lady being mugged by a young man, I am observing a chain of events that automatically elicits a reaction of empathic distress. Provided that I do not come across additional information about the incident that mitigates my reaction (if it turns out, for instance, that the old lady actually is a young man, disguised as an old lady, and that he stole the other young man's purse right before), I will take my emotional reaction—as the evidence for the necessity-thesis predicts—as a defeasible "ought not-appearance" and judge that what the young man did was wrong. Can I reflectively endorse the method with which I arrived at my verdict? According to the above proposal, I am entitled to do so if it is reliable, that is, likely to produce correct moral judgments even under slightly different circumstances. Given that under normal circumstances, empathic distress that is caused by the observation of a *prima facie* harmful incident is likely to "pick up" on features of the situation that are morally relevant, namely, that a harmful action has been done, it indeed is. It responds to the fact that a presumably innocent person is being stolen from without good reason, and the disposition to respond to such acts in that way reliably, though not infallibly, leads to a correct normative attitude.

[4] The concept of reliability employed here may strike some as odd, because reliability doesn't seem to be about getting it right in a range of circumstances, but about getting it right on most occasions. In this sense, my use of the concept of reliability is stipulative to a certain extent. Roughly, what I have in mind is that a method of judgment-formation is reliable if it satisfies a *sensitivity*-condition (a judgment that p is sensitive iff $\sim p \rightarrow \sim B_x(p)$ and $p \rightarrow B_x(p)$, cf. Nozick 1981, 176). Both conditions are important because, given the empirical evidence, disgust reponses that pick up on extraneous features are ruled out by the first conjunct, cases of artifical mood induction that prevents people from picking up on morally relevant features are ruled out by the second conjunct. For an overview over the concept of reliability in general epistemology, see Pritchard 2005.

As for the second question, the task is to see whether *Normative Emotionism* can successfully cope with cases that, intuitively, don't seem to meet the constraints for proper moral judgment. Subjects in the disgust-condition of an experiment (for instance, people who are placed behind a filthy desk) predictably judge morally wrong actions more severely, and people under the influence of hypnosis might even judge actions wrong that aren't morally wrong at all. But what these people respond to in experiencing an emotional reaction—like the word 'often' in a story vignette—is not a morally significant feature of the situation at hand. Imagine a person that correctly judges a morally wrong act to be wrong, but does so by picking up on a trigger word used in the description of the story. Clearly, that person would have judged a different action to be wrong as well, even though it might not have been, as long as its description had contained the trigger word. The method that person has used to arrive at her judgment does not reliably "track" moral wrongness. A person that reacts outraged upon reading an article about female mutilation does. The proposed criterion can cope with deviant cases.

Take the following case, presented by Gilbert Harman (2007): you witness a group of children setting a cat on fire for fun. They pour gasoline on it, and ignite it. Now suppose you are participating in a psychological experiment and a story vignette is presented to you that contains a description of the incident. It might read something like this:

> *Scenario I*
> You walk around the corner and see a group of young people. You witness one of them catch a cat, another one hold it to the ground, another one pour gasoline all over it and another one pull out a matchbox and set it on fire. They observe the cat while it tries to escape its misery, occasionally, they laugh about the animal's screams and its futile attempts to survive, and they stay until the charred, dead body stops burning.

I expect most people to be horrified even by imagining that scene, chilled by the youngsters' callousness and outraged upon their wanton cruelty. The judgment that what these young people did was wrong literally forces itself upon us. Now suppose that, via post-hypnotic suggestion, you have been primed to feel a pleasant warm feeling upon reading the arbitrary trigger word 'often'. You are now presented with a slightly different version of the story (*Scenario II*) where the trigger word has been inserted into the last sentence. Due to the fact that your feelings have been manipulated through hypnosis, the judgment that what they did wasn't that bad at all forces itself upon you instead. You don't know why, but what these youngsters did suddenly seems quite kind to you. Unbeknownst to you, you have not responded to the morally relevant features of the situation, but to a trigger word that you have been primed to pick up on. But clearly, using that method proved to be unreliable. Under slightly different circumstances—only the word 'often' has been added to the story, after all—where the judged action is still wrong, or at least no different in all important respects, you diverted from your initial judgment and changed your mind. But the described action *is* wrong. It is the unreliable method you used in arriving at your judgment that made you think otherwise. A competent moral judge would thus dismiss the method upon reflection under ideal conditions in which he has been informed about the causal genesis of his judgment.

