

Does Disgust Influence Moral Judgment?

Joshua May

Australasian Journal of Philosophy vol. 92, no. 1 (2014), pp.125-141.

DOI: [10.1080/00048402.2013.797476](https://doi.org/10.1080/00048402.2013.797476)

(This is a preprint of an article whose final and definitive form is published in the AJP, which is available online at: <http://www.tandf.co.uk/journals/>.)

Abstract: Recent empirical research seems to show that emotions play a substantial role in moral judgment. Perhaps the most important line of support for this claim focuses on disgust. A number of philosophers and scientists argue that there is adequate evidence showing that disgust significantly influences various moral judgments. And this has been used to support or undermine a range of philosophical theories, such as sentimentalism and deontology. I argue that the existing evidence does not support such arguments. At best it suggests something rather different: that moral judgment can have a minor emotive function, in addition to a substantially descriptive one.

Keywords: emotionism, social intuitionism, cognitivism, emotivism, expressivism, moralization

1. Introduction

There has been something of an ‘affective turn’ in the interdisciplinary study of moral psychology. Scientists have apparently amassed converging evidence that emotions play a substantial role in the production of most, if not all, of our moral judgments. One of the most cited and most surprising lines of support for this claim focuses in particular on the emotion of disgust. Largely due to the work of Jonathan Haidt and other psychologists, researchers across disciplines have declared there is sufficient empirical evidence to establish that disgust plays a substantial role in the production of moral judgment.

Of course, it seems rather uncontroversial that disgust influences *some* moral judgments to at least *some* extent, perhaps in concert with other factors. Many of us believe that condemnations of homosexuality are a prime example, and there is some empirical evidence in support of it [e.g. Inbar, Pizarro, Knobe, and Bloom 2009]. The more controversial claim is that empirical research shows disgust can *substantially* influence a *surprising range* of moral judgments, including those concerning actions we don’t ordinarily find revolting (e.g. embezzlement). As Daniel Kelly [2011] puts it in his recent book, disgust is ‘powerful’ [124] as it ‘can have dramatic effects’ on our moral views [130]. Similarly, Alexandra Plakias [2013]

contends that ‘disgust is strongly implicated in moral judgment’ [261] and that such claims are ‘well-established’ [264] by the empirical data.

Various roles for disgust in moral judgment can have theoretical import. The most obvious for philosophers is the support it lends to various brands of sentimentalism that emphasize the role of emotions in moral judgment [Haidt 2001; Prinz 2007], which often feeds into an evolutionary story of its origin [Joyce 2006; Haidt and Bjorklund 2008; Kelly 2011].¹ One can even attempt to undermine epistemological theories, such as forms of ethical intuitionism or foundationalism that rely on some moral intuitions with default warrant. Walter Sinnott-Armstrong, for example, writes that, because ‘emotions can distort moral beliefs in such ways, moral believers need confirmation in order to be justified in holding their moral beliefs’ [2006: 204]. This is especially plausible for moral judgments influenced by disgust because it is arguably a morally irrelevant emotion, which would hinder rather than help the warrant of the resulting judgments (but cf. Plakias [2013]).

Disgust can also seem to debunk a surprising range of moral judgments. Joshua Greene [2008] argues that intuitions driving forms of non-consequentialism, especially ‘rationalist deontology’, arise from such emotional reactions or otherwise less cognitive, rational processes. He points to various pieces of empirical work in support of the claim that emotions play a worrisome role in condemning ‘harmless moral violations’, which utilitarians and other consequentialists want to excise. While Greene has conducted relevant studies of his own, he rightly maintains that the disgust experiments from Haidt and colleagues are the more ‘powerful and direct evidence’ [2008: 58], a trend that crops up among a variety of theorists. After all, the studies point to a strong causal, rather than merely correlational, connection between moral judgment and an emotion that seems morally irrelevant in at least most contexts. Moreover, the studies purport to show that experiencing disgust is sufficient for a change in moral judgment.

On the face of it, the evidence for such claims is impressive. It doesn’t seem especially implausible that disgust can influence moral judgment to some extent or other, and this could serve as a potential building block for understanding the evolution of moral judgment. Moreover, there are now nearly a dozen experiments apparently converging on the idea that this effect is general and substantial. Yet I will argue to the contrary that the existing evidence, while important, does not provide sufficient support for such contentions. Nevertheless, if we can draw any weaker conclusions, they indicate something else about moral judgment, namely that it has some minor emotive function, in addition to a more substantially cognitive or descriptive one. Arriving at these conclusions will require delving into some of the details of the experiments, but the philosophical payoff is considerable. Of course, since this issue is at the mercy of subsequent empirical research, my goal is simply to show that we do not have sufficient evidence at the moment, even if we may in the

¹ Haidt should probably no longer be classified as a sentimentalist. In more recent work he claims not to be contrasting reason and emotion, but rather ‘two kinds of cognition’, neither of which is ‘always a part of an emotional response’ [Haidt and Bjorklund 2008: 200]. His concern is more to show that moral judgment has less to do with explicit *conscious* reasoning.

future. However, critically examining the state of the art uncovers directions for future research and how best to interpret it.

2. Hypnotically-Induced Disgust

Arguably the most often cited evidence for a causal connection between disgust and at least certain moral judgments comes from experiments involving hypnotism. In their first experiment, Thalia Wheatly and Haidt [2005] presented ‘highly hypnotizable subjects’ with six vignettes, all involving a moral transgression. The participants were hypnotized to experience a pang of disgust upon hearing a morally neutral word (half for ‘often’ and the other half for ‘take’). All subjects rated the morality of the protagonist’s act on a line anchored at the end-points with ‘not at all morally wrong’ and ‘extremely morally wrong’, which the experimenters subsequently divided into a 100-point scale. As predicted, participants tended to provide harsher moral judgments when the disgust-eliciting word was present in the vignette. More specifically, the mean morality ratings for the vignettes with the relevant trigger word were higher than the means from stories lacking it, and these differences were statistically significant. So it is unlikely that the differences are due merely to chance.

