

# Moral Judgements and Emotions: A Less Intimate Relationship than Recently Claimed

THOMAS PÖLZLER

Department of Philosophy, University of Graz, Austria

---

**ABSTRACT:** It has long been claimed that moral judgements are dominated by reason. In recent years, however, the tide has turned. Many psychologists and philosophers now hold the view that there is a close empirical association between moral judgements and emotions. In particular, they claim that emotions (1) co-occur with moral judgements, (2) causally influence moral judgements, (3) are causally sufficient for moral judgements, and (4) are causally necessary for moral judgements. At first sight these hypotheses seem well-supported. In this paper I show, however, that appearances are deceiving. If one considers the relevant scientific studies in detail, one finds that in many interpretations the above hypotheses are either not supported or even contradicted by the available evidence. This conclusion is significant both for our understanding of moral judgements *qua* empirical phenomena and for normative ethics and metaethics.

**KEYWORDS:** emotionism, moral psychology, moral judgements, emotions

Despite some obvious exceptions (see, e.g., Hume, 1978, 1983), philosophers have typically considered moral judgements to be dominated by reason rather than emotion (see, e.g., Kant, 1998; Seneca, 1978). When empirical moral psychology emancipated itself from philosophy, it first inherited philosophy's emphasis on reason (see, e.g., Kohlberg, 1984; Piaget, 1965). In recent years, however, the tide has turned. More and more researchers engaged in empirical moral psychology are now sympathetic to the view that it is actually emotions which are in charge of moral judgements. Summing up the current state of research, Joshua Greene and Jonathan Haidt, two of the main proponents of this new "emotionist" paradigm (as I will call

it here)<sup>1</sup>, write: “recent evidence suggests that moral judgment is more a matter of emotion and affective intuition than deliberative reasoning” (2002, p. 517; also quoted in Joyce, 2008, p. 375).

The truth of emotionism would have significant implications for our understanding of moral judgements *qua* empirical phenomena. Moreover, there may also be repercussions for normative ethics and metaethics. Joshua Greene (2008), for example, has recently claimed that as deontological moral judgements are strongly influenced by emotions, and these emotions would lead us to make these judgements independently of whether they are true, deontological moral judgements are epistemically unjustified. We should adopt consequentialism. According to Jesse Prinz (2006, 2007, 2010), emotionism supports a certain view in philosophical moral psychology, namely the view that moral judgements are essentially constituted by emotions.<sup>2</sup> Is emotionism plausible, then? Is the empirical association between moral judgements and emotions really as close as what many psychologists and philosophers have suggested?

At first sight, scientific research really seems to support emotionism, and in a fascinating and entertaining way. Relevant studies involve hypnosis and fart sprays, sex between siblings and the eating of dead pet dogs. This article shows, however, that appearances are deceiving. In section 1 I will clarify various key notions of the emotionist paradigm. Usefully, this paradigm is divided into four sub-hypotheses: (1) the co-occurrence hypothesis, (2) the causal influence hypothesis, (3) the causal sufficiency hypothesis and (4) the causal necessity hypothesis. In sections 2 to 5 I will exemplarily consider scientific studies that have been claimed to support these four hypotheses. It will turn out that on closer consideration, the results of most of these studies do not support emotionism in any significant way. Some of

---

This article is based on my PhD thesis in Philosophy at the University of Graz. I would like to thank Johann Marek (my doctoral advisor), as well as the editor and two anonymous reviewers of the *Journal of Theoretical and Philosophical Psychology* for their helpful comments.

Correspondence concerning this article should be addressed to Thomas Pölzler, Department of Philosophy, University of Graz, Attemsgasse 25/II, 8010 Graz, Austria. E-Mail: thomas.poelzler@uni-graz.at

<sup>1</sup> The hypothesis that moral judgements are closely associated with emotions in the sense at issue here has sometimes been referred to as “emotivism” (Clarke, 2008, p. 800; Haidt, 2001, p. 816) or “sentimentalism” (Prinz, 2013, p. 302). I avoid these labels because “emotivism” is traditionally used to denote a certain metaethical position (the non-cognitivist position that moral judgements are constituted by, and moral sentences express emotions which do not purport to represent moral facts), and “sentimentalism” has been used to denote all kinds of metaethical and empirical claims about the relation between morality and the emotions. For two other philosophers who use the term “emotionism,” though in somewhat different senses, see Jesse Prinz (2007, pp. 14, 16) and Hanno Sauer (2012, p. 105).

<sup>2</sup> Here I do not want to take any stance on the soundness of emotionist arguments in favor of normative ethical or metaethical conclusions. At another occasion, however, I criticized several metaethical arguments of this kind, including the above one by Prinz, see Pölzler 2014, pp. 247-266.

these results even contradict (stronger versions of) the hypotheses of emotionism. Thus, while moral judgements are to some degree empirically associated with emotions, their relationship is not as intimate as what has recently been claimed.

## 1. Clarifications

According to the emotionist paradigm, moral judgements are closely empirically associated with emotions. Below we will see that one's assessment of the empirical relation between moral judgements and emotions significantly depends on what one means by a *moral judgement*. For now, however, let us focus on understanding the other two key notions of emotionism, that is, what it is for a mental state to be an *emotion* and what emotionists mean when they claim that emotions are *closely associated* with moral judgements.

Emotion theorists generally agree that emotions are responses to the perception of some stimulus, that they come along with physiological activation, and that they are therefore closely tied to action (see Johnson, 2009; but see, e.g., Nussbaum, 2004, p. 195). Consider, for example, the emergence and nature or implications of fear. First, I perceive some stimulus: say, I observe or imagine a big dog barking at me (stimulus). I tremble, my heart rate increases, I start sweating (physiological activation), and this physiological activation leads me to run away from the dog, stand still, or shout for help (action).

While there is agreement that emotions are responses to stimuli; that they involve physiological activation; and that they often lead to action, various other aspects of their nature are highly contested. Most importantly, philosophers and psychologists hold fundamentally different views about how the perception of the relevant stimulus is connected to our physiological activation and our subjective experience. William James (1884) and Carl Lange (1885), for example, famously claimed that fear is nothing but one's awareness that one trembles, that one's heart rate increases, that one starts sweating, and so on. Cognitivist theories of emotion, on the other hand, identify emotions with judgements about (e.g., Nussbaum, 2004; Solomon, 1993), or appraisals of (e.g., Lazarus, 1991) the stimulus that persons respond to when they have an emotion. To feel fear in the face of a barking dog, for example, on this view simply means to judge that the dog is dangerous. In other (Pözlner, 2014, p. 265) I suggested that which of these general theories of emotions one assumes significantly affects the plausibility of deriving certain normative ethical or metaethical conclusions from emotionism. With one exception noted below, however, one's general theory of emotions should not have much influence on the plausibility of emotionism itself.

Some proponents of emotionism have argued that moral judgements are closely associated with “full-fledged” emotions, such as anger, disgust, guilt, and shame. In recent years, however, much research has instead focused on what have been called “moral intuitions”: quick, automatic, and strongly motivating affective reactions (Haidt, 2001). Jonathan Haidt, the main proponent of the role of intuitions in moral judgement, explains:

Moral intuition is defined as: *the sudden appearance in consciousness, or at the fringe of consciousness, of an evaluative feeling (like-dislike, good-bad) about the character or actions of a person, without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion [...].* (Haidt & Björklund 2008, p. 188)

According to some theories of emotion, in particular more cognitivist ones, moral intuitions in this sense may not qualify as emotions: they may be too unconscious, too much beyond our control, or have too little cognitive content. And this is where I believe the general debate about the nature of emotions may have some influence on the degree of plausibility one is willing to ascribe to emotionism. In what follows, however, I will grant that moral intuitions — in the sense of unconscious, quick, automatic, and strongly motivating affective reactions — count as emotions as well.

After having clarified the notion of emotions, let us ask what emotionists mean when they hold that moral judgements are “closely empirically associated” with emotions, or, as they have also put it, empirically “grounded” in emotions, “a matter of” emotions, or “based on” emotions (e.g., Prinz, 2006, 2007; Sauer, 2012). Empirical associations between moral judgements and emotions may take various forms. In this article I will focus on four particularly prominent emotionist sub-hypotheses (as distinguished by Prinz, see 2006, pp. 30-33):

1. *The Co-Occurrence Hypothesis*: Emotions co-occur with moral judgements
2. *The Causal Influence Hypothesis*: Emotions causally influence moral judgements
3. *The Causal Sufficiency Hypothesis*: Emotions are causally sufficient for moral judgements
4. *The Causal Necessity Hypothesis*: Emotions are causally necessary for moral judgements

All of these hypotheses can be true in weaker or stronger senses. For example, we may find that moral judgements *sometimes* correlate with emotions, or that they *always* do so; that they

correlate with *weak* or with *strong* emotions; that they are *sometimes* or *often* influenced by emotions; that this influence is *weak* or *strong*, and so on. Contemporary proponents of emotionism typically hold several of the above hypotheses, and in rather strong forms (see, e.g., Prinz 2006, 2007). In the next sections I will examine the plausibility of this emotionist stance, discussing each of the above hypotheses in turn.