Whether a subject has arrived at her judgment using a method she can reflectively endorse and whether, although the subject didn't use a reliable method, she is prepared to reconsider her judgment depending on how it fares under reflective scrutiny, are two different things, of course. Due to its "openness" to rational reflection, however, the second case can be classified as proper moral judgment as well. This, I suggest, can be accounted for disjunctively:

Normative Emotionism'
It is necessary and sufficient for making a moral judgment to have a disposition to an emotional attitude of (dis)approbation towards certain actions that either

i) causes the judgment in a way that can be reflectively endorsed by the judging person under ideal conditions of full information and rationality or
ii) allows the subject to withdraw her judgment in the light of undermining evidence or back it up with appropriate further grounding.

It makes sense to classify judgments that fall under the second disjunct as genuine moral judgments, too, because although in ii), a subject does not respond to a morally significant feature of the situation she judges about, she nevertheless possesses the cognitive virtues that are needed for genuine moral competence which, in a way, is another reliable method of forming moral beliefs.

One can also make the above point the other way round. Suppose the test subjects are psychopaths, who typically score low on susceptibility to empathic distress. You confront them with *Scenario I*. They react pretty untouched and judge the case accordingly. As we have seen, they lack the necessary "perceptual" capacities to pick up on the morally relevant features of the issue at hand. Psychopaths are, however, not deprived of all disposition to feeling. So you prime a second group of psychopaths to respond with disgust to the trigger word in *Scenario II*, and the subjects in that group judge the children's violent acts to be wrong. Even though in the latter case, the emotional response did in fact lead to a correct moral judgment, the method that was used cannot be reflectively endorsed under conditions of full information and rationality. Responses to extraneous features of a situation do not guarantee that one's judgments track the available moral reasons.

On the other hand, a morally competent judge will respond to relevant features of the situation in question, and his judgments will not be subject to manipulation by random influences. Suppose a normal, healthy adult individual is presented with a third story that frames roughly the same chain of events as in *Scenario I* in a completely different way. We can expect a normal person—a person who thinks what these kids did was terribly wrong—to be insensitive to the changes in wording between the first and the third scenario when making her judgment, because these are features she did not—and should not—respond to in making her judgment. Rather, her judgment was caused by her empathic distress, her feelings for the suffering animal. This is how she arrived at her judgment, and due to the fact that this "method" is likely to lead to correct judgments, it conveys the justificatory force necessary to render a judgment a moral judgment, and a moral judge competent.[5]

7 Reliability, Normativity, and Moral Objectivism

Moral judgments can be based on emotions without this being a threat to a suitably moderate and empirically feasible rationalism about the psychology of moral judgment. The

[5] Neo-Sentimentalist accounts of moral judgment analyze moral judgments similarly, namely in terms of conditions of appropriateness for emotional reactions: on that view, to make a moral judgment is to think it appropriate to have an emotional reponse (of guilt, resentment etc.) towards an action, person, or event (see, for example, Wiggins 1998; Gibbard 1990; McDowell 1998). I have argued elsewhere (Sauer 2011) that this account is not successful, particularly because it does not offer a satisfactory response to the so-called "conflation problem" (D'Arms and Jacobson 2000; Rabinowicz and Rønnow-Rasmussen 2004; Olson 2004).

emotion/perception analogy shows how rationalism can account for the necessity-thesis, and the normative constraint on emotional processes outlined above smoothly incorporates the sufficiency-thesis.

Both building blocks of my argument, however, are subject to the same kind of tension. On the one hand, the account tries to avoid commitment to a strong from of moral objectivism, and locates moral properties in (rationally amenable) emotional responses. On the other hand, the account relies heavily on the normative distinction between emotions that do convey justificatory force to the judgments they cause and emotions that do not. Prima facie, it is hard to see how the normative goals of the proposal and its rejection of the idea that there can be an emotion-independent standard of reliability can be reconciled. It seems that the account falls prey to a substantial inconsistency: how can emotions be both judge and party in the moral court (de Sousa 2001)?