These results may seem dubious from the outset for two reasons. First, one might worry about the use of hypnotism as a source of evidence. But it is difficult to discount the results when the researchers did find a statistically significant effect. In other words, if hypnotism shouldn’t be trusted to provide good evidence here, why did it produce a detectable difference? Moreover, checks for the effectiveness of the manipulation and statistically significant differences among ratings of disgust bolster the case that something is going on here that we shouldn’t discount. A second worry that immediately arises is that all of the vignettes in the experiments involved a moral transgression, so this does not show that disgust would affect our moral judgments in a case that we would otherwise consider morally permissible [cf. Huebner et al. 2009: 3].

In a second experiment, however, Wheatley and Haidt attempt to address such an issue by including an additional ‘Student Council’ scenario, in which a student performs a mundane, morally neutral action:

Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion. [2005: 782]

Those who read this *without* their disgust-inducing word present tended to rate Dan’s action as ‘not at all morally wrong’ (providing marks near this end of the scale), whereas the ratings were significantly elevated for those who read the version *with* the trigger word. Moreover, the experimenters offered participants an opportunity to explain their judgments, and some wrote of Dan that ‘It just seems like he’s up to something’ or that he seems like a ‘popularity-seeking snob’ [783]. Wheatley and

Haidt conclude that disgusted subjects ‘condemned Dan’ and made ‘severe judgments’ [783].

Now if this is an accurate description of the results, this study would clearly be powerful and surprising, as many have noticed. Plakias, for example, deems it a ‘striking demonstration of the power of disgust [to affect moral judgment]’ [2013: 264]. Moreover, Wheatley and Haidt’s studies apparently provide support for the provocative claim that disgust or emotions generally are sufficient for a change in moral judgment. After all, as Prinz points out, many theorists could happily admit a causal role for disgust by holding, for instance, that ‘negative emotions merely draw our attention to morally relevant features of a situation’ [2006: 31] at which point reasoning processes could play a substantial role [cf. Huebner et al. 2009]. What is especially striking about Wheatley and Haidt’s experiments is that they seem to show that disgust does more than simply colour or intensify moral judgments whose presence are due to some other factor. But this popular account of the hypnotism experiments is problematic for various reasons, which do not stand or fall together.

An immediate worry concerns generalizing merely from a group of subjects with a particular feature that not everyone shares. Notice that the only subjects used in Wheatley and Haidt’s experiments were ‘highly hypnotizable’. Given this, we may rightly generalize from their participants to all highly hypnotizable people. But based on these data alone we cannot conclude with great confidence that disgust influences those who are not highly hypnotizable.

Of course, it might seem implausible that being highly hypnotizable would make one’s moral judgments more susceptible to influence from disgust. One might argue that the hypnotisability of their subjects only allowed the researchers to induce disgust. Perhaps, but the experiments alone don’t help us reach a verdict here. Still, if further experiments reliably produce an effect for more than a particular type of person, we can reasonably ignore the worry in this case. As we’ll see, however, the majority of experiments suffer from a similar problem with generalizing from subgroups.

There is, however, another problem with the popular description of Wheatley and Haidt’s results that is more difficult to address. Recall that subjects recorded their moral judgments in response to a range of cases. Yet Wheatley and Haidt didn’t observe an effect of disgust on moral judgment for most vignettes (not a statistically significant one at least). There were some, but only a small minority: 2 out of 6 in Experiment 1 (Cousin Incest and Bribery), and only 1 out 7 in Experiment 2 (the Student Council case). So, there was *no* effect found on the mean morality ratings for the vast majority of the vignettes, but there was an effect for the mean of those means in both experiments. While this is a legitimate way to approach one’s data, it obscures the fact that the researchers didn’t detect a difference in the vast majority of individual vignettes and that some factor specific to some cases could be driving the overall effect observed. (A similar worry arises for Greene’s imaging studies: see McGuire et al. [2009].) These points are lost when the literature focuses on these means of the means. For example, Kelly writes regarding these studies: ‘Across the board, ratings [of moral wrongness and disgustingness] were more severe when

disgust was induced' [2011: 25]. This isn't accurate, assuming 'across the board' to mean: the same for everything in the relevant class.

Admittedly, it could be that disgust does affect our moral judgments about most of the individual vignettes but the researchers didn't find it in their sample. After all, failing to find an effect doesn't mean there isn't one—unless of course the study has the statistical 'power' to accept the null hypothesis (that there isn't an effect). But experiments in the social sciences are often underpowered, which precludes this inference. At best, then, we have no evidence either way, in which case we still shouldn't say there is an effect 'across the board' when one wasn't found.

Still, it may be significant that the average response of disgusted participants is more severe when we total up their responses to every vignette and average them. But there remains a problem with how significant any of the differences are themselves, whether regarding these means averaged across all cases or an individual vignette. Consider their important Student Council case, for example. The mean rating of moral wrongness for those who did not receive the version of this story with their disgust-inducing word was 2.7 (recall: 0 = 'not at all morally wrong' and 100 = 'extremely morally wrong'). Disgusted participants, however, had a mean rating of 14, which still seems to count the action as *not* morally wrong [cf. Mallon and Nichols 2010: 317–8].

