

Does incidental disgust amplify moral judgment?

A meta-analytic review of experimental evidence

Justin F. Landy and Geoffrey P. Goodwin

University of Pennsylvania, Department of Psychology

IN PRESS AT *PERSPECTIVES ON PSYCHOLOGICAL SCIENCE*

### Abstract

The role of emotion in moral judgment is currently a topic of much debate in moral psychology. One specific claim made by many researchers is that irrelevant feelings of disgust can amplify the severity of moral condemnation. Numerous studies have found this effect, but there have also been several published failures to replicate this effect. Clarifying this issue would inform important theoretical debates between rival accounts of moral judgment. We meta-analyzed all available studies, published and unpublished, that experimentally manipulated incidental disgust prior to or concurrent with a moral judgment task ( $k = 50$ ). We found that there is evidence for a small amplification effect of disgust ( $d = .11$ ), which is strongest for gustatory/olfactory modes of disgust induction. However, there is also some suggestion of publication bias in this literature, and when this is accounted for, the effect disappears entirely ( $d = -.01$ ). Moreover, prevalent confounds mean that the effect size that we estimate is best interpreted as an upper bound on the size of the amplification effect. The results of this meta-analysis argue against strong claims about the causal role of affect in moral judgment and suggest a need for new, more rigorous research on this topic.

The psychological study of moral judgment has undergone something of a revolution in the past 25 years. Prior to this time, Lawrence Kohlberg's pioneering theorizing dominated research in moral judgment. Kohlberg viewed moral judgment as resulting primarily from reasoned deliberation; "the moral force in personality is cognitive," he argued (Kohlberg, 1971 p. 230). Kohlberg did not deny the role of affect in moral judgment (see Kohlberg, 1971, pp. 230-231), but he saw it as considerably more peripheral than cognition, a general perspective that we will call "rationalist." Lately, however, there has been a paradigm shift within moral psychology, such that many theorists now emphasize the role of automatic emotional processes in moral judgment, largely downplaying the role of reasoned deliberation. For instance, Haidt's (2001) highly influential Social Intuitionist Model posits that "moral intuitions (including moral *emotions*) come first and *directly cause* moral judgments," (p. 814, emphasis added). Moral reasoning, insofar as it occurs at all, is seen as post-hoc, justifying a judgment that has already been made.<sup>1</sup> According to this model, the reason why a person judges a proscribed act such as incest as morally wrong is that "one feels a quick flash of revulsion at the thought of incest, and one knows intuitively that something is wrong (p. 814)." We will refer to the general class of theories that posit a causal role of affect in moral judgment as "neo-sentimentalist," because the scholars who argue for them position themselves as the intellectual (and empirical) descendants of sentimentalist philosophers like David Hume (2003; see, e.g., Haidt, 2001, p. 816).

Both psychologists (e.g., Nabi, 2002; Pizarro & Bloom, 2004) and philosophers (e.g., Fine, 2006; Saltzstein & Kasachkoff, 2004) have challenged the neo-sentimentalist approach on theoretical grounds, and some researchers (e.g., Royzman, Goodwin, & Leeman, 2011; Royzman, Leeman, & Baron, 2009) have argued for a renewed emphasis on rationalist (or, more broadly, cognitive) inputs to moral judgment. The divide between rationalists and neo-

sentimentalists arguably constitutes the major theoretical debate in modern moral judgment research (although cf. Monin, Pizarro, & Beer, 2007). This debate is important because it cuts at the heart of the human moral faculty. The competing accounts of human morality offered by the neo-sentimentalist and rationalist perspectives are strikingly different, and give rise to very different conceptions of the origin, development, and malleability of human moral judgment. Advancing this debate is therefore of considerable interest to psychologists. The present meta-analysis aims to contribute to this debate by examining the evidence for one key claim of the neo-sentimentalist approach – that unrelated physiological disgust makes judgments of moral transgressions harsher than they otherwise would be.

### **Why Focus on Disgust?**

The emotion of disgust has received particular attention for its putative role in moral judgment. An early indication that disgust might play a role in moral judgment emerged from a seminal study in which people were found to treat disgusting, taboo acts that do not cause any harm – for example, having sexual intercourse with a dead chicken before cooking and eating it – as moral transgressions, rather than as violations of social convention (Haidt, Koller, & Dias, 1993). This finding contradicted one of the foundational assumptions of Turiel's (1983) social domain theory, according to which causing harm is a necessary and signature element of moral transgressions. This result showed that judgments of immorality can occur without the presence of harm, and was therefore seen as a major challenge for the prevailing rationalist models of moral thinking, and it sparked a considerable amount of subsequent research on the role of disgust in moral judgment.