In order to see that this is not so, one needs to take a closer look at how the concept of reliability figures in *Normative Emotionism'*. I have said that an emotional process is sufficient for justification just in case it can be endorsed upon (counterfactual) reflection, and that a subject is rationally entitled to do so if the process is reliable. Reliability is a normative concept, and the standards for assessing whether a particular emotional reaction is reliable seem to presuppose a strong form of moral objectivism: an emotion is reliable if it tracks the moral truth. But the account is supposed to meet the emotionist about morality half way, and work without there being any objective moral truths, or any response-independent normative facts that make moral judgments true, and thus the notion of reliability seems to be a forbidden fruit. What counts as a correct moral judgment, I suggested, depends on reflective endorsement as well. But this seems to be circular: an emotion can be endorsed if it is reliable, and the standards of reliability are determined by what can be reflectively endorsed.

The analogy between emotion and perception suffers from a different but related problem. Perceptions convey justificatory force on the beliefs they cause because the facts there are (say, that the sun is shining) cause true beliefs about those facts. For moral emotions to be a potential source of reasons for the moral judgments which are based on them, something similar has to happen; otherwise, an emotion does not justify a judgment any more than any other causal, judgment-eliciting process does. Here, too, the concept of reliability is of crucial importance.

Both the problem of circularity and the problem of normativity disappear once we see that the concept of reliability the account makes use of is twofold. I shall, for lack of a better term, refer to those two types of reliability as reliability$_1$ and reliability$_2$. Let me first explain how, with this distinction in hand, the problem of circularity disappears. Roughly, the idea is that an emotional process is reliable$_1$ iff it can be reflectively endorsed under ideal conditions, and it can be reflectively endorsed iff it is reliable$_2$. Reliability$_1$ is about which emotional processes that cause the subject to make a certain moral judgment convey justification on that judgment. The distinction between processes that have this kind of normative force and those that do not is needed to capture the intuition that reactions of, say, empathy or disgust which *merely happen to be* about morally salient scenarios, but do not pick up on the morally relevant features of those scenarios, do not count as genuine moral judgments. Reliability$_2$ is about the features that ground this reflective endorsement. Suppose a subject has formed a moral judgment about an action X, but has done so using a "method" which—oblivious to her—picks up on an arbitrary feature of the situation, such as ambient noise, a bad smell or a trigger word. This method is unreliable not because it is insensitive to the objective moral truth (because it cannot be said to "track" it), but in a way that is internal to the subject: using the same method, the subject could and would have formed the opposite judgment about the case at hand if her emotional state had been manipulated differently; and she would have made the very same judgment, given the same

kind of manipulation, even if the case had been entirely different. That fact alone rules out the method as unreliable, because it is not sufficiently robust and insensitive to arbitrary features. This lack of robustness can be detected simply by comparing two different scenarios. One need not invoke an objective standard of moral truth here.

An emotional process is reliable$_1$ iff it can be reflectively endorsed: if the subject came to know about how she arrived at her judgment, would she still accept it? And an emotional process can be reflectively endorsed iff it is reliable$_2$: in arriving at her judgment, has the subject been causally influenced by features she deems relevant to the issue in question? The first type of reliability explains why my account is not committed to a strong form of moral objectivism: there are no moral facts in a straightforward sense, over and above the emotional reactions we would have if we were fully rational (cf. Smith 1994, 182ff.). The second type of reliability holds that it is in order to reflectively endorse an emotional process by which one has arrived at a moral verdict iff that process responded to the morally relevant features of the situation, action, or person the judgment is about. Why call this second type reliability? The previous section shows that picking up on morally irrelevant features cannot be endorsed because doing so can—and will, as the experiments demonstrate—lead one to form a judgment one might not be willing to accept in the light of information about its actual causal genesis, unless one can back it up with appropriate further grounding that *reconnects* one's judgment with features one *could have* responded to in arriving at one's moral belief. To say that there are morally relevant features is not to say that there are moral facts. Moral facts are mind-independent facts about what is morally right or wrong. I do not think there are such facts. What is morally right or wrong is determined not by facts but by the rationally amenable emotional reactions we have. The considerations which are relevant, in a domain-specific way, for the normative assessment of those reactions consist of ordinary facts together with considerations that refer to the morally relevant features of a situation. That what is morally relevant or not—considerations of harm or fairness, for example—is also due to the emotional capacities of human beings does not introduce an inconsistency into the proposal. It is a trivial fact without which morality would be pointless to begin with.