Some researchers are not concerned about whether their participants' responses tend to fall on opposite sides of the midpoint, so long as the difference is statistically significant. For example, in their study of how moral judgments affect various intuitions in folk psychology, Dean Pettit and Joshua Knobe explicitly propose to disregard whether responses tend to straddle the midpoint [2009: 589–90]. While this may be a fine approach to some research questions, it can over-inflate the import of certain results, and the experiments on disgust are a clear example. It's of course unclear whether we should take the scales used in such research to have a genuine midpoint at all, or to otherwise clearly deliver information about whether participants tended to judge the action as right or wrong, rather than being uncertain or ambivalent. This further poses a problem for the sorts of claims many have made regarding these studies, especially the Student Council case. Still, it is useful to consider where on these various scales subjects were tending to fall, even if it is difficult to determine a polarity for the mean response.

Consider how the data conflict with the usual descriptions of Wheatley and Haidt's hypnotism studies. Jesse Prinz, for example, summarizes their second experiment thus: 'when the trigger word is used in [morally] neutral stories, subjects tend to condemn the protagonist'—'[they] find this student morally suspect' (2007: 27–8). Likewise, Richard Joyce writes that the subjects responding to the Student Council story: 'were often inclined to follow up with a negative moral appraisal' [2006: 130]. Kelly similarly writes: 'Participants maintained their unfavourable judgment of Dan despite their complete lack of justification for it....' [2011: 25]. And Plakias says, 'subjects who had been hypnotized judged Dan's actions morally wrong' [2013: 264]. Contrary to all of the above descriptions of Wheatley and

Haidt's results, if anything it appears their subjects tended to regard the representative's action as *not* morally wrong. The studies certainly don't provide evidence that 'disgust is sufficient to bring about an appraisal of moral wrongness even in the absence of a moral violation' [Plakias 2013: 264].

While the different morality ratings between the groups may not straddle the midpoint, one might contend that the effect is nonetheless substantial. Kelly, for example, claims Wheatley and Haidt's disgust-manipulation 'increased judgments of disgustingness and moral wrongness by factors of roughly 10 and 6, respectively' [2011: 25]. While it's true that the morality ratings of subjects increased by a factor of 6 (means were 2.7 vs. 14 in the Student Council case) in the *direction* of the 'extremely morally wrong' end of the scale (100), again this looks if anything to be on the side of counting Dan as *not* having done something wrong. The factor by which it increased along the 'moral wrongness' scale would have to be higher just to get it barely in the realm of being judged somewhat morally wrong (i.e. above 50). So, while disgust may have made participants' judgments more 'harsh' (as some more carefully put it), we do not have evidence that it tended to alter their polarity—e.g. from permissible to wrong. Such data only warrant something like the conclusion that disgust *amplifies* moral judgments in the direction of condemnation [cf. Pizarro, Inbar, and Helion 2011].

At this point, some would retort that in fact some of the disgusted participants rated Dan's action as immoral. Greene, for example, says, 'Many subjects who received matching posthypnotic suggestions indicated that his behaviour was somewhat wrong, and two subjects gave it high wrongness ratings' [Greene 2008: 58]. Such claims are apparently based on an earlier version of the manuscript that circulated prior to publication, which discusses some additional details about earlier versions of the data.² But the 'many' to which Greene refers was a *minority* of the group (about 20% by my calculations), and their ratings are only reported (in the manuscript) as being 'above 15' which is still well on the '*not* morally wrong' side of the 100-point scale. Furthermore, the two subjects (out of 63) who allegedly provided 'high wrongness ratings' were at most in the area of judging the act *somewhat* morally wrong ('above 65'). More importantly, these data points are mere outliers—the kind that are often removed from analysis in experimental work. However, even if we included the data points from the older manuscript and the authors' description of them, Greene's gloss is fairly misleading and the outliers are irrelevant anyhow. What matters are the central tendencies of subjects' ratings, which we can subject to statistical analysis to determine whether the results were likely due to chance. And the means from both groups are still quite low (14 in the published article; 15 in the prior manuscript), indicating either way a tendency to count the act as morally permissible.

There is a further worry about the way Wheatley and Haidt's experiments have been characterized in subsequent literature. Contra Prinz and others who refer

² Thanks to Thalia Wheatley (via Walter Sinnott-Armstrong) for clarifying this issue and providing the earlier version of the paper.

to neutral stories (plural), there was only one morally neutral vignette presented to participants (the Student Council case). This isn't an isolated misunderstanding. Cushman, Young, and Greene refer to two stories: the Student Council case and 'a story about cousins who often visit the zoo' [2010: 57]. But this is nowhere in the published article or the earlier manuscript that I have seen.

Finally, to further support the alleged effect of disgust, many authors also point to the written explanations subjects provided regarding the Student Council story. While some disgusted participants did explain their morality ratings by reporting suspicions of Dan and so forth, Wheatley and Haidt don't report the percentages. They only tell us that 'some participants' engaged in this apparently post-hoc 'search for external justification' [2005: 783]. And these existential generalizations can be true even if only a small minority of participants provided such explanations (e.g. the two outliers). Indeed, while Wheatley and Haidt provide no explicit indication either way, it is likely that only a small minority provided these rationalizations, since only a small minority provided harsher moral judgments, and only two outliers provided a response that indicates a condemnation of Dan's behaviour. So we shouldn't be led into thinking that the above reports from some of their subjects are representative of those who reacted to the Student Council case with disgust.³

3. Non-Hypnotic Disgust

A second popular line of evidence for the role of disgust in moral judgment involves inducing disgust without hypnosis. Schnall, Haidt, Clore, and Jordan [2008] conducted four experiments that manipulated disgust in four ways, respectively: a foul smell, a disgusting workspace, recollection of an experience involving physical disgust, and watching a revolting film clip involving a toilet (from the film *Trainspotting*). After the manipulation of disgust, the experimenters had participants record, among other things, their moral judgments about people in a number of vignettes. These were then compared to the mean ratings from a control group who were spared the disgust manipulation. For any differences in means, statistical analysis would determine whether we can be confident that the differences weren't simply due to chance.