Disgust is, at first glance, an odd candidate for an emotion that would influence moral judgment, because, as reviewed below, it is usually elicited by things that are contaminating but

have little to do with morality – things like feces, insects, and decaying corpses. In other words, disgust is “pre-normative” – that is, it can be evoked in the absence of a violated moral prohibition (Royzman, Leeman, & Baron, 2009). In this way, it differs from other negative emotions such as anger, which is more typically evoked in response to the appraisal of some offense or transgression, thus giving it an in-built normative character. However, this pre-normative feature of disgust is precisely what affords it a major advantage when examining the role of emotion in moral judgment: it allows a cleaner test of the “pure” role of emotion. If disgust, in the absence of moral violation, were found to influence moral judgments, this would provide clearer evidence for such a “pure” role of emotion than would demonstrating that a normatively-laden emotion like anger influences moral judgment. Studying disgust therefore offers the possibility of de-confounding affective and normative elements, thus potentially allowing clearer inferences about the role of affect *per se* on moral judgment.

### **What is Disgust?**

Before examining what role, if any, disgust plays in moral judgment, we briefly review what disgust is and what it does. Disgust is a withdrawal-motivating emotion, prompting avoidance of its eliciting stimulus (Ekman & Friesen, 1975). It is often accompanied by the sensation of nausea. Although there are divergent theories of the original purpose and function of disgust, there is considerable agreement on what sorts of stimuli typically elicit it: bodily fluids and waste, decay, filth, certain animals such as insects and rats, and certain classes of sexual acts (Oaten, Stevenson, & Case, 2009; Rozin, Haidt, & McCauley, 2008; Tybur, Lieberman, & Griskevicius, 2009; Tybur, Lieberman, Kurzban, & DeScioli, 2013).

Two prominent theories provide accounts of the origin and function of disgust. In both of these theories, disgust is seen as a multifaceted construct, involving a core disgust response,

which can then be applied to stimuli that do not intrinsically elicit disgust, such as moral offenses. The oral rejection theory of disgust (Rozin, Haidt, & Fincher, 2009; Rozin, et al. 2008) posits that disgust has its roots in distaste, the rejection of bitter substances in the mouth. Bitter taste, which is an indicator of plant poisons, evokes a physiological rejection response, the “gape” response, meant to expel the offending taste object. This response is present from birth, and has analogs in other mammals, such as rats (Travers & Norgren, 1986), which strongly suggests that it is an evolved response that protects the body from harmful or poisonous substances. Proponents of the oral rejection theory argue that this disgust response was co-opted to protect humans from other contaminating threats beyond toxins, including typical core disgust elicitors such as feces, insects, wounds, and the like, but also people who commit immoral acts. These immoral acts need not involve any core disgust elicitors, though of course they could. On this theory, moral disgust is a preadaptation – the disgust response being expanded to serve functions for which it did not originally evolve (for a much more in-depth treatment of this theory, see Rozin et al., 2008, and for a critique of it, see Tybur et al., 2013).

The other major theory of disgust posits that the disgust response does not have its roots in distaste, but rather that it evolved specifically to help us avoid disease. On this theory, disgust is elicited by sensory cues that indicate the probable presence of contagious pathogens (Oaten et al., 2009). Accordingly, activities that pose direct disease threats elicit the physiological disgust response (i.e., nausea, gagging, loss of appetite; see Royzman, Leeman, & Sabini, 2008), and are moralized. However, these disgust-eliciting stimuli do not have to be physically present to produce a disgust response – merely thinking about them is sufficient. Indeed, on this theory, the fact that mere thoughts of disgust elicitors can produce disgust is central to how disgust becomes broadly involved in moral judgment. Consequently, other activities that themselves do not pose

direct disease threats, for instance, deviant sexual practices, can become moralized, because they bring to mind thoughts of concrete disgust-eliciting stimuli via mental association. For example, necrophilia, which may in some instances consist only of sexual attraction to corpses, or of non-harmful contact with those corpses, evokes disgust and moral condemnation because it brings to mind the thought of dangerous contact with (potentially diseased) corpses.