## 2. The Co-Occurrence Hypothesis

According to emotionism's first sub-hypothesis, emotions *co-occur* with moral judgements, that is, they *correlate* with them or *accompany* them (for proponents see, e.g., Prinz 2006, p. 30; Goodwin & Darley, 2010, p. 179; Sauer, 2012, p. 97). One might attempt to show this hypothesis true on purely conceptual grounds. In particular, it might be argued that to judge something morally right, wrong, good, bad, and so forth simply *means* having a certain emotion towards it, and that a judgement that is unaccompanied by emotions therefore cannot possibly qualify as moral (for a classic non-cognitivist version of this view see Ayer 1952, p. 108; for a prominent contemporary cognitivist version see Prinz, 2006, p. 34). As this metaethical claim is likely wrong (see, e.g., Egan, 2012: 566; Geach, 1965; Mackie 1980: 69; Miller, 2003, pp. 43–51), and at least very controversial, however, I will here exclusively focus on the available empirical evidence. Advocates of the co-occurrence hypothesis have cited various scientific studies in its favor. In what follows I will consider three of the most often-mentioned ones.

### 2.1. Ultimatum Game 1

The first study to be claimed to provide evidence for the co-occurrence hypothesis — conducted by Alan Sanfey, James Ritling, Jessica Aronson, Leigh Nystrom, and Jonathan Cohen (2003) — focuses on so called “ultimatum games”. Ultimatum games are situations in which two players have to decide on how to divide a given sum of money. The first player makes a suggestion as to how to divide the sum. If the second player then accepts the first player's offer, the money is divided accordingly. If the second player rejects the offer, none of the players get anything.

In their study Sanfey et al. asked subjects to respond to various kinds of ultimatum game offers, some of them fair (the first player proposed an equal split), others to greater or lesser degrees unfair (unequal split to the proposer's advantage). What is unique about the study and has been supposed to make it relevant for assessing emotionism is that, while sub-

jects played the ultimatum game, they were in a functional magnetic resonance imaging (fMRI) scanner which recorded their brain activity. It turned out that when subjects rejected unfair offers, there was increased activity in areas associated with emotions, in particular, with negative emotions (Sanfey et al., 2003, p. 1758). Likely, the rejection of unfair offers in ultimatum games is either motivated or at least accompanied by moral considerations. The second player may think, for example, that a certain split is so unfair that she or he would rather punish the proposer at her or his own expense rather than take the small sum the first player would grant. The result of the study has thus been claimed to suggest that moral judgements reliably come along with emotions.

## **2.2. Ultimatum Game 2**

A second study often claimed to support the co-occurrence hypothesis — conducted by Mascha van't Wout, René Kahn, Alan Sanfey, André Aleman (2006) — also focuses on ultimatum games. Subjects were again asked to respond to various kinds of fair and unfair offers in such games. This time, however, experimenters did not attempt to measure their emotional engagement by neuroscientific methods, but by recording subjects' skin conductance responses (which are indicative of the activity of sweat glands, and thus of emotions or stress). Van't Wout et al. hypothesized that the rejection of unfair offers would be accompanied by significantly higher skin conductance levels than the acceptance of (unfair or fair) offers. And indeed, the prediction was found to hold. Subjects who rejected offers to the proposer's advantage seem to have experienced significantly stronger emotions than subjects who accepted unfair or fair offers (2006, p. 566).

## **2.3. Trolley Dilemma**

The study probably most often claimed to support the co-occurrence hypothesis is Greene, Sommerville, Nystrom, Darley, and Cohen's famous Trolley dilemma study (2001). A moral dilemma is a situation in which one (seems to) have moral reasons to perform two or more different actions, but can actually only perform one of them. Greene et al. presented their subjects with various dilemmas that either involved inflicting harm in some personal or in a more impersonal way. Their most famous dilemmas are the so-called Trolley dilemmas, as they were originally introduced by Philippa Foot (1967), and later famously discussed by the Judith Jarvis Thomson (1985):

## SWITCH

A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one? (Greene et al., 2001, p. 2105)

## FOOTBRIDGE

As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death? (Greene et al., 2001, p. 2105)

While subjects were thinking about these moral dilemmas, Greene et al. had their brain activity recorded by a fMRI scanner. In line with their predictions it turned out that when subjects thought about inflicting harm in an impersonal way, as in turning the trolley (which was typically judged permissible) there was significantly less activity in brain areas typically associated with emotions, and significantly more activity in brain areas typically associated with reasoning than when they considered personal violations such as pushing the stranger off the footbridge (which they typically judged impermissible) (Greene et al., 2001, p. 2107).

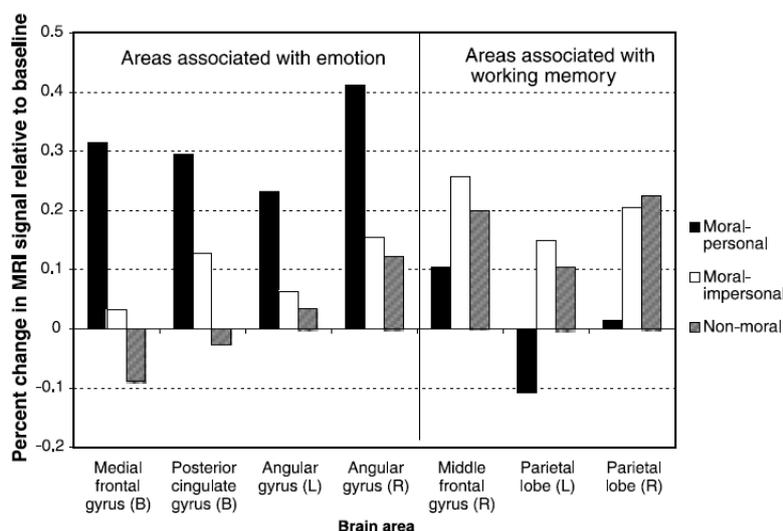
Proponents of the co-occurrence hypothesis typically stress other aspects of the study's results, though. First, they stress that in both impersonal and personal moral dilemmas subjects showed at least some activity in brain areas associated with emotions. And second, they stress that when Greene et al. presented their subjects non-moral dilemmas (e.g., situations in which one has both reason to travel by bus or by train, or to use one or another coupon at a store) subjects showed less activity in brain areas typically associated with emotions than in both moral cases (see Prinz, 2007, pp. 24-25, 2010, p. 3, and in press, p. 9).

## 2.4. Discussion

Taken together, the three studies considered above provide some evidence in favor of the co-occurrence hypothesis. Contrary to how they have sometimes been interpreted, however, they only support the hypothesis in the sense that moral judgements reliably correlate with *weak* emotions. One cannot infer from them that moral judgements correlate with *strong* emotions. In fact, the studies rather seem to undermine this stronger version of the co-occurrence hypothesis. Consider, first, the ultimatum game studies. Subjects in these studies were found to

show strong emotions when they rejected unfair offers (i.e., offers according to which they would receive a significantly smaller amount of the sum to be divided than the proposer). When subjects accepted unfair offers or accepted fair offers they only showed weak emotional activity. However, it is likely that many subjects made moral judgements when they *accepted* offers as well. For example, when they accepted an unfair offer they may have thought, “This guy is depraved, but I nevertheless will take the money!” When they accepted a fair offer, they may have thought, “Wow, that’s admirable!” So if moral judgements reliably correlated with strong emotions, experimenters should have found strong emotions in these cases as well.

Greene et al.’s study contradicts strong versions of the co-occurrence hypothesis even more clearly. As pointed out above, fMRI scans of subjects’ brains suggested significant variation in the activation of areas typically associated with emotions. When subjects thought about the permissibility of causing harms in personal ways, such as pushing the stranger off the footbridge, they likely experienced strong emotions. In SWITCH and other impersonal dilemmas, in contrast, the emotional arousal was only weak. In fact, it was *very* weak, with the level hardly exceeding the base line and being almost as low as when subjects thought about non-moral dilemmas (Greene et al., 2001, p. 2106).



**Figure 1:** Neural activity in subjects considering personal moral, impersonal moral, or non-moral dilemmas in Greene et al.’s 2001 study. Reprinted from “An fMRI investigation of emotional engagement in moral judgment,” by J. D. Greene, B. R. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen, 2001, *Science*, 293(5537), 2106. Copyright 2001 by The American Association for the Advancement of Science. Reprinted with permission.