What those features are, and what renders a feature morally relevant or irrelevant, is a question for normative, not for metaethical inquiry. A possible suggestion, however, would be to start from intuitively compelling examples—bad smells and filthy desks are paradigm examples for features that are morally irrelevant, violations of accepted moral principles and considerations of harm are good examples for things that are morally relevant—and work one's way up from those distinctions. What counts as morally relevant and irrelevant is still under negotiation: it is both essentially contested and historically variable. Disgust, for example, can pick up on arbitrary trigger words, the pollution of a sanctuary or the description of an appalling violent crime. Whether only the third or also the second thing are morally relevant is a normative question, a question about what one ought to do. But the distinction between morally relevant and irrelevant factors alone involves no commitment to metaethical objectivism. The story told in this paper is thus not only a story about what moral judgment is, and what the psychological basis of genuine moral judgments is, but also a story about how metaethical questions naturally lead into questions of normative ethics.

8 Conclusion

Are emotions necessary and sufficient for moral judgment? I have argued that they are, but that that does not have to worry a moderate rationalist about moral judgment. Emotions are

necessary for moral judgment in the same way as perceptions are necessary for judgments about the external world, and emotions are sufficient for moral judgments only if they cause them in a normatively acceptable way.

One might worry that the argument I have presented uses a conceptual magic trick, designed to simply discard the empirical evidence by "fiat". It seems that the conceptual argument just presented offers a refutation of empirical models of moral judgment simply by saying that what is studied in the respective experiments isn't genuine moral judgment *by definition*. This is not what I want so say, and it wouldn't be a plausible objection anyhow. When people are judging about cannibalism or incest, they really are making moral judgments, even in cases where it can be shown that their judgments are nothing more than expressions of disgust. Compare the aesthetic judgment by a person who thinks Otto Dix' triptych *The War* is "ugly" because it depicts "ugly" things—shredded bodies, burning land, soldiers with gas masks. This may well be poor reasoning, but still: we have to admit that that person really is making an aesthetic judgment, although a deficient one. I suggest to describe the moral case in the same way: people whose judgment's are not responsive to relevant moral considerations, but are triggered by uncontrollable and hence unreliable emotional responses, are making deficient moral judgments. They are *trying* to make moral judgments, they *think* that they do, but if they show themselves not to stand in any kind of autonomous—that is, rational—relation to their judgments, they cannot be said to engage in the practice of moral judgment *in the right way*. Constantly failing to live up to these standards renders a subject morally incompetent, and ultimately deprives it of its status as a morally judging subject altogether (although it remains the legitimate addressee of obligations and possessor of rights, of course). A full-blown moral judge must eventually take an interest in the trackingness (as specified in the first disjunct of the above account) and rational answerability (as specified in the second disjunct) of her moral beliefs.

It remains an open question whether the account developed in this paper describes what subjects actually do or whether it specifies what agents ought to do. This is not an accident: the argument above is supposed to be psychologically realistic and empirically adequate; at the same time, however, it is also supposed to be demanding enough to make normative claims on moral agents, claims that these agents can fail to meet. It can be shown empirically that subjects do engage in normative reflection about their immediate emotional attitudes and automatic intuitions (Schwartz and Clore 1983), and that they are prepared to discount distorting influences on their judgments and correct for extraneous influences to their beliefs, once these are made accessible to them. The influence of filthy desks on people's moral judgments can be seen as a form of unwanted mental contamination (Wilson and Brekke 1994).

In fact, the above conditions specify norms of moral competence test subjects themselves hold to be adequate. The findings by Wheatley and Haidt about the connection between hypnotic disgust and the severity of moral judgment nicely illustrate that certain standards of correctness are not externally imposed on people by philosophical theory, but are actually written into the patterns of moral reasoning by ordinary subjects themselves. Remember that the research conducted by Wheatley and Haidt used the method of post-hypnotic suggestion to prompt people to experience a quick flash of disgust after hearing (or reading, respectively) an arbitrary word like 'often'. Here is how some of the subjects described their experience: "'When 'often' appeared I felt confused in my head, yet there was turmoil in my stomach. It was as if something was telling me that there was a problem with the story yet I didn't know why.' One non-amnesic participant commented. 'I knew about 'the word' but it still disgusted me anyway and affected my ratings. I would wonder why and then make up a reason to be disgusted" (Wheatley and Haidt 2005: 783). People

are confused about an emotional reaction they can find no plausible source for. They hesitate to make the judgment they are inclined to make, and rightly so. They even admit that they would "make up" a reason, implying that they know that there actually is none. Subjects are aware of the fact that their response to an irrelevant cue does not reliably indicate the moral wrongness of an action or the blameworthiness of a person.