A variant of this theory says that the disgust response originally evolved to motivate pathogen avoidance but was later co-opted to motivate avoidance of sexual partners with low reproductive value, and, separately, to coordinate punishment of those who violate moral norms within groups (Tybur et al., 2013). On this theory, disgust can influence moral judgment because it is a cue to the fitness costs of engaging in an action. Therefore, when one feels disgust, one concludes that an act is wrong and therefore ought to be punished (Tybur et al., 2013). Regardless of the specifics, proponents of this theory generally agree that disgust has its roots in an adaptation that helps us to avoid potential pathogens, but can be co-opted to support condemnation in the moral domain.

Both of these theories -- oral rejection and pathogen avoidance -- provide an account of how disgust may become involved in moral judgment. And, as will be seen presently, they are relevant to understanding the existing literature on disgust and moral judgment, particularly when considering potential moderators of existing findings.

### **How Might Disgust Be Involved in Moral Judgment?**

Theorists of morality have also considered how disgust might be relevant to moral judgment. Pizarro, Inbar, & Helion (2011) articulated three ways that disgust could relate to moral judgment (for a similar treatment concerned with negative affect in general, see Prinz, 2006). The first is that disgust results from at least some moral judgments -- that is, when we

judge something to be a moral transgression, this evokes disgust, an idea we will refer to as the “elicitation hypothesis.” As indicated above, this idea is endorsed by both the oral rejection and the pathogen avoidance theories of disgust. The second possibility reverses this causal direction: the experience of disgust might make our moral appraisals more negative than they otherwise would be, a more controversial claim that we will call the “amplification hypothesis.” Third, disgust might *moralize* actions. This moralization hypothesis, which is the most controversial of all, states that experienced disgust is, by itself, enough to produce moral condemnation of otherwise neutral acts – that is, disgust is *sufficient* for moral condemnation. These three hypotheses are summarized schematically in Figure 1.

**[Insert Figure 1 about here.]**

### **Evidence for the Elicitation Hypothesis**

Before examining whether disgust exerts a causal influence on moral judgments, we first consider whether seeing or contemplating moral transgressions can produce disgust. If the elicitation hypothesis is not correct, then the stronger claims of amplification or moralization would seem unlikely, because these claims presuppose the regular co-occurrence of disgust and moral judgment. That is, on both of these stronger hypotheses, the reason why disgust influences moral judgment is either because it is seen as a reliable cue to the moral gravity of an offense (akin to the affect-as-information approach, see Schwarz & Clore, 1988), or because disgust is partly constitutive of moral disapprobation itself (see, e.g., Haidt, 2001; Prinz, 2006). Either way, one would expect disgust regularly to be elicited by appraisals of moral wrongness.

Three sources of evidence support the elicitation hypothesis. First, in response to at least some moral violations, individuals explicitly report feeling “disgusted”. Some studies suggest that this effect is limited to so-called “purity violations” – proscribed activities that involve the



defilement or degradation of a person's body or soul, including inappropriate sex acts, food taboos, misuse of the body, and crimes against nature (e.g., Haidt & Graham, 2007; Horberg, Oveis, Keltner, & Cohen, 2009, Study 1). Other studies have found this effect to extend to a range of moral offenses (Hutcherson & Gross, 2011), although this latter finding may have arisen because participants reported feeling "morally disgusted", rather than "disgusted", per se (see Russell, Piazza, & Giner-Sorolla, 2013).

Second, moral violations also appear to activate the concept of disgust implicitly. Having read about a moral violation, participants were more likely to complete ambiguous word stems such as REVOL\_ING with disgust related words (e.g., REVOLTING) rather than disgust-unrelated words (e.g., REVOLVING) (Jones & Fitness, 2008). An ambiguity with this result, and others like it, is that in lay parlance the word "disgust" appears to capture elements of anger as well as core disgust (see e.g., Nabi, 2002; Russell & Fehr, 1994), so it remains possible that the disgust evoked by moral offenses is primarily rooted in anger rather than core disgust. However, participants who were offered a gift to take home after the experiment were more likely to take a cleaning-related product when they had read about a moral violation, suggesting that they experienced core disgust which they then desired to cleanse (Jones & Fitness, 2008, Study 1).