In sum, then, while the studies considered in this section suggest that most moral judgements reliably correlate with (very) weak emotions, they do not support the view that all or most such judgements correlate with strong emotions. Judgements about hitting a switch in order to save five lives at the expense of one, say — or, as one might speculate, about issues such as climate change or copyright infringement — rather seem to be dominated by reasoned mental processes (see Greene et al., 2001; Paxton et al. 2012).

### **3. The Causal Influence Hypothesis**

Proponents of the causal influence hypothesis claim that emotions do not only correlate with moral judgements, but also stand in a particular causal relation to them. A person's anger, disgust, "moral intuitions" and so forth *causally influence* his/her moral judgements: they lead the person to judge things significantly more right, wrong, good, bad, and so forth than he/she would judge them if he/she did not have these emotions. In a weak sense this hypothesis seems intuitively appealing. Most people would acknowledge that their emotions *sometimes* make them see things in a *somewhat* more negative or positive moral light. Many emotionists believe, however, that the causal influence hypothesis is true in a much stronger sense: *many* or *most* of our moral judgements are influenced by our emotions, and they are influenced in *very substantial* ways. According to Jonathan Haidt's famous social intuitionist model of moral judgement (2001), for example, we tend to judge things morally right, wrong, good, bad, and so forth because of moral intuitions that overcome us when we think about or observe these things. Reason, Haidt believes, typically only comes into play in the form of *post-hoc* rationalization (2001, pp. 814, 817). Such strong versions of the causal influence hypothesis have been thought to be supported by various psychological studies. In what follows I will again consider three of the most often cited studies.

#### **3.1. Disgust 1**

A first study often cited in favor of the causal influence hypothesis — conducted by Schnall, Haidt, Clore, and Jordan (2008a) — focuses on the influence of disgust on our moral judgements. Subjects of the study were divided into two groups. The first group was confronted with stimuli likely to make them feel disgusted. In particular, they were seated in a room treated with "fart spray" (experiment 1), seated in a room that was dirty and untidy (experiment 2), asked to recall a physically disgusting experience (experiment 3), or shown a video clip involving a dirty toilet (experiment 4). The second group was not confronted with these

stimuli. Subjects of both groups were then asked to rate the moral wrongness of a number of actions, for example sex between first cousins or driving instead of walking a short distance. The results of the study are commonly presented as showing that subjects in the disgust conditions judged the actions significantly more wrong than those in the neutral conditions. Thus, disgust seemed to have a significant influence on subjects' moral judgements.

### **3.2. Disgust 2**

The second study — conducted by Schnall, Benton, and Harvey (2008b) — focused on the influence of disgust as well. All subjects of the study had to watch a film clip that involved a dirty toilet. After that they were divided into two groups. One group was asked to complete an otherwise unrelated task involving words associated with cleanliness, for example “pure”, “clean”, or “pristine” (experiment 1), or to wash their hands with soap and water (experiment 2). The other group completed a task involving only neutral words (experiment 1), or were not given the opportunity to wash their hands (experiment 2). When subjects were then asked to evaluate actions such as turning the trolley in SWITCH, or eating one's dead dog, those in the cleanliness conditions are often said to have judged these actions to be significantly less wrong than those in the neutral conditions. Thus, again, disgusting stimuli seem to have had a significant influence on subjects' moral judgements.

### **3.3. Positive Emotions**

Piercarlo Valdesolo and David DeSteno (2006), finally, conducted a study on the influence of positive emotions on moral judgements. Subjects were divided into two groups. One group was shown a comedy video (an episode of Saturday Night Live), the other had to watch an emotionally neutral documentary about a Spanish village. After they had watched the clips, both groups were confronted with the trolley dilemmas, as they were described above. That is, subjects were asked whether they thought it permissible to stop a runaway trolley heading towards five persons by turning it on an alternative set of tracks, where it would only kill one person (SWITCH); or by pushing a stranger off a bridge, where his dead body would block the trolley's way (FOOTBRIDGE). Valdesolo and DeSteno found that those subjects who had watched the comedy video were significantly more likely to judge it permissible to push the stranger off the bridge in FOOTBRIDGE than those who had watched the neutral documentary. Thus, they suggested, positive emotions have a significant influence on our moral judgements (2006, p. 477).

### 3.4. Discussion

Do the three studies considered above really support the view that emotions influence many of our moral judgements, and do so to a substantial degree? Contrary to common interpretations of the studies (see, e.g., Haidt, 2001; Joyce, 2007; Prinz, 2010), I will show that this conclusion is premature.

The first problem with the common interpretation of the above studies suggests that studies employing their methodology may not have much relevance for assessing the causal influence hypothesis in general. In all three studies, let us grant for the moment, subjects who had a certain emotion reported significantly more or less severe judgements than subjects who did not have that emotion. These findings *may* of course be explained by the subjects' emotions influencing their moral judgements. As has been pointed out by several critics of emotionism, however, the data are consistent with various alternative explanations as well. For example, it might be claimed that with regard to some or many of the actions subjects were presented with (say, incest), subjects' emotions only influenced their perception of these actions rather than their moral judgements. They only made them more aware of these actions having certain morally relevant non-moral features (say, of the fact that children from incestuous relationships may be disabled) (Huebner et al., 2009, pp. 1-3).<sup>3</sup> According to another alternative explanation, subjects may have become aware of their having the emotion they were induced to have and incorporated this fact into their (unconscious) reasoning: "I feel disgust when I think of that action. So even though I can't say what it is, there must be something wrong with that action" (see Jones, 2006, p. 50).<sup>4</sup> Such alternative explanations of subjects' diverging responses may be deemed not particularly powerful. Yet in order for studies of the above kind to support (strong versions of) the causal influence hypothesis the studies would nevertheless have to give us reason to consider these explanations inferior to the explanation that subjects' emotions (strongly) influence their moral judgements.

The second main problem with the above studies is that, as they are commonly presented, and as they have been presented here, their results are grossly exaggerated. Proponents of the causal influence hypothesis have often suggested that in all of the above studies, emotions affected the severity of subjects' moral judgements with regard to all or most of the

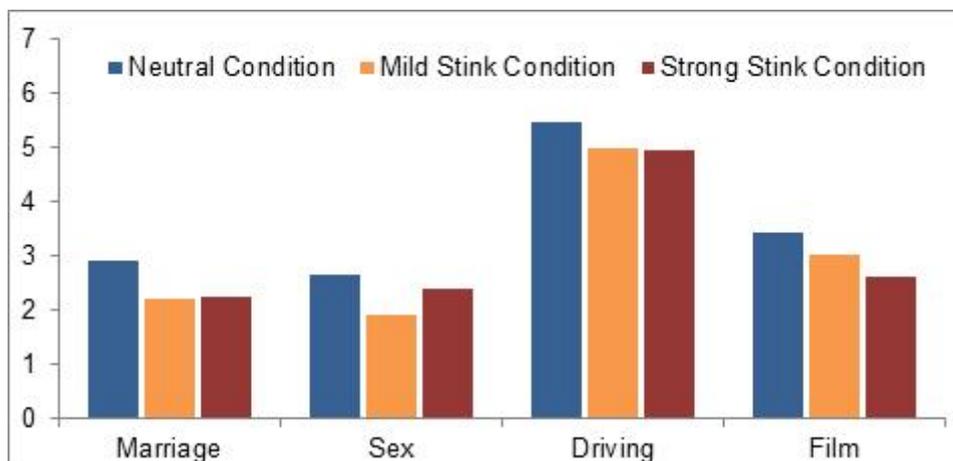
---

<sup>3</sup> This possibility is sometimes even conceded by emotionists themselves, for example by Prinz 2006, p. 31.

<sup>4</sup> On both of the above mentioned explanations emotions influenced subjects' moral judgements in some sense, namely indirectly, via influencing their awareness or reasoning. It might thus be objected that these explanations are consistent with the causal influence hypothesis. As the causal influence hypothesis is understood here, and as it is commonly understood, though, it claims that the effect of emotions is *directly* on people's moral judgements (see, e.g., Huebner et al., 2009, pp. 2-3). I thank an anonymous reviewer for pressing me on this point.

scenarios they were presented with, and to a rather strong degree. But this suggestion is clearly wrong. Consider, first, *Disgust 1*. The first of the four experiments of this study showed a significant general effect of disgust on the severity of moral judgements of all subjects (Schnall et al., 2008a, p. 1099). In experiments two, three, and four, however, only disgusted subjects with a particular feature tended to judge the actions to be significantly more wrong than non-disgusted ones, namely those who were particularly aware of changes in their body (Schnall et al., 2008a, pp. 1100-1103). Moreover, even the results of experiment 1 do not support the causal influence hypothesis in any strong sense.