Acknowledgements I would like to thank Tom Bates, Pauline Kleingeld and Markus Schlosser for their very helpful comments to an earlier version of this paper. I am also indebted to the organizers and participants of the workshop *Philosophical Implications of Empirically Informed Ethics* at the University of Zürich, March 2010, especially to Anne Burkard, Markus Christen, Jan Gertken, Bert Musschenga, Hichem Naar and Shaun Nichols. Two anonymous referees from *Ethical Theory and Moral Practice* have made very useful suggestions, for which I would like to thank them as well.

References

Appiah KA (2008) Experiments in ethics. Harvard UP, Cambridge
Arpaly N (2003) Unprincipled virtue. An inquiry into moral agency. Oxford University Press, New York
Blair RJR (1995) A cognitive developmental approach to morality: investigating the psychopath. Cognition 57:1–29
Blair J, Mitchell D, Blair K (2005) The psychopath. Emotion and the brain. Blackwell, Malden
Brady MS (2008) The irrationality of recalcitrant emotions. Philos Stud 145(3):413–430
Chapman HA, Kim DA, Susskind JM, Anderson AK (2009) In bad taste: evidence for the oral origins of moral disgust. Science 323:1222–1226
Cima M, Tonnaer F et al (2010) Psychopaths know right from wrong but don't care. Soc Cogn Affect Neurosci 5:59–67
Damasio A (1994) Descartes' error. Emotion, reason, and the human brain. Gossett/Putnam, New York
Damm L (2010) Emotions and moral agency. Philos Explor 13(3):275–292
D'Arms J (2005) Two arguments for sentimentalism. Philos Issues 15:1–21
D'Arms J, Jacobson D (2000) Sentiment and value. Ethics 110:722–748
Deigh J (1994) Cognitivism in the theory of emotions. Ethics 104(4):824–854
Deonna JA (2006) Emotion, perception and perspective. Dialectica 60(1):29–46
de Sousa R (1987) The rationality of emotion. MIT Press, Cambridge
de Sousa R (2001) Moral emotions. Ethical Theory Moral Pract 4(2):109–126
Döring S (2007) Seeing what to do: affective perception and rational motivation. Dialectica 61(3):363–394
Gerrans P, Kennett J (2010) Neurosentimentalism and moral agency. Mind 119(475):585–614
Gibbard A (1990) Wise choices, apt feelings. A theory of normative judgment. Oxford University Press, New York
Goldie P (2000) The emotions. A philosophical exploration. Oxford University Press, New York
Goldie P (2004a) Emotion, reason, and virtue. In: Evans D, Cruse P (eds) Emotion, evolution, and rationality. Oxford UP, Oxford, pp 249–267
Goldie P (2004b) Emotion, feeling, and knowledge of the world. In: Solomon R (ed) Thinking about feeling: contemporary philosophers on emotion. Oxford UP, Oxford
Goldie P (2007) Seeing what is the kind of thing to do. Perception and emotion in morality. Dialectica 61 (3):347–361
Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. Science 293:2105–2108
Greenspan PS (1988) Emotions and reasons: an inquiry into emotional justification. Routledge, Chapman and Hall, New York
Gunther Y (2003) In: Gunther Y (ed) Emotions and force. Essays on nonconceptual content. MIT Press, Cambridge, pp 279–288
Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108(4):814–834
Haidt J, Rozin P, McCauley C, Imada S (1997) Body, psyche, and culture: the relationship between disgust and morality. Psychol Dev Soc 9(1):107–130
Hare RD (1999) Without conscience. The disturbing world of the psychopaths among us. Guilford, New York
Harman G (2007) Ethics and observation. In: Shafer-Landau R (ed) Ethical theory. An anthology. Blackwell, Malden, pp 36–41