Third, moral transgressions, such as unfair offers in the ultimatum game, appear to activate facial expressions that are characteristic of core disgust rather than anger (Chapman, Kim, Susskind, & Anderson, 2009), which in turn predict the likelihood of rejecting such offers. Facial responses indicating disgust appear particularly prevalent in response to purity violations and violations of fairness norms, rather than actions that are directly harmful (Cannon, Schnall, & White, 2011). Whether such findings show that core disgust is directly activated by moral

transgressions is disputed, however: Royzman & Kurzban (2011a, b) argue that facial expressions of this sort may constitute intentional signals of disapproval rather than direct readouts of emotional responses (though see Chapman & Anderson, 2011 for an alternative perspective).

In sum, there is accumulating, though imperfect, evidence that disgust can be elicited by an appraisal that a moral transgression has occurred, and some evidence suggests that this effect is more robust for purity offenses. This evidence thus establishes an important precondition for the amplification and moralization hypotheses. We now turn to evaluating the evidence for the primary focus of this paper: the amplification hypothesis. There is considerable evidence that the experience of disgust can increase the severity of moral condemnation, as well as some conflicting evidence.

### **Evidence for the Amplification Hypothesis**

#### **Correlational evidence**

Although the present meta-analysis is concerned with experimental studies of the amplification effect, relevant correlational results also exist, which are consistent with claims of a causal link between disgust and moral judgment. Some studies have shown that individuals' trait sensitivity to disgust positively predicts the severity of their moral and punitive judgments (Chapman & Anderson, 2014; Jones & Fitness, 2008, but see Landy & Piazza, 2014), especially those regarding violations of bodily and sexual purity (Horberg et al., 2009), and that it also predicts conservative attitudes towards abortion and gay marriage (Inbar, Pizarro, & Bloom, 2009).

These correlational results are consistent with a dispositional form of the amplification hypothesis, at least in the domain of purity violations. However, they each leave open the

possibility of reverse causation – it could be that the harshness of people’s moral judgments causes disgust sensitivity, rather than the other way around. We are exposed to many disgust elicitors – for instance, body odor or dog feces – only when someone else breaches a social norm – for instance, by failing to maintain standards of personal hygiene or to clean up after their pet. People who are generally more severe in their moral judgments might be especially attuned to such hygiene-related offenses – both noticing and attending to such offenses more assiduously, as well as judging them more severely – thereby experiencing an elevated level of disgust in their daily lives as a result. Moreover, the reviewed correlational evidence leaves open any number of third variable explanations for the association between trait disgust and moral harshness. To name just one, it is known that political conservatism predicts disgust sensitivity (Inbar, Pizarro, & Bloom, 2009) – particularly in the domain of sexual disgust (Tybur, Merriman, Caldwell Hooper, McDonald, & Navarrete, 2010), and that conservatives also tend to be more punitive in their moral thinking than liberals, at least under certain conditions (Gromet & Darley, 2011; Tetlock et al., 2007). The association between overall disgust sensitivity and moral harshness could therefore be driven by political ideology predicting both variables independently.

### **Experimental evidence**

Because experimental designs can rule out these alternative explanations and provide direct evidence for a causal effect of disgust on moral judgment, we focus on these designs in the present meta-analysis. The amplification hypothesis is best tested in experiments in which disgust is manipulated using an emotion induction that is unrelated – or *incidental* (Bodenhausen, 1993; Loewenstein & Lerner, 2003) – to the moral judgment. Such designs can show that the mere subjective experience of physiological disgust amplifies moral judgments.

Indeed, studies of incidental disgust arguably provide the best current evidence for the neo-sentimentalist idea that moral judgment is largely driven by emotion.

Several such studies have found that incidental disgust amplifies moral judgment. In the earliest published study of this kind, participants were hypnotized to feel “a brief pang of disgust... a sickening feeling in your stomach” when reading an otherwise innocuous word (*take* or *often*). Participants then read vignettes describing moral transgressions, half of which contained the hypnotic trigger word. Participants rated behaviors described in vignettes that included the trigger word as more immoral than identical behaviors described in vignettes that did not include the trigger word (Wheatley & Haidt, 2005).