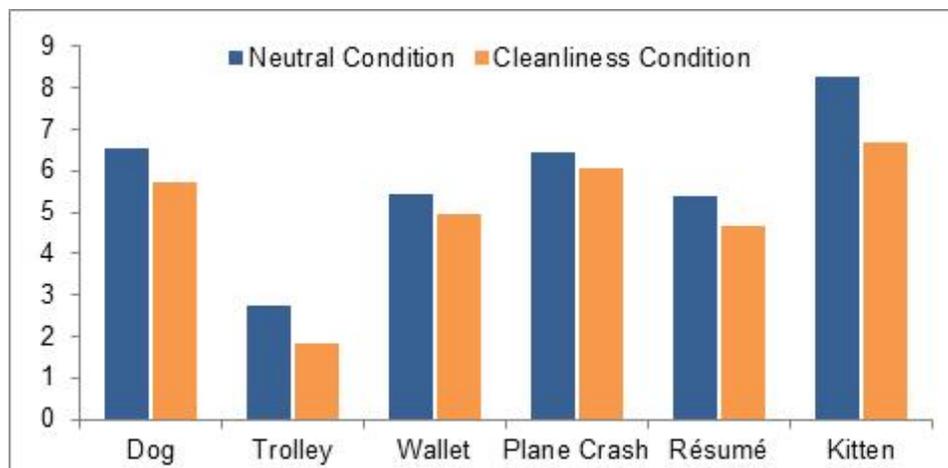
Subjects in experiment 1 were asked to rate the wrongness of actions on a seven-point-scale, with low scores indicating condemnation and high scores indicating permissibility. Those who conducted the survey in rooms that were mildly or significantly treated with fart spray on average judged the actions more wrong than those who conducted it in an olfactorily neutral environment. However, first, the effect was significant only with regard to two out of the four vignettes of the experiment (Marriage and Sex); and second, it was quite weak. In the neutral condition the wrongness of the actions was on average rated 3.75, in the mild stink condition 3.15, and in the strong stink condition 3.18 (2008a, p. 1099). That is, disgusted subjects judged the actions to be only slightly more wrong than non-disgusted ones (see May, 2014, pp. 131-133).



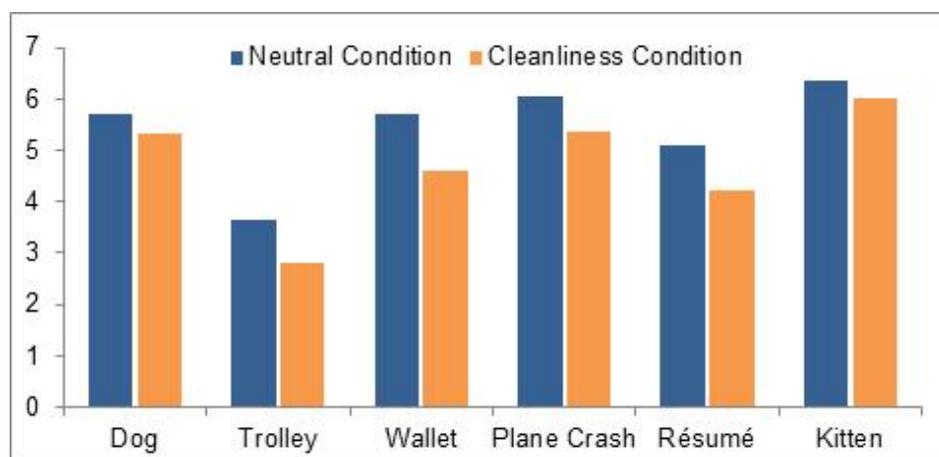
**Figure 2:** Moral judgements in experiment 1 of Schnall et al.’s 2008 study on the causal influence of disgust on moral judgements. Low scores indicate moral condemnation; high scores indicate moral permissibility. Data from “Disgust as embodied moral judgment,” by S. Schnall, J. Haidt, G. L. Clore, and A. H. Jordan, 2008, *Personality and Social Psychology Bulletin*, 34(8), 1099.

The very same problems pertain to *Disgust 2* as well. The results of this study are often claimed to suggest that the suppression of disgust makes our moral judgements milder to a

considerable degree. Actually, however, the effect of coming across cleanliness words/handwashing on subject’s wrongness ratings was found to be significant only with regard to 1 out of 6 of the scenarios presented in experiment 1 (“Kitten”) and 2 out of 6 in experiment 2 (“Trolley” and “Résumé”); and it was again quite small. On the nine-point wrongness scale of experiment 1 subjects in the non-disgust condition rated the actions at an average of 4.98 and subjects in the disgust condition at 5.81 (Schnall et al., 2008b, p. 1220). On the seven-point wrongness scale of experiment 2, subjects in the disgust condition rated the actions at an average of 4.73 and subjects in the non-disgust condition at 5.43 (Schnall et al., 2008b, p. 1221).



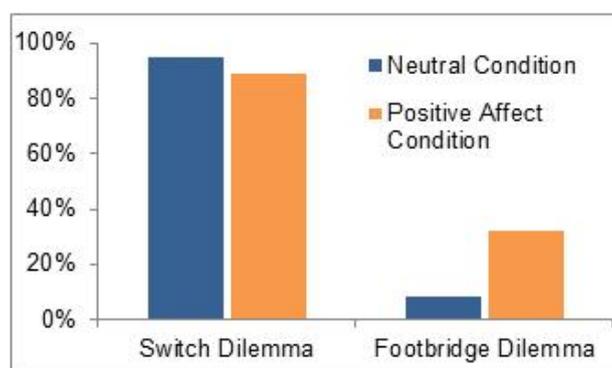
**Figure 3:** Moral judgements in experiment 1 of Schnall et al.’s 2008 study on the causal influence of the suppression of disgust on moral judgements. Data from “With a clean conscience: Cleanliness reduces the severity of moral judgments,” by S. Schnall, J. Benton, and S. Harvey, 2008, *Psychological Science*, 19(12), 1220.



**Figure 4:** Moral judgements in experiment 2 of Schnall et al.’s 2008 study on the causal influence of the suppression of disgust on moral judgements. Scales ranged from “nothing wrong at all” (1) to

“extremely wrong” (7). Data from “With a clean conscience: Cleanliness reduces the severity of moral judgments,” by S. Schnall, J. Benton, and S. Harvey, 2008, *Psychological Science*, 19(12), 1221.

Finally, the results of *Positive Affect* have often been exaggerated as well. Valdesolo and DeSteno’s study shows that positive affect condition subjects judged it more permissible than neutral condition subjects to save five lives at the cost of one in FOOTBRIDGE. Their permissibility ratings did not differ to any significant extent with regard to SWITCH, though (Valdesolo & DeSteno, 2006, p. 477).



**Figure 5:** Proportion of “appropriate” responses in Valdesolo and DeSteno’s 2006 study on the causal influence of positive effect on moral judgements. Data from “Manipulations of emotional context shape moral judgment,” by P. Valdesolo, and D. DeSteno, 2006, *Psychological Science*, 17(6), 477.

In sum, our discussion of the above three studies suggests that the effects obtained in them are not sufficiently large and widespread to support the causal influence hypothesis in any strong sense. In fact, the available evidence even seems to contradict the claim that emotions frequently and strongly influence our moral judgements. Often, the emotions considered in the above studies did not exhibit any or only a limited effect.

## 4. The Causal Sufficiency Hypothesis

According to the causal influence hypothesis, emotions lead us to judge things significantly more or less right, wrong, good, bad, and so forth than we would have judged them without having these emotions. The next emotionist hypothesis discussed here, the causal sufficiency hypothesis, takes this claim one step further. Proponents of the hypothesis argue that the influence of emotions on our moral judgements is so strong that they sometimes or often even causally *guarantee* that we make such judgements. Emotions make us judge things in a positive moral light (right, good, etc.) that we would otherwise have judged negatively (wrong,

bad, etc.); and they make us judge things in a negative moral light (wrong, bad, etc.) that we would otherwise have judged positively (right, good, etc.). Supposedly, emotions can even make us judge things in moral terms that we otherwise would not have judged in these terms at all.

Critics of emotionism have occasionally suggested that the causal sufficiency hypothesis can be dismissed on conceptual grounds (e.g., Jones, 2006, pp. 48-50; Rachels, 1993, p. 483; Sauer, 2012, p. 106). For example, it might be argued that in order for a person's judgement to qualify as moral (as opposed to, e.g., being merely about his/her preferences), it must be based on (relevant) reasons. The person must have come to hold the judgement by considering the interests of human beings, what is just, what is respectful, and so forth. As judgements that are exclusively caused by disgust, happiness, or other emotions do not fulfil this condition, we can rule out *a priori* that any such judgement qualifies as moral, and therefore that there is any *moral* judgement exclusively caused by emotions.