In the majority of their experiments (all but Experiment 1), Schnall and her colleagues observed an effect of disgust on the severity of moral judgments only among those participants who were more sensitive to changes in their bodily state. These subjects scored high on the Private Bodily Consciousness scale, tending to agree with statements such as 'I know immediately when my mouth or throat gets dry' and 'I am quick to sense the hunger contractions of my stomach'. In one experiment, for example, disgust was induced by having subjects fill out their

³ Eskine, Kaciniak, and Prinz [2011] have recently extended the findings of Wheatley and Haidt by using their stimuli but manipulating gustatory disgust. As they didn't use hypnotism, the effect observed was not limited to a subgroup of highly hypnotizable participants. However, their study and results possess all the other limitations found in Wheatley and Haidt's hypnotism experiments.

questionnaire at a desk that was clearly disgusting—it was dirty, sticky, had remnants of old food near it, and so forth. Among other things, participants recorded their moral judgments regarding six moral vignettes, some of which many would independently find disgusting (e.g. cannibalism), while others not (e.g. providing false information on a resume), but none of which were morally-neutral (like the Student Council case). Those who tested high on the Private Bodily Consciousness scale provided more severe moral judgments than those in the control group who completed the task at a clean desk. The same basic finding was replicated in two of their subsequent experiments using different methods of inducing disgust.

An obvious worry for those making claims about the role of disgust in moral judgment is that the effect was only observed among a subset of subjects: those testing high on the Private Bodily Consciousness scale. As with Wheatley and Haidt's studies, this doesn't provide strong evidence that disgust affects the moral judgments of people generally (even if it does for a certain subgroup). Of course, as before, it could be that we can only observe this phenomenon in those who are especially sensitive to their bodily state, even though it is there for everyone. But there is no independent support for this hypothesis, experimental or otherwise.

To be fair, in their first experiment, Schnall and colleagues did find an effect for their participants generally, not just a subset of them. This study induced disgust by using a commercial 'fart spray' which was present while participants registered their moral judgments about four hypothetical cases (distinct from the ones in Experiments 2–4), some of which one might independently find disgusting (e.g. sex between first cousins), but none of which were morally-neutral. However, this is but one experiment, and it is unclear whether it saves the other experiments from the worry, especially given that subjects responded to a different set of vignettes in Experiment 1. So it looks as though we are left with fewer experiments that avoid generalizing from subgroups, and thus less in the way of converging evidence.

In any event, there are further issues that apply to all their studies, though we can focus on Experiment 1 for simplicity. First, as with Wheatley and Haidt, there only appear to be minor differences between the two key groups. On a 7-point Likert scale (with lower numbers indicating condemnation), the mean composite response regarding the morality of the acts in question for the control group was 3.75, compared to 3.15 and 3.18 in the Mild and Strong Stink groups, respectively. This does not seem substantial on the face of it. Yet all of the observed differences between the control group and the disgusted groups are fairly minor in this way. Moreover, the means don't straddle the midpoint (4.0) or in any way suggest one group condemned while the other didn't. Again, the disgusted group only tends to provide somewhat harsher moral judgments on the scale. The differences between the sampled groups are statistically significant, so we can confidently conclude that they are not explained merely by chance. But the results don't warrant strong claims to the effect that disgusted participants tended to count the actions as wrong. Greene,

for example, writes that ‘the disgust manipulation made people more likely to condemn these actions [in the stories]’ [2008: 58].⁴

One might object that the average response from the control group is quite close to the midpoint but not for the other two groups. And this might seem to suggest that those not especially disgusted tended to be ambivalent or unsure and disgust then shifted the central tendency to condemnation. This is still not, however, evidence of the stronger claim that there was a shift from one moral judgment to another, which would more clearly support the claims many have made with respect to these experiments. Moreover, it is difficult to even draw the weaker conclusion for several reasons. First, something like the following would have to be true for this defence to work: the difference between the midpoint and the average response from the disgusted groups was statistically significant while this was not true for the controls. But we have no indication either way in the authors’ article. Second, the means here are the mean composite, which reflects the average response to all the vignettes combined, as in Wheatley and Haidt’s hypnotism studies. And this raises a worry of its own.

So the second major issue is that, while focusing on the mean composite is a legitimate procedure, it again obscures the fact that an effect was not found for each vignette. An effect was detected in only 2 out of the 4 (Marriage and Sex) for the Mild Stink group and only 1 (Film) for the Strong Stink group. This issue arises for every one of their five experiments. In fact, for some, matters are worse. In their last two experiments, for example, only an analysis of the mean composite produced a statistically significant difference, not any individual vignette.

Also of note is a similar group of experiments involving cleanliness, which presumably reduces the experience of disgust. Schnall, Benton, and Harvey [2008] sought to make participants’ moral judgments *less* severe by making them feel clean, or to even just think about purity and related notions. The results were roughly as the researchers predicted, and an important feature of these experiments is that the effects were found for normal subjects, not just ones who tested high on the Private Bodily Consciousness. Nevertheless, all of the other issues from the previous experiments involving disgust arise with these as well. The same goes for a related study by Chen-Bo Zhong and colleagues [2010], who report similar findings (but suggesting that the target of the cleanliness and reduced disgust is important). I leave the details for the reader.⁵

⁴ One might merely read Greene as claiming that disgust makes condemnation ‘more likely’ in that disgust increased mean morality ratings in that direction, even if the mean (or any subject’s rating) doesn’t amount to something we should count as condemnation (e.g. by being significantly above the midpoint). But the reading of ‘more likely’ that is required for his main argument is something like the stronger one preceding the quotation.