Helm BW (2001) Emotional reason. Deliberation, motivation and the nature of value. Cambridge University Press, Cambridge
Jacobson D (2008) Does social intuitionism flatter morality or challenge it? In: Sinnott-Armstrong W (ed) Moral psychology vol. 2: the cognitive science of morality: intuition and diversity. MIT Press, Cambridge, pp 219–233
Jones K (2006) Metaethics and emotions research. A response to Prinz. Philos Explor 9(1):45–53
Joyce R (2007) The evolution of morality. MIT Press, Cambridge
Kennett J (2009) Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology. Ethical Theory Moral Pract 12:77–96
Kennett J, Fine C (2008) Internalism and the evidence from psychopaths and "acquired sociopaths". In: Sinnott-Armstrong W (ed) Moral psychology vol. 3: the neuroscience of morality: emotion, brain disorders, and development. MIT Press, Cambridge, pp 173–191
Knobe J, Nichols S (2008) Experimental philosophy. Oxford UP, New York
Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgments. Nature 446:908–911
McDowell J (1985) Values and secondary qualities. In: Honderich T (ed) Morality and objectivity. Routledge, London, pp 110–129
McDowell J (1994) Mind and world. Harvard University Press, Cambridge
McDowell J (1998) Projection and truth in ethics. In: McDowell J (ed) Mind, value and reality. Harvard University Press, Cambridge, pp 151–167
Nichols S (2004) Sentimental rules. On the natural foundations of moral judgment. Oxford UP, New York
Nozick R (1981) Philosophical explanations. Belknap, Cambridge
Nucci L (1985) Children's conceptions of morality, social conventions and religious prescription. In: Harding C (ed) Moral dilemmas: philosophical and psychological reconsiderations of moral reasoning. Precedent, Chicago, pp 137–174
Nussbaum M (2001) Upheavals of thought. The intelligence of emotions. Cambridge University Press, New York
Olson J (2004) Buck-passing and the wrong kind of reasons. Philos Q 54:295–300
Prinz J (2006) The emotional basis of moral judgment. Philos Explor 9(1):29–43
Prinz J (2007) The emotional construction of morals. Oxford UP, New York
Pritchard D (2005) Epistemic luck. Oxford University Press, New York
Rabinowicz W, Rønnow-Rasmussen T (2004) The strike of the demon: on fitting pro-attitudes and value. Ethics 114:391–424
Roskies A (2003) Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy". Philos Psychol 16(1):51–66
Sauer H (2011) The appropriateness of emotions. Moral judgment, moral emotions, and the conflation problem. Ethical Perspectives 18(1):107–140
Saver JL, Damasio AR (1991) Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. Neuropsychologia 29(12):1241–1249
Schnall S, Haidt J, Clore GL, Jordan A (2008) Disgust as embodied moral judgment. Pers Soc Psychol Bull 34:1096–1109
Schwartz N, Clore GL (1983) Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. J Pers Soc Psychol 45(3):513–523
Slote M (2004) Moral sentimentalism. Ethical Theory Moral Pract 7:3–14
Smith M (1994) The moral problem. Blackwell, Malden
Solomon RC (1976) The passions. Emotions and the meaning of life. Hackett, Indianapolis
Stocker M, Hegeman E (1996) Valuing emotions. Cambridge University Press, Cambridge
Tappolet C (2011) Emotions, perceptions, and emotional illusions. In: Calabi C (ed) The crooked oar, the moon's size and the Kanisza triangle. Essays on perceptual illusions. MIT Press, Cambridge
Turiel E (1983) The development of social knowledge: morality and convention. Cambridge UP, Cambridge
Valdesolo P, DeSteno D (2006) Manipulations of emotional context shape moral judgment. Psychol Sci 17 (6):476–477
Wheatley T, Haidt J (2005) Hypnotic disgust makes moral judgments more severe. Psychol Sci 16(10):780–784
Wiggins D (1998) A sensible subjectivism? In: Wiggins D (ed) Needs, values, and truth. Clarendon, Oxford, pp 185–214
Williams B (1981) Persons, character and morality. In: Williams B (ed) Moral luck. Cambridge UP, Cambridge, pp 1–19
Wilson TD, Brekke N (1994) Mental contamination and mental correction: unwanted influences on judgments and evaluations. Psychol Bull 116:117–142