While there may indeed be some conceptual constraints on whether and to what extent a person's moral judgements can be largely or completely caused by emotions, in this paper I start from the plausible assumption that any such constraints can be expected to be rather minor. The question of whether emotions are sufficient for moral judgements is thus primarily an empirical one. So does the available scientific evidence support the hypothesis that emotions are sometimes or often causally sufficient for making moral judgements? In what follows I will consider those two studies that emotionists have most often cited in support of this causal sufficiency hypothesis (see, e.g., Prinz, 2006, p. 31).

#### **4.1. Hypnosis**

The study that proponents of the causal sufficiency hypothesis most often appeal to in support of their view was conducted by Thalia Wheatley and Jonathan Haidt (2005).<sup>5</sup> Similar to some of the research considered in the last section, Wheatley and Haidt's study focuses on the effects of disgust. The main difference to other studies about the relation between moral judgements and emotions concerns the method that Wheatley and Haidt employ in inducing this emotion. Rather than, say, putting subjects in a dirty and untidy room or making them watch video clips, the study draws on the procedure of hypnotic induction. Subjects were hypnotized and implanted with the suggestion that as soon as they awake, they will feel dis-

---

<sup>5</sup> Note that Wheatley and Haidt themselves do not suggest that their study supports the sufficiency hypothesis. They are also more cautious in the interpretation of their data than many of their emotionist commentators.

gust every time they come across a particular word (“often” or “take”). Wheatley and Haidt then presented subjects descriptions of a number of actions and asked them to rate their wrongness on a scale from 0 (“not at all morally wrong”) to 100 (“extremely morally wrong”). Some of the scenarios involved what could plausibly be qualified as a moral transgression, such as second-cousins incest, bribery, or theft (2005, pp. 780-781). Other scenarios did not seem to contain any wrong-doing at all. For example, subjects were told about a student council who picks topics for discussion that appeal to both professors and students alike (2005, p. 782).

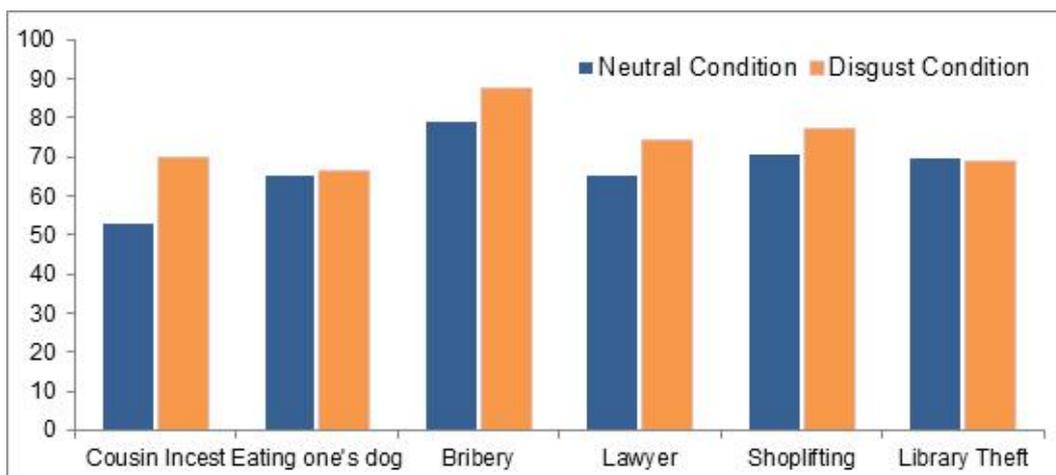
As they are most commonly described, the results of Wheatley and Haidt’s study are astonishing. When subjects received descriptions of the actions that contained their trigger word, and were thus likely to have felt disgust, they rated the actions as significantly more wrong than subjects who had not been confronted with their trigger word. The judgements of disgusted subjects are said to have become harsh even in cases in which we normally would not see any wrong-doing at all. In particular, subjects are often said to have even judged it wrong for the student council to pick discussion topics that appeal to professors and students alike. Asked why they had made these judgements, Wheatley and Haidt report subjects to have written comments such as “it just seems like he’s up to something,” that the student council is a “popularity-seeking snob,” or “I don’t know [why it’s wrong]; it just is” (2005, p. 783).

At first sight the results of Wheatley and Haidt’s hypnosis study seem to provide strong evidence for the causal sufficiency hypothesis. In particular, they seem to suggest that disgust can make us judge things wrong that we would otherwise judge right, or would not judge in moral terms at all. On closer consideration, however, it is doubtful whether such a conclusion can be drawn. Let us first look at subjects’ responses to those actions that can plausibly be said to involve moral transgressions. Here, we find both of the problems pointed out with regard to *Disgust 1*, *Disgust 2* and *Positive Emotions* considered above: the methodological and the presentational one.

First, all that can be understood from the results of Wheatley and Haidt’s study is that disgusted subjects were reported to have made more severe moral judgements. They rated first-cousin incest, bribery, theft, and so forth to be more wrong than those who did not receive their trigger-word and thus did not feel disgust. This finding, however, cannot only be explained by the hypothesis that disgust influenced subjects’ moral judgements, but also by the hypothesis that the disgust made subjects more aware of morally relevant non-moral fea-

tures of the actions they were presented with, or by the hypothesis that subjects interpreted the presence of their disgust as a reason for judging these actions wrong.

Second, Wheatley and Haidt’s results with regard to their transgression cases have also often been presented in a notably exaggerated way (as also pointed out by May, 2014, p. 128). Those subjects who had received descriptions that contained their disgust-inducing word on average made only *somewhat* harsher judgements than those who had not come across their trigger-word. In particular, while non-disgusted subjects rated the actions at 64.7 (experiment 1) and 69.6 (experiment 2), those who were confronted with the disgust-inducing words rated them at 73.9 (experiment 1) and 73.4 (experiment 2) (see 2005, pp. 781-782). Moreover, the effect of the disgust did not occur consistently either. With regard to 4 out of the 6 scenarios of experiment 1 and 6 out of the 7 in experiment 2, the presence of the disgust-triggering word did not have any significant effect on subjects’ wrongness-ratings (see 2005, pp. 781-782).



**Figure 6:** Combined mean wrongness ratings of transgression cases in experiment 1 and 2 of Wheatley and Haidt’s 2005 study on the causal influence of hypnotically induced disgust on moral judgements. Scales ranged from 0 (“not at all morally wrong”) to 100 (“extremely morally wrong”). Data from “Hypnotic disgust makes moral judgments more severe,” by T. Wheatley, and J. Haidt, 2005, *Psychological Science*, 16(10), 781-782.

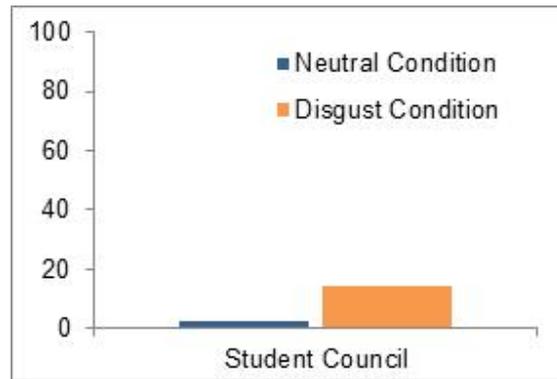
Given these problems, it is not surprising that many proponents of the causal sufficiency hypothesis have in particular appealed to Wheatley and Haidt’s results of their non-transgression part. Here the first of the above-mentioned problems — the methodological one — seems less virulent than in their transgression cases. We cannot rule out that subjects became aware of their being disgusted and incorporated the presence of this emotion into their

reasoning when they thought about the wrongness of the actions they were presented with. However, as the student council's action of picking discussion topics that appeal to both professors and students alike does not seem to have any wrong-making features, subjects' responses at least cannot be attributed to their disgust making them more aware of such features. It is thus more likely than in the transgression cases that the disgust had a direct effect on subjects' moral judgements (see Prinz, 2006, p. 31). That said, the problem of the distorted presentation of the study's results even persists in the non-transgression case.