⁵ Some might also point to Shaun Nichols’s work on disgust and the moral/conventional distinction. But Nichols didn’t manipulate disgust; he created two groups based on their disgust sensitivity. More importantly, he found no statistically significant difference between these groups’ judgments about the permissibility of the relevant transgressions, which comes closest to a moral judgment [2002: 231].

4. Further Issues

The data gathered to date do not support the strong claim that disgust substantially influences a surprising range of moral judgments. Some of the reasons for this are specific to a given set of experiments, while others afflict them all. Recapping the overarching concerns will clarify where we stand.

- (1) *Generalizing from Subgroups*: The vast majority of the studies reveal effects only for a sub-group within their sample of normal subjects.
- (2) *Scarce Effects*: Statistically significant differences are observed in only a small minority of vignettes (if any) and then the mean of these means.
- (3) *Small Differences*: When researchers do find differences in moral judgment between disgusted subjects and controls, they are quite small.

This last problem is, to my mind, the most substantial. No study has demonstrated that the polarity of a moral judgment tends to shift with disgust (e.g. from ‘moral’ or ‘okay’ to ‘immoral’). As we saw, Wheatley and Haidt’s second experiment comes close, but their subjects by and large tended to judge the student council representative’s action as on the ‘not at all morally wrong’ side of the scale. So this study does not constitute empirical evidence in support of sentimentalist claims such as: ‘we can form the belief that something is morally wrong by simply having a negative emotion directed towards it’ [Prinz 2006: 31].⁶

It is especially difficult to tell how substantial a difference in mere severity disgust can cause without anything to serve as a clear comparison to these results. There has been some preliminary work relevant to this issue. Laham, Alter, and Goodwin (2009) showed that subjects’ moral judgments can be made less severe if they are presented with a vignette in a font and background that made it easier for them to read in comparison to the previous three vignettes (a matter of ‘processing fluency’). In this experiment, subjects rated six cases for moral wrongness using a 10-point scale (1 = *not at all*; 10 = *very much*). For half of subjects, the first three cases were difficult to read while the others were easy; the reverse occurred for the other group. For analysis, the researchers focused on the fourth vignette, which would either be refreshingly easy for participants to read compared to the previous three or annoyingly difficult, depending on which group they happened to be in. For the refreshingly legible cases in this crucial fourth position, which appeared directly after three that were more difficult to read, the mean was 7.54, whereas the reverse order yielded a mean of 8.70, and this was quite statistically significant ($p < 0.01$). It seems that the relief of no longer having to read the difficult font tended to make the participants’ moral judgments less severe. What’s interesting for comparison is that the observed differences are roughly 1 point on the same half of a 10-point scale. This looks similar in magnitude to differences found in the disgust experiments. A more complete investigation is surely required, which examines effect sizes and other aspects of the data. The point is simply to highlight the possibility that the

⁶ I don’t believe there is any reason, however, to think that scales for measuring attitudes (like the Likert scale) are inherently problematic. One just needs to be sensitive to the issues involved [cf. May and Holton 2012].

effect of disgust on moral judgment is about as potent as the feelings that unexpected legibility produces. If correct, this further suggests that the significance of the disgust effect has been overstated.

The issue calls out for further research. As it stands, we have little evidence about the substance of the scales used in moral judgment research. We need something like a calibration of these rather artificial instruments for measuring moral judgments, presumably by using benchmarks in clearer, less controversial cases.⁷ One would need to conduct similar experiments as Haidt and his colleagues have done, but involving a range of independent variables. Some of these should clearly have an important effect on moral judgment (e.g. whether the victim is an animate versus inanimate object), while others shouldn't (e.g. readability of the font the vignette is presenting in). In this way, we could hopefully achieve standards for the significance of results using these measures for moral judgment. Given the results of Laham et al. [2009], we might expect that disgust plays as much of a role in moral judgment as the feelings of relief resulting from unexpected legibility, rather than factors like harm, violations of rights, and so forth.

Finally, there is a further worry about the work on disgust by Haidt, Schnall, and their collaborators given recent empirical work. A group of experiments provide evidence that disgust, if it is to have any causal influence at all, is restricted to moral judgments in a certain domain some psychologists have come to call 'purity' [Horberg, Oveis, Keltner, and Cohen 2009]. This includes moral issues 'directed at protecting the sanctity of the body and soul' [964], such as appropriate sexual behaviour. We needn't belabour this point, as the idea of domain-specificity fits with research done previously by Haidt, along with other colleagues [Rozin, Lowery, Imada, and Haidt 1999].⁸ At any rate, such results suggest another limitation on the effect of disgust. Instead of being able to influence moral judgments generally, disgust seems to at most slightly affect our moral judgments in a domain-specific way. We might need more data to gain a complete understanding of the extent of this domain, but at the moment it seems to be something like purity.⁹

⁷ The general idea here is that we should have a clearer calibration of our instruments in new areas of psychological research. My thinking about this was influenced by a presentation of Jonathan Weinberg's at a mini-conference on Experimental Epistemology in 2011.

⁸ Some of the vignettes that yielded an effect for Haidt and colleagues don't seem to fall within the purity domain (e.g. Bribery, Film, Student Council), which may conflict with the conclusion drawn by Horberg et al. The problems we have raised here for the experiments conducted by Haidt and his collaborators cast some doubt on the conflict, but further research is surely required to fully adjudicate the issue.

⁹ Of course, we must also pay close attention to whether the results connecting disgust and moral judgment can be replicated, especially after making relevant improvements. For some preliminary evidence suggesting problems with replication, see Case, Oaten, and Stevenson [2012], although their experiments aren't necessarily the last word on the matter.