Since this study, several other methods of inducing disgust have been used, many of which were pioneered by Schnall, Haidt, Clore, and Jordan (2008). In one study, participants exposed to a noxious ambient odor (a commercially available novelty “fart spray”) rated a variety of transgressions as more immoral than participants not exposed to any odor. Similarly, in other studies, participants who completed a study in a filthy work area, or who wrote about a time they had encountered something physically disgusting, or who had just watched a disgusting film clip (the infamous toilet scene from the movie *Trainspotting*), all rated various moral transgressions as more wrong than control participants. These latter three results were observed only among participants high in Private Body Consciousness (PBC; Miller, Murphy, & Buss, 1981), a dispositional tendency to attend to one’s internal bodily states (Schnall, Haidt, et al. 2008). Using the same video induction, less harsh moral judgments were observed among participants who had washed their hands after the disgust induction than participants who had not – seemingly because the hand-washing alleviated the felt disgust, thereby eliminating the amplification effect (Schnall, Benton, & Harvey, 2008). Amplification effects in moral

judgment have been replicated using other videos (Ugazio, Lamm, & Singer, 2012) and even bitter tastes (Eskine, Kacinik, & Prinz, 2011) as disgust inductions. Other researchers have replicated the amplification effect of disgust (Cheng, Ottati, & Price, 2013), while also observing similar amplifying effects for several different arousing emotions beyond disgust.

Several other studies have observed the amplification effect, but only in a moderated way. For instance, a study using the same video induction as Schnall, Haidt, et al. (2008), replicated the amplification effect, but only for purity violations; directly harmful actions were unaffected (Horberg et al., 2009). Similarly, when participants listened to the sound of a person vomiting while making moral judgments, only judgments of the wrongness of purity offenses were amplified (Seidel & Prinz, 2013). In contrast, anger – induced by an irritating sound – amplified judgments of the wrongness of harmful actions and violations of fairness norms, but did not affect judgments of purity violations. In another partial replication, the amplification effect was found to hold only among participants low in attentional control (Van Dillen, van der Waal, & van den Bos, 2012). It has also been found to replicate only among participants who were unskilled at differentiating between their emotional states (Cameron, Payne, & Doris, 2013). Lastly, Japanese participants who were high in mindfulness – specifically, those who described themselves as aware and “present” in the moment – did not show the amplification effect, but low-mindfulness participants did (Sato & Sugiura, 2014).

Finally, other investigations have used more behavioral measures of moral judgment to investigate the amplification effect. In one study, participants who had viewed disgusting video clips were more likely to reject unfair offers in an ultimatum game – a form of costly punishment (Harlé & Sanfey, 2010). Though this is a behavioral response, it presumably depends on moral judgment to some degree, and is thus pertinent to our investigation. Similarly, a study using

disgusting images as an induction showed that disgusted participants were more likely to reject unfair offers in the ultimatum game, but only when participants believed that they were playing with human partners; when they believed that they were playing with computer programs, disgust had no effect (Moretti & di Pellegrino, 2010).

### **Countervailing evidence to amplification**

Thus, there is considerable evidence from the published literature supporting the amplification hypothesis. Nonetheless, there have also been several published failures to replicate these results. One such failure came from a pilot study mentioned by Schnall, Haidt et al. (2008), in which participants immersed their arms in a gooey, disgusting substance prior to making moral judgments. No effect of this disgust induction on moral judgment was observed. A failure to replicate the amplification effect when using a fart spray as a disgust induction has also been reported (Ugazio, Lamm, & Singer, 2012). Moreover, a direct replication of Schnall, Benton, and Harvey's (2008) hand-washing study failed to find any difference in judgments of moral severity between the hand-washing condition and the no-washing condition, despite having .99 power to detect such an effect (Johnson, Cheung, & Donnellan, 2014). In a separate investigation comprising two independent studies, participants saw disgusting images, neutral images, or negative but non-disgusting images prior to a moral judgment task. The judgments of participants in the disgust condition did not differ from those of participants in the other two conditions (Case, Oaten, & Stevenson, 2012). Another study used an evaluative conditioning procedure to condition participants to feel disgust in response to an innocuous word (*part*). Participants then read several vignettes describing moral transgressions, half of which contained the conditioned word and half of which did not. Participants' ratings of the wrongness of the described transgressions did not differ as a function of the presence or absence of the conditioned

word (David & Olatunji, 2011). Lastly, when participants were subliminally primed with disgust facial expressions, they actually rated utilitarian harms – i.e., directly killing one person to save several others – as *less* morally wrong, contra Schnall, Haidt, et al., though this was moderated by disgust sensitivity – highly sensitive participants rated the harms as less wrong, while less sensitive participants rated them as more wrong (Ong, Mullette-Gillman, Kwok, & Lim, 2014).