Many emotionists have taken Wheatley and Haidt's study to show that subjects' disgust affected their judgements about various different non-transgressions, and did so in a substantial way. Their disgust made them judge things wrong that they otherwise would have judged right, or that they would not have judged in moral terms at all (see, e.g., Joyce, 2007, p. 130; Prinz, 2006, p. 31; Sauer, 2012, p. 101). However, this common way of stating Wheatley and Haidt's results is clearly misleading (see May, 2014, pp. 128-131). Across both of the study's experiments, subjects were presented only *one single case* in which an action cannot plausibly be claimed to be wrong: the student council case mentioned above. And even in this single case disgust did not change the severity of subjects' judgements to any large extent, let alone affected their polarity. Those subjects who had received a version of the story that did not contain their disgust-inducing word on average rated what the student council did at 2.7. For those for whom the word was present, the mean value was still only 14 (2005, pp. 781-782).<sup>6</sup> So on average, even disgusted subjects thought hardly anything was wrong with the student council trying to accommodate both the wishes of students and professors.

---

<sup>6</sup> Of all of the 63 subjects of Wheatley and Haidt's study, only two rated the wrongness of the student council's action above the mid-point, and even here only mildly above the mid-point, meaning that they took the action to be somewhat wrong. Presumably, these were the subjects who expressed that the student council "seems like he's up to something," and that it is a "popularity-seeking snob" (see May, 2014, p. 130).



**Figure 7:** Mean wrongness ratings of the non-transgression case in experiment 2 of Wheatley and Haidt’s 2005 study on the causal influence of hypnotically induced disgust on moral judgements. Scales ranged from 0 (“not at all morally wrong”) to 100 (“extremely morally wrong”). Data from “Hypnotic disgust makes moral judgments more severe,” by T. Wheatley, and J. Haidt, 2005, *Psychological Science*, 16(10), 782-783.

In sum, neither the transgression nor the non-transgression part of Wheatley and Haidt’s study provide any significant evidence for the claim that emotions are causally sufficient for moral judgements.

## 4.2. Dumbfounding

Two other studies that have been claimed to support the causal sufficiency hypothesis (conducted by Haidt in collaboration with Koller and Dias in 1993, and in particular with Björklund and Murphy in 2000) have become known under the heading of “moral dumbfounding”. Subjects of both studies were presented descriptions of actions that are likely to evoke disgust, but were supposed not to involve any harmful consequences, neither for the person performing the actions nor for any other persons (see Haidt et al., 1993, p. 617; Haidt, 2001, p. 5). Here are three of the most well-known of these descriptions:

### INCEST

Julie and Mark, who are brother and sister, are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. (Haidt et al., 2000, p. 16)

## FLAG

A woman is cleaning out her closet, and she finds her old [American or Brazilian] flag. She doesn't want the flag anymore, so she cuts it up into pieces and uses the rags to clean her bathroom. (Haidt et al., 1993, p. 617)

## DOG

A family's dog was killed by a car in front of their house. They had heard that dog meat was delicious, so they cut up the dog's body and cooked it and ate it for dinner. (Haidt et al., 1993, p. 617)

For each of the actions described in these vignettes, subjects of Haidt et al.'s studies were asked whether they thought it was wrong. With regard to most of the actions most of the subjects within seconds declared that it was wrong. When they were then asked *why* they thought that the action was wrong, answers took a far less definitive form, though. Most subjects started out by suggesting that the action was wrong because it causes harm. With regard to INCEST, for example, it was pointed out that Julie might become pregnant and bear a disabled child, or that the siblings sleeping with each other might harm their relationship. As the experimenter explained to the subjects, however, these justifications were bad. The actions did not involve any such harm (see Haidt et al., 2000, pp. 8-10). Haidt et al. assumed that if subjects had judged the actions wrong because of their supposed harmful consequences, the experimenter's explanations should have led them to withdraw their judgements and to acknowledge that what Julie and Marc did, or what the persons in FLAG and DOG did, was not wrong after all. However, only a few subjects actually did withdraw their judgements (17% in Haidt et al.'s 2000 study, see 2000, p. 1). The broad majority was reported to have expressed surprise and confusion at it being unable to justify its judgement. They laughed, stuttered, and showed signs of uncertainty. In the end, however, they nevertheless held on to their views, insisting that they *knew* the actions were wrong (2000, p. 9). As Haidt et al. put it, these subjects were "dumbfounded."<sup>7</sup>

Many emotionists have taken Haidt et al.'s finding that subjects held on to their judgements without being able to cite reasons in their favor to suggest that these judgements were not based on reason, but exclusively on disgust. Subjects' disgust was *sufficient* for causing their moral judgements (see, e.g., Haidt et al., 2000, p. 1; Prinz, 2006, p. 31). However, there is reason to be sceptical about the dumbfounding studies' supporting the causal suf-

---

<sup>7</sup> More precisely, Haidt et al. define moral dumbfounding as "the stubborn and puzzled maintenance of a judgment without supporting reasons" (2000, p. 1).

efficiency hypothesis. The fact that subjects held on to their judgements even though they could not cite reasons in their favor *may* be explained by the effect of disgust. There is at least one plausible non-emotionist explanation as well, though. Subjects may have had reasons for judging the actions described in INCEST, FLAT, DOG, and the other scenarios they were presented with to be wrong, but they may have simply been *unable to articulate* these reasons (say, because they did not have enough time to put their thoughts in order, or their judgements were based on the automatic and non-conscious application of moral principles that they could not consciously access; see Jacobson, 2012; Jones, 2006, p. 50; Kauppinen, 2013, pp. 6-7; Wielenberg 2014, pp. 96-107).

Initially, at least Haidt et al.'s main study (2000) seems to rule out this alternative explanation. All of the study's subjects were undergraduate students from a US university. This suggests that the subjects on average had a high socio-economic status, a high level of education, and liberal political views. In other research, Haidt et al. had found that people who have a high socio-economic status, a high level of education, and liberal political views tend to justify their moral judgements predominantly by reference to harm (see in particular Haidt & Joseph, 2007). As the actions described in INCEST, FLAG, DOG, and the other scenarios that subjects were presented with (supposedly) did not involve any harm, this additional evidence seems to imply that subjects probably did not have *any* reason to condemn these actions, not even inarticulate ones. Their judgements were exclusively due to disgust (see Haidt et al., 2000, p. 6).

Following Jacobson (2012), however, I doubt that this appeal to moral foundations research really shows the causal sufficiency explanation to be superior to the inarticulateness explanation. First, some subjects of Haidt et al.'s study could simply have had *bad* inarticulate harm-based reasons for condemning the actions. They could have thought that the actions involve harm in a way that they actually did not involve this harm (Jacobson, 2012, p. 292). Second and more importantly, indeed contrary to what Haidt et al. assume, there may even be *good* harm-based reasons for condemning Julie and Mark's having sex, cleaning one's toilet with the national flag, eating one's dead pet dog, and so on. In INCEST, for example, Mark and Julie take a great risk of compromising their relationship. Although things turned out well, one can easily imagine that their having sex could have introduced an awkwardness that may have stayed between them for the rest of their lives. This recklessness of the siblings, which predicts their actually bringing about harm in other cases, may well justify condemning what they did (see Jacobson, 2012, p. 298-302). Cleaning one's toilet with the national flag or eating one's dead pet dog may also be said to cause harm, in particular, symbolic

harm. These actions may damage the meaning that people ascribe to the flag of a particular country or to the relation between a family and its pet dog (see Jacobson, 2012, p. 309-314; Kauppinen, 2008, p. 71, fn. 130). As Haidt et al.'s results are thus to some degree compatible with the hypothesis that subjects did have reasons for their moral judgements, reasons that they simply could not articulate, they do not lend any strong support to the causal sufficiency hypothesis either.

In sum, *Hypnosis* and *Dumbfounding* provide some evidence for the hypothesis that emotions sometimes causally influence our moral judgements; but they do not by themselves show that emotions are causally sufficient for these judgements, let alone that they are often causally sufficient.

## **5. The Causal Necessity Hypothesis**

The last emotionist hypothesis that will be discussed here is the causal necessity hypothesis. According to this hypothesis, emotions are *causally necessary* for moral judgements. A person *must* have them (in a non-conceptual sense of “must”) in order to be able to judge anything right, wrong, good, bad, and so forth. Following Prinz (2006, pp. 31-32), we can distinguish between two versions of the causal necessity hypothesis: (1) a synchronic and (2) a diachronic version.

### **5.1. The Synchronic Causal Necessity Hypothesis**

On its synchronic reading the causal necessity hypothesis claims that we must have an emotion any time we make a moral judgement: either at the particular time we make the judgement, immediately before we will make it, or immediately after we made it. This entails, for example, that it is only possible to judge torturing puppies for fun wrong, if one feels (say) disapproval towards torturing puppies for fun around the time of that judgement; or that it is only possible to judge helping others good, if one feels (say) approval for helping others around the time of that judgement. Like with the co-occurrence hypothesis, one might argue that the synchronic causal necessity hypothesis can be shown true on conceptual grounds. If it were conceptually impossible for a person to make a judgement in the absence of emotions at that time, then it would be impossible for there to be any causal way of making such judgements in the absence of emotions either. Again, however, I will ignore this likely false, and at least very controversial metaethical claim and stick with the available empirical evidence.