5. Ethical Implications

It is undoubtedly important in its own right to determine whether disgust plays a role in moral judgment, but there are further implications, as we've seen. Most saliently, many theorists appeal to precisely the studies on disgust, though not always those studies alone, as a key pillar of evidence that emotions considerably influence all kinds of moral judgments or deontological judgments in particular. The relevant premises of such arguments can hardly take refuge in experiments which provide only *minor* support for the claim that disgust *slightly* affects a *narrow* (and arguably unsurprising) class of moral judgments.

Consider first the debunking-style arguments, which attempt to use disgust's purported influence on moral judgment to undermine the warrant of some moral intuitions. Sinnott-Armstrong [2006] attempts to undermine moral intuitionism by arguing that moral beliefs fail to enjoy warrant without independent confirmation because they can be distorted by emotions. Greene [2008] argues that deontological intuitions are unwarranted in the same way, as he believes the distortion is roughly limited to that class. The empirical evidence under discussion provides a key line of support for these arguments by apparently showing that the relevant moral beliefs can be biased by emotions like disgust.

A proper conception of the results puts pressure on such debunking arguments. First, if disgust's influence is limited in scope, then so are the corresponding arguments. Greene claims that emotions like disgust lead to deontological judgments, such as intuitions apparently driven by the doctrine of double effect, which is problematic for Kantians and other deontologists. Yet some core applications of deontological principles, such as in cases of self-defence, are not plausibly in the purity domain. Thus, it is only if disgust's influence stretches beyond a narrow domain that it then raises problems for deontological intuitions. Similarly, Sinnott-Armstrong's debunking argument cannot apply to moral judgments generally if we only have evidence that emotions like disgust influence a more limited class. After all, then only that class would be questionable, and intuitionists could maintain that some other core moral judgments enjoy default warrant.

A second problem with the debunking arguments is that disgust presumably renders moral judgments unwarranted by preventing them from tracking the truth. In the experiments above, however, subjects were asked about the morality of an act depicted in a scenario, and the *polarity* of their judgments didn't appear on average to vary with disgust. The experimenters did use scales, but it is difficult to decide how exactly we should interpret movement along them. Whatever way we should eventually go on that issue, slight movement on one side of such scales seems clearly irrelevant to whether one's belief is accurate concerning whether an action is simply right or wrong.

Perhaps moral rightness and wrongness come in degrees and the scales were picking up on participants' judgments about these gradable properties. In that case, it may seem that the studies show that disgust distorts moral judgments about *how* right

or *how* wrong some act is.¹⁰ For example, the evidence may suggest that elevated disgust tends to make us count bribery as extremely wrong when in reality it is only moderately wrong. It's not obvious that morality comes in degrees this way, at least for the properties of right and wrong (or immoral), as opposed to evaluative properties, such as good and bad, which weren't represented on the scales. Talk of acts that are 'extremely wrong' may be about as loose as talk of acts that are 'extremely illegal'. Nevertheless, even granting this, it is unclear whether the scales are measuring beliefs about degrees of morality rather than degrees of belief. While the scales don't mention anything like confidence, subjects could be opting for a slightly different spot on the line precisely because they are more or less confident and simply express it this way.

Even if we set both of these worries aside, the existence and reporting of gradable moral properties doesn't necessarily yield bad news for the relevant targets. Greene's deontologists are primarily concerned, for example, to build their theories on *which* acts are right and wrong, not *how* right or wrong they are. Ethicists don't normally test their theories by comparing, say, trolley cases that differ slightly in their degree of morality. (Imagine: 'It's clearly wrong to flip the switch in Case A; but it's slightly *more* wrong in Case B; therefore, utilitarianism is true'.) Similarly, in order to guide their own and others' actions, ordinary moral agents are chiefly concerned to know which acts are right or wrong, rather than about how right or wrong they are. Perhaps comparing the morality of acts in terms of degrees is a primary concern in certain contexts, but the case against deontology and intuitionism at this point becomes quite strained. Moreover, as we've seen, we currently lack the empirical data to fully determine the significance and extent of disgust's influence on moral judgment. So far the evidence we have seems to suggest that it's rather minor, and thus its debunking power is likewise limited.

Next, consider philosophical arguments that rely on a causal role for emotions, such as disgust, regardless of how it impacts the warrant of the beliefs they generate. Prinz [2006; 2007] claims that emotions are sufficient for moral judgment. And he supports this in large part by appeal to the disgust experiments, claiming that elevated repugnance can make us moralize scenarios we'd normally consider to be lacking a moral dimension. This is important because, as we've seen, rationalists can explain the fact that some moral judgments are influenced by disgust to some extent. When certain emotions are relevant to the scenario (e.g. disgust at sexual crimes), sentiments may simply draw one's attention to morally relevant features of the situation (e.g. profound visceral harm). But this explanation isn't available to the rationalist if disgust can influence a wider and more surprising class of moral judgments, such as those concerning acts that normally wouldn't elicit disgust at all (e.g. stealing) or morally neutral acts (e.g. choosing topics for discussion). Yet, as we've seen, only something like Wheatley and Haidt's Student Council case has the ability to rule out such a rationalist explanation, and it utterly fails to do so. The distinction between altering the polarity of the judgment and altering its severity is

¹⁰ Thanks especially to Daniel Nolan for discussion on this point.

again of the utmost importance. The weakness of disgust's influence can hardly support such brands of sentimentalism.

Of course, Greene, Sinnott-Armstrong, and Prinz all point to a wide range of empirical evidence. However, as we have emphasized, the experiments under consideration are arguably the most important here. There is a wealth of research on disgust, but only some of it directly addresses the causal impact on moral judgment [see Pizarro et al. 2011]. Some researchers also point to Haidt's famous 'moral dumbfounding' studies [Haidt, Koller, and Dias 1993], which involved interviewing participants about various 'harmless taboo violations', some of which were disgusting. But these studies didn't manipulate disgust experimentally [for further worries see Kennett 2012]. Likewise, brain imaging data only provide correlations; and lesion studies have thus far only provided some suggestive data concerning complex areas of the brain associated in part with emotional processing [Huebner et al. 2009]. Such evidence is useful when combined with the disgust experiments, but the additional studies do not alone establish the contentious claims required by the philosophical arguments at issue.

There are some experimental manipulations of emotions other than disgust, indicating that they have some influence on moral judgment. These too have limitations [for a critical review, see Huebner et al. 2009]. However, even if we take them at face value, the results do not provide much help for the various philosophical arguments on offer. First, they indicate nothing about disgust or emotions like it. In fact, the relevant emotions are primarily 'prosocial', such as feelings associated with empathy, the causal relevance of which is far from incompatible with rationalism. Moreover, feelings of empathy cannot so easily be used to argue that the resulting moral intuitions are in some way faulty (as with Sinnott-Armstrong and Greene). Second, such studies do not establish the stronger claim that emotions can influence moral judgments concerning actions we otherwise don't consider morally wrong, which is an especially crucial claim for versions of sentimentalism akin to Prinz's. Thus, Greene and others rightly acknowledge the importance of the disgust experiments on which we've focused [cf. also Sinnott-Armstrong 2006: 203–4]. Without them, the metaethical arguments rest on a much weaker foundation.

In any event, insofar as we have evidence that disgust slightly affects the severity (but not necessarily polarity) of a certain class of moral judgments, the data do show *something* and they are certainly not *incompatible* with various metaethical theories. However, mere compatibility doesn't allow us to provide any support for, say, sentimentalism that will count against its competitors. Rationalists could argue that emotions like disgust merely draw our attention to morally relevant features of the situation. This may not seem like a plausible approach in some cases, however, similar as it is to thinking there is wisdom in mere repugnance. Consider Wheatley and Haidt's student council case. Given that it is a morally neutral scenario, there doesn't seem to be anything to which a disgust reaction can point.

But there are other explanations available. A more relevant issue than rationalism versus sentimentalism is the function of moral judgment—whether it expresses a cognitive state or not. Non-cognitivists have tended to maintain that

moral judgments *never* express propositions (to put it semantically) or cognitive mental states whose function is to accurately represent (to put it psychologically). Yet both rationalists and sentimentalists could agree that *a* function of moral judgment is non-cognitive, and particularly emotive—to express the emotions we feel. As Alasdair MacIntyre once put it:

Those emotive theorists who said that the function of moral utterance was to evince emotion would... have been correct if they had substituted the indefinite for the definite article. [1989: 101]

Such ‘hybrid theories’ are becoming more common among metaethicists [e.g. Copp 2001]. The slight effect disgust has on moral judgments might plausibly be the result of this emotive function coming to the fore, which is independent of the corresponding moral belief expressed.¹¹

I take it this is an antecedently plausible hypothesis about moral judgment, as it captures two long-standing desiderata. First, moral judgments seem intimately connected with motivation and our feelings toward certain actions. Simply reflecting on our ordinary practices suggests that we often signal our emotional attitudes with our moral judgments. Likewise, if moral judgments can express emotional states, they may be able to account for the apparently intimate connection between judging an action wrong, say, and being averse to it. Second, moral judgments seem to express claims that can be true or false. After all, we do vigorously argue about morality, and this only seems to make sense if moral judgments can sometimes express attitudes that can be assessed for their truth or falsity.

If we are to use the results from the disgust experiments to address a metaethical debate, I suggest they are better directed here. After all, if one of the functions of moral judgment is to express relevant emotional states, then we might predict that participants with elevated emotions will accordingly register elevated moral judgments. However, the evidence suggests that the descriptive function of moral judgment is more prominent than any expressive one. Again the Student Council case is instructive. Disgusted participants were not apparently moved to condemn the morally neutral action of picking topics that would appeal to others. Instead, the average response was slightly shifted toward the negative side of an extremely fine-grained scale. The same idea applies as well in the cases involving a morally suspect act: whether disgusted or not, participants were apparently inclined to provide the expected moral condemnation, although disgust slightly elevated this. So, while we may sometimes use moral judgments to indicate our emotional or other non-cognitive states, the polarity of ordinary people’s moral judgments is apparently driven more by one’s perception of whatever facts they are meant to describe. We are primarily moved to ascribe the property of rightness or wrongness with our judgments, but to some extent we express certain emotional states. In the end, the empirical results under discussion can provide some evidence for this philosophical debate, not disputes about sentimentalism and the warrant of moral judgments.

¹¹ Many thanks to Nic Southwood for discussion of these issues.

6. Conclusion

We have uncovered an array of issues with the experiments widely cited in support of the strong claim that disgust substantially influences a range of moral judgments. However, the evidence does support the weaker claim that disgust slightly influences the severity (but not necessarily polarity) of a relevant (but narrow) class of moral judgments. I have argued that this weak claim does not support the various arguments philosophers have tended to make with reference to this research. We especially lack any support for the claim that the experience of an emotion alone is necessary or sufficient for generating a moral judgment or altering its polarity.