So far no direct scientific research on the causal necessity of emotions for moral judgements has been conducted. Some of the studies we looked at in the previous sections, however, clearly do have implications for this issue. A strong version of the synchronic causal necessity hypothesis claims that any time we make a moral judgement we must have a strong emotion. If this strong version of the synchronic causal necessity hypothesis was true, then for us to judge anything right, wrong, good, bad, and so forth would be causally impossible without having a strong emotion at that particularly moment, or slightly before or slightly after. This prediction, however, is contradicted by the studies on the co-occurrence of emotions and moral judgements considered above. People sometimes do make moral judgements in the absence of strong emotions. The Ultimatum Game studies, for example, suggest that we may only have weak emotions when we make moral judgements that are related to the acceptance of unfair or fair offers on how to split a certain good. Greene et al. found that when people consider impersonal moral dilemmas (such as whether it is permissible to save five lives at the cost of one by hitting a switch), there is only very weak activity in brain areas associated with emotions as well. At best, the synchronic causal necessity hypothesis could therefore turn out true in a weak sense. It could only be the case that we must have *weak* emotions in order to be able to make moral judgements.

## **5.2. The Diachronic Causal Necessity Hypothesis**

Proponents of the diachronic version of the causal necessity hypothesis claim that for a person to be able to make a moral judgement only requires that the person has had (certain) emotions at *any* time of her/his cognitive development. These emotions are supposed to be causally necessary for acquiring a moral capacity, that is, for understanding what it means that something is right, wrong, good, bad, and so forth in the first place. The claim that emotions are causally necessary for moral judgements in this sense implies that it is causally impossible for a person to make moral judgements if she has never had the relevant emotions. An obvious test case for the diachronic causal necessity hypothesis is thus presented by persons who suffer from impairments of morally relevant emotions. If it could be shown that these persons lack our ordinary moral concepts, and that their lack of moral concepts is due to emotional impairments, this result would provide some evidence in favor of the hypothesis.

Proponents of the diachronic necessity hypothesis have argued that there are indeed persons that fit the above description. Their typical example is people suffering from psychopathy: a personality disorder characterized by various anti-social affective and behavioural patterns, such as exaggerated self-worth, manipulative charm, lack of guilt, remorse and

empathy, irresponsibility, impulsivity, and criminal tendencies. Jesse Prinz, for example, writes:

Psychopaths are [...] profoundly deficient in negative emotions, especially fear and sadness. [...] I think that psychopaths behave badly because they cannot make genuine moral judgments. They give lip-service to understanding morality, but there is good reason to think that they do not have moral concepts—or at least they do not have moral concepts that are like the ones that normal people possess. [...] Research on psychopathy suggests that emotions are developmentally necessary for acquiring the capacity to make moral judgments. (Prinz, 2006, p. 32; see also, e.g., Gill & Nichols 2008, p. 144; Greene, 2002, pp. 157-161; Haidt, 2001, p. 15; Nichols, 2002; Sauer, 2012, pp. 98-99)

Proponents of the diachronic causal necessity hypothesis are right that psychopaths show impairments of morally relevant emotions. This claim is supported by various psychological studies. For example, compared to non-psychopaths, psychopaths have been found to show reduced activity in brain areas commonly associated with emotions (e.g., Blair, 2007), to be less aroused by certain kinds of distress cues (e.g., images of screaming children; e.g., Blair, 1997; Blair et al., 1999), and to be worse at recognizing fear in the faces of strangers (e.g., Marsh and Blair, 2008). Psychopaths' inability to feel emotions such as guilt, remorse and empathy is even part of the very criteria that are used to diagnose the condition (see Robert Hare's Psychopathy Checklist – Revised; Cooke & Michie, 2001, p. 172).

If psychopaths lacked our ordinary moral concepts, this fact could also plausibly be explained by their emotional deficits (see Nichols, 2002, pp. 295-299). Some researchers have suggested that psychopaths lack ordinary moral concepts because they are unable to take the perspective of others (see, e.g., Nagel, 1986, p. 140). Two of the main characteristics of psychopaths' condition, however, are that they frequently try to manipulate others and are extremely skilled in doing so (see again Hare's Psychopathy Checklist – Revised; Cooke & Michie, 2001, p. 172). Behavior of this kind clearly requires considering issues from the angle of third persons.

According to another popular alternative explanation, the reason for psychopaths' lack of moral concepts is that they are irrational (e.g., Maibom, 2005). This explanation is worse than ones that appeal to emotional deficits as well. First, the kind of practical irrationality that at least some psychopaths have been found to show is likely a product of their emotional impairment itself (Damásio, 1994, p. 144). And second, a significant proportion of psychopaths do not appear to be irrational at all. This is suggested, among others, by their

success in business and other realms of society (see Babiak & Hare, 2006, p. 193; Board & Fritzon, 2005).

As psychopaths do seem to suffer from impairments of morally relevant emotions and their having these impairments would provide the best explanation of their lacking moral concepts, the crucial question in determining their status as evidence for the diachronic necessity hypothesis is whether they indeed lack ordinary moral concepts. Are psychopaths truly unable to understand what it means for a thing to be right, wrong, good, bad, and so forth? Proponents of the diachronic causal necessity hypothesis have sometimes assumed that this question can be settled by considering scientific evidence as well. For example, it has been argued that psychopaths likely lack ordinary moral concepts because psychological studies have shown them to apply moral concepts in the following unusual ways: (1) psychopaths are more likely than non-psychopaths to advocate pushing a stranger off a bridge in order to save five others in the FOOTBRIDGE version of the trolley dilemma (Koenigs et al., 2011); (2) psychopaths are less likely than non-psychopaths to universalize their moral judgements to social groups or humanity as a whole (see Maibom, 2005); and (3) psychopaths ascribe significantly higher or lower degrees of seriousness, universality, authority-independence, and harm-relatedness to moral norms than non-psychopaths, that is, on one conception of these norms, they are worse at distinguishing between moral and conventional norms (e.g., Blair, 1995; Blair et al., 1997).

Contrary to proponents of such arguments, however, it is doubtful that scientific inquiry of the above kind by itself can determine whether psychopaths have ordinary moral concepts. All that the above studies show (if their results can be trusted) is that psychopaths apply moral concepts *differently* from non-psychopaths. Whether these differences also count as evidence for their lacking ordinary moral concepts depends on what we think those concepts mean. In order for the particular findings mentioned above to support the conclusion that psychopaths lack moral concepts, for example, it would have to be the case that judging it impermissible to cause certain kinds of personal harm, being willing to universalize one's judgements, and ascribing to these judgements particular degrees of seriousness, universality, authority-dependence, and harm-relatedness are essentially related to having moral concepts. Whether these claims hold is metaethically controversial. This means that although the available scientific evidence is consistent with the hypothesis that emotions are necessary to de-

velop moral concepts, this hypothesis must not be accepted without additional philosophical arguments.<sup>8</sup>

## Conclusion

Many psychologists and philosophers nowadays hold that moral judgements are closely empirically associated with emotions. In particular, they claim that emotions (1) co-occur with moral judgements, (2) causally influence moral judgements, (3) are causally sufficient for moral judgements, and (4) are causally necessary for moral judgements. In this article I examined various scientific studies that have been claimed to support the above hypotheses. Even after inadequate methodological assumptions, interpretations and presentations have been corrected, the results of these studies must not be overestimated. Most of them are based on rather unrepresentative samples and involve unrepresentative (sensational and unrealistic) moral issues (e.g., Haidt & Kesebir, 2010, p. 808; Joyce, 2009; Kauppinen, 2008, pp. 92-93). The interpretation of their data often significantly depends on controversial philosophical questions about the meaning of moral judgements (see, e.g., *The Diachronic Causal Necessity Hypothesis*). And the neuroscientific studies that were considered have recently been criticized on specific methodological grounds as well (Klein, 2011).

That said, the above studies provide at least some evidence about the plausibility of emotionism. Although moral judgements are associated with emotions to some degree, this association is less close than has recently been suggested. Moral judgements reliably co-occur with weak emotions, but only sometimes with strong emotions. Moral judgements are sometimes and to some weak degree causally influenced by emotions, but currently no convincing evidence proves that this happens often and that we are led to judge things *considerably* more right, wrong, good, bad, and so forth than we would have done in the absence of emotions. None of the above studies suggests that emotions are causally sufficient for moral judgements, let alone that they are often sufficient. Finally, while it is not clear yet whether emotions are necessary for developing a capacity to make moral judgements, or whether weak emotions are necessary to exercise this capacity in particular cases, we can confidently rule out that we need to have strong emotions in order to be able to make particular moral judgements.

---

<sup>8</sup> As argued by proponents of experimental philosophy, certain kinds of scientific evidence (in particular, evidence about the conceptual intuitions of ordinary people) may be relevant to determining the meaning of our concepts. Conceptual analysis must always include some extent of *a priori* philosophical reflection as well, though (see, e.g., Kauppinen, 2007).

These findings do not only enrich our understanding of moral judgments *qua* psychological phenomena, they also suggest that most attempts to derive normative ethical or metaethical conclusions from strong versions of emotionism are doomed. Such attempts can typically be rejected on purely empirical grounds. At best, philosophical claims may be supported by weak versions of emotionism, but then these philosophical claims are likely to be weak — in the sense of less significant and interesting — themselves.

## References

- Ayer, A. J. (1952). *Language, Truth and Logic*. New York: Dover.
- Babiak, P., & Hare, Richard D. (2006). *Snakes in Suits: When Psychopaths Go to Work*. New York, NY: Harper Collins.
- Blair, R. J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1-29.
- Blair, R. J. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences*, 22(5), 731-739.
- Blair, R. J. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, 11(9), 387-392.
- Blair, R. J., Jones, L., Clark, F., & Smith, M. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences*, 27(1), 135-145.
- Blair, R. J., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack or responsiveness to distress cues? *Psychophysiology*, 34(2), 192-198.
- Board, B. J., & Fritzon, K. (2005). Disordered personalities at work. *Psychology, Crime & Law*, 11(1), 17-32.
- Clarke, S. (2008). Sim and the city: Rationalism in psychology and philosophy and Haidt's account of moral judgment. *Philosophical Psychology*, 21(6), 799-820.
- Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, 13(2), 171-188.

- Damásio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
- Egan, A., 2012. Relativist Dispositional Theories of Value. *Southern Journal of Philosophy*, 50(4), 557–582.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Geach, P. T. (1965). Assertion. *Philosophical Review*, 74(4), 449–465.
- Gill, M. B., & Nichols, S. (2008). Sentimentalist pluralism: Moral psychology and philosophical ethics. *Philosophical Issues*, 18(1), 143-163.
- Goodwin, G. P., & Darley, J. M. (2010). The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. *Review of Philosophy and Psychology*, 1(2), 161-188.
- Greene, J. D. (2002): *The Terrible, Horrible, No Good, Very Bad Truth about Morality and What to Do About It*. Princeton University: PhD Thesis.
- Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology Vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (pp. 35-79). Cambridge: MIT Press.
- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517-523.
- Greene, J. D., Sommerville, B. R., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Haidt J., Björklund F., & Murphy, S. (2000). *Moral Dumbfounding: When Intuitions Find No Reason*. Retrieved from <http://faculty.virginia.edu/haidtlab/articles/manuscripts/haidt.bjorklund.working-paper.when%20intuition%20finds%20no%20reason.-pub603.doc>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.
- Haidt, J., & Björklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed), *Moral Psychology, Vol.2: The Cognitive Science of Morality: Intuition and Diversity* (pp. 181-217). Cambridge: MIT Press.

- Haidt, J., & Joseph, C. M. (2007). The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind. Vol. 3: Foundations and the Future* (pp. 367-391). New York: Oxford University Press.
- Haidt, J., & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert & L. Gardner (Eds.), *Handbook of Social Psychology* (pp. 797-832). Hoboken, NJ: Wiley.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613-28.
- Huebner, B., Dwyer, S., & Hauser, M. D. (2009). The role of emotion in moral psychology. *Trends in Cognitive Science*, 13(1), 1-6.
- Hume, D. (1978). *A Treatise of Human Nature*. Oxford: Clarendon Press.
- Hume, D. (1983). *An Enquiry Concerning the Principles of Morals*. Indianapolis: Hackett.
- Jacobson, D. (2012). Moral dumbfounding and moral stupefaction. In M. Timmons (Ed.), *Oxford Studies in Normative Ethics: Volume Two* (pp. 289-316). Oxford: Oxford University Press.
- James, W. (1884). What is an emotion? *Mind*, 9, 188-205.
- Johnson, G. (2009). Theories of emotion. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved from <http://www.iep.utm.edu/evol-psy/>.
- Jones, K. (2006). Metaethics and emotions research: A response to Prinz. *Philosophical Explorations*, 9(1), 45-53.
- Joyce, R. (2007). *The Evolution of Morality*. Cambridge & London: The MIT Press.
- Joyce, R. (2008). What neuroscience can (and cannot) contribute to metaethics. In W. Sinnott-Armstrong (Ed.), *Moral Psychology Vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (pp. 371-394). Cambridge: MIT Press.
- Joyce, R. (2009). Review: Jesse J. Prinz: The Emotional Construction of Morals. *Mind*, 118(470), 508-518.
- Kant, I. (1993): *Grounding for the Metaphysics of Morals*. Indianapolis: Hackett.
- Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, 10(2), 95-118.

- Kauppinen, A. (2008). *Essays in Philosophical Moral Psychology*. University of Helsinki: Ph.D. Thesis.
- Kauppinen, A. (2013). Sentimentalism. In H. LaFollette (Ed.), *International Encyclopedia of Ethics*. Oxford: Wiley-Blackwell. Retrieved from <http://philpapers.org/rec/KAUSIE.pdf>
- Klein, C. (2011). The dual track theory of moral decision-making: A critique of the neuroimaging evidence. *Neuroethics*, 4(2), 143-162.
- Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2011). Utilitarian moral judgment in psychopathy. *Social, Cognitive, and Affective Neuroscience*, 7(6), 708-714.
- Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of Moral Stages*. New York: Harper & Row.
- Lange, C. G. (1885). *On Emotions: A Psycho-Physiological Study*. Baltimore: Williams and Wilkins.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- Mackie, J. L. (1980). *Hume's Moral Theory*. London: Routledge and Kegan Paul.
- Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind & Language*, 20(2), 237-257.
- Marsh, A. A., & Blair, J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 32(3), 454-465.
- May, J. (2014). Does disgust influence moral judgment? *Australasian Journal of Philosophy*, 92(1), 125-141.
- Miller, A. (2003). *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.
- Nagel, T. (1986). *The View from Nowhere*. New York and Oxford: Oxford University Press.
- Nichols, S. (2002). How psychopaths threaten moral rationalism. *The Monist*, 85(2), 285-303.
- Nussbaum, M. C. (2004). Emotions as judgments of value and importance. In R. C. Solomon (Ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions* (pp. 183-199). Oxford: Oxford University Press.
- Paxton, J. M., Ungar, L. & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- Piaget, J. (1965). *The Moral Judgment of the Child*. New York: The Free Press.

- Pözlner, T. (2014). *Moral Reality and the Empirical Sciences*. University of Graz: PhD Thesis.
- Prinz, J. J. (2006). The emotional basis of moral judgements. *Philosophical Explorations*, 9(1), 29-43.
- Prinz, J. J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prinz, J. J. (2010). *Constructive Sentimentalism: Legal and Political Implications*. Retrieved from <http://www.mit.edu/~shaslang/mprg/PrinzCS.pdf>
- Prinz, J. J. (2013). Emotion and moral judgement. In H. Pashler (Ed.), *Encyclopedia of the Mind* (pp. 300-304). Thousand Oaks: SAGE.
- Prinz, J. J. (in press). *Naturalizing Metaethics*. Retrieved from <http://facultad.pucp.edu.pe/generales-letras/files/2012/02/Naturalizing-Metaethics.pdf>
- Rachels, J. (1993). Subjectivism. In P. Singer (Ed.), *A Companion to Ethics*. Oxford: Blackwell, 432-441.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(5626), 1755-1758.
- Sauer, H. (2012). Psychopaths and filthy desks: Are emotions necessary and sufficient for moral judgment? *Ethical Theory and Moral Practice*, 15(1), 95-115.
- Schnall, S., Haidt, J., Clore, G. L., & Jordan A. H. (2008a). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096-1109.
- Schnall, S., Benton, J., & Harvey, S. (2008b). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19(12), 1219-1222.
- Seneca (1978). *Ad Lucilium: Epistulae Morales*. Cambridge, MA: Harvard University Press.
- Solomon, R. C. (1993). *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett.
- Thomson, J. J. (1985). The trolley problem. *Yale Law Journal*, 94(6), 1395-1415.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17(6), 476-477.
- van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006): Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research*, 169(4), 564-568.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16(10), 780-784.

Wielenberg, E. J. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford: Oxford University Press.