However, the numerous experiments do support a conclusion, albeit weaker, on an issue which has garnered less attention in this arena. I suggest we take this evidence as confirming the idea that moral judgments have both an expressive and descriptive function, although the latter is more substantial than the former. This interpretation of the evidence predicts what has been consistently observed across all these experiments: an emotion can sometimes slightly influence a specific class of moral judgments in a domain related to that emotion.¹²

¹² For comments on drafts of this paper, I thank C. Daniel Batson, Wesley Buckwalter, Zachary Horne, Andy Lamey, Walter Sinnott-Armstrong, Aaron Zimmerman, and the anonymous reviewers for this journal. Many thanks also to audiences at Monash University, the Australian National University, Macquarie University, and the Australasian Association of Philosophy conference. I received exceptional feedback from so many, but especially Trevor Case, David Chalmers, Ben Fraser, Catriona Mackenzie, Jonathan McGuire, John Maier, Christina Majoinen, Daniel Nolan, and Nicholas Southwood.

References

- Case, T. I., Oaten, M. J., and R. J. Stevenson 2012. Disgust and Moral Judgment, in *Emotions, Imagination and Moral Reasoning*, ed. C. Mackenzie and R. Langdon, New York: Psychology Press: 191–214.
- Copp, David 2001. Realist-Expressivism: A Neglected Option for Moral Realism, *Social Philosophy and Policy* 18/2:1–43.
- Cushman, F., Young, L. and J. Greene 2010. Multi-System Moral Psychology, in *The Moral Psychology Handbook*, ed. J. M. Doris and The Moral Psychology Research Group, New York: Oxford University Press: 47–71.
- Eskine, Kendall J., Kacinik, N. A., J. J. Prinz 2011. A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment, *Psychological Science* 22/3: 295–299.
- Greene, Joshua 2008. The Secret Joke of Kant’s Soul, in *Moral Psychology, Vol. 3*, ed. W. Sinnott-Armstrong, Cambridge, MA: MIT Press: 35–117.
- Haidt, Jonathan 2001. The Emotional Dog and Its Rational Tail, *Psychological Review* 108/4: 814–834.
- Haidt, J. and F. Bjorklund 2008. Social Intuitionists Answer Six Questions About Morality, in *Moral Psychology Vol. 2*, ed. W. Sinnott-Armstrong, Cambridge, MA: MIT Press: 181–218.
- Haidt, J., Koller, S. H., and M. G. Dias 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65/4: 613–628.
- Horberg, E. J., Oveis, C., Keltner, D., and A. B. Cohen 2009. Disgust and the Moralization of Purity, *Journal of Personality and Social Psychology* 97/6, 963–976.
- Huebner, Bryce, Dwyer, Susan, and Marc Hauser 2009. The Role of Emotion in Moral Psychology, *Trends in Cognitive Sciences* 13/1: 1–6.
- Inbar, Y., Pizarro, D. A., Knobe, J. and P. Bloom 2009. Disgust Sensitivity Predicts Intuitive Disapproval of Gays, *Emotion* 9/3: 435–43.
- Joyce, Richard 2006. *The Evolution of Morality*, Cambridge, MA: MIT Press.
- Kelly, Daniel 2011. *Yuck!: The Nature and Moral Significance of Disgust*, Cambridge, MA: MIT Press.
- Kennett, Jeanette 2012. Living with one’s choices: Moral judgment in vivo, in *Emotions, Imagination and Moral Reasoning*, ed. C. Mackenzie and R. Langdon, New York: Psychology Press.
- Laham S., Alter A., G. Goodwin 2009. Easy on the Mind, Easy on the Wrongdoer: Discrepantly Easy to Process Violations are Deemed Less Morally Wrong, *Cognition* 112/3:462–466.
- MacIntyre, Alasdair 1989. What Morality Is Not, in his *Against the Self-Images Of the Age*, Notre Dame, IN: University of Notre Dame Press.
- Mallon, Ron and Shaun Nichols 2010. Rules, in *The Moral Psychology Handbook*, ed. J. M. Doris and The Moral Psychology Research Group, New York: Oxford University Press: 297–320.
- May, Joshua and Richard Holton 2012. What in the World Is Weakness of Will? *Philosophical Studies* 157/3:341–360.
- McGuire, J., Langdon, R., Coltheart, M., C. Mackenzie 2009. A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research, *Journal of Experimental Social Psychology* 45/3: 577–580.
- Nichols, Shaun 2002. Norms with Feeling: Towards a Psychological Account of Moral Judgment, *Cognition* 84/2: 221–236.

- Pettit, Dean and Joshua Knobe 2009. The Pervasive Impact of Moral Judgment, *Mind and Language* 24/5:586–604.
- Pizarro, D., Inbar, Y., and C. Helion 2011. On Disgust and Moral Judgment, *Emotion Review* 3/3: 267–268.
- Plakias, Alexandra 2013. The Good and the Gross, *Ethical Theory and Moral Practice* 16/2: 261–278.
- Prinz, Jesse 2006. The Emotional Basis of Moral Judgments, *Philosophical Explorations* 9/1: 29–43.
- Prinz, Jesse 2007. *The Emotional Construction of Morals*, Oxford University Press.
- Rozin, P., Lowery, L., Imada, S., J. Hadit 1999. The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity), *Journal of Personality and Social Psychology* 76/4: 574–86.
- Schnall, S. Benton, J., and S. Harvey 2008. With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments, *Psychological Science* 19/12: 1219–1222.
- Schnall, S., Haidt, J., Clore, G. L., and A. H. Jordan 2008. Disgust as Embodied Moral Judgment, *Personality and Social Psychology Bulletin* 34/8: 1096–1109.
- Sinnott-Armstrong, Walter 2006. *Moral Skepticisms*, New York: Oxford University Press.
- Wheatley, Thalia and Jonathan Haidt 2005. Hypnotic Disgust Makes Moral Judgments More Severe, *Psychological Science* 16/10: 780–784.
- Zhong, C. B., B. Strejcek and N. Sivanathan 2010. A Clean Self Can Render Harsh Moral Judgment, *Journal of Experimental Social Psychology* 46/5: 859–862.