In almost all of these cases, the authors report successful manipulation checks – that is, they did successfully induce feelings of disgust, and yet the experienced disgust did not reliably influence moral judgment. There were only two exceptions. First, Ong et al. (2014) found that disgust increased during the experiment in both their disgust-prime and neutral-prime conditions. Second, no manipulation check was reported in Schnall, Haidt et al.'s (2008) pilot study. However, the disgust induction used in that study (plunging one's hand into a gooey substance) was, *prima facie*, very potent and disgusting, so it is reasonable to assume that it induced disgust successfully. Indeed, the authors argue that this manipulation may have been *too* effective – that it was so salient a source of disgust that participants attributed all of their experienced disgust to the induction task and thus discounted it when making their moral judgments.<sup>2</sup>

In addition to these failures to replicate, other recent empirical results also call into question the role of disgust in moral judgment. When disgust is pitted against another moral emotion such as sympathy in a moral dilemma, the neo-sentimentalist account would predict that individuals' moral judgments should be driven by whichever emotion is felt most strongly. In a study testing this prediction, participants read dilemmatic vignettes in which one course of action required the protagonist to engage in a disgusting act (incest) in order to prevent a serious harm, while the other course of action was to refrain from the disgusting act at the cost of allowing serious harm (Royzman, Goodwin, & Leeman, 2011). Contrary to the neo-sentimentalist

prediction, participants' judgments of the morally right course of action were best predicted by their beliefs about which course of action would cause the least harm, broadly construed, rather than by their relative levels of subjective disgust and sympathy. A similar result showed that judgments about the immorality of a putatively harmless violation of social etiquette (spitting into a napkin at a dinner party) were better predicted by whether participants felt that "someone was negatively affected" by the action than by how disgusting they found it (Royzman et al., 2009).

In sum, these results suggest that the amplification hypothesis might not be supported as robustly as it might appear at first glance. A full examination of the available evidence should therefore yield a clearer picture of the overall support for the amplification hypothesis. There have only been a handful of studies examining the moralization hypothesis, but we examine the current evidence for this as well in our meta-analysis.

### **Overview of the Meta-Analysis**

Accordingly, we carried out a meta-analysis that examined experiments in which incidental disgust was induced in participants prior to or concurrent with their making moral judgments. Our aim was to estimate the magnitude of the amplifying effect of incidental disgust on moral judgment, through a survey of all relevant literature, both published and unpublished, thus informing the broader theoretical debate in moral psychology between the neo-sentimentalist and rationalist positions. If strong evidence for the amplification hypothesis were to be found, this would clearly support the neo-sentimentalist paradigm, but if only weak or no evidence were to be found, this would suggest that a rethinking of the neo-sentimentalist approach would be in order.



The meta-analysis is divided into four related sections. We first derive a point estimate for the size of the effect of incidental disgust on moral judgment from both published and unpublished studies (the amplification hypothesis). We then evaluate the published literature alone to see if there is any indication of publication bias. Next, we examine the role of two theoretically important moderator variables – type of disgust induction and type of violation. Finally, we compute a point estimate for the size of the effect of incidental disgust on judgments of non-moral actions as a test of the theoretically stronger moralization hypothesis.

### **Literature Search**

We began our search for relevant studies by consulting several online databases: ISI Web of Knowledge, PsycINFO, PsycARTICLES, and the Social Science Research Network. In each of these four databases, we conducted a full-text search using the intentionally broad search terms *disgust moral judgment*. We then narrowed down the search results based on the following criteria. To be included:

1. A study had to manipulate the presence of disgust experimentally such that the resulting disgust was ostensibly unrelated to the moral judgments – that is, the disgust had to be incidental.
2. A study's dependent variable(s) had to be some form of morally relevant judgment or action, such as ratings of immorality, assessments of character, recommendations for punishment, or actual punishment in economic games.<sup>3</sup>

It was usually possible to determine whether a paper contained at least one study that met these criteria by examining the abstract. When there was any ambiguity or doubt about a paper, we examined the method sections to determine whether any studies met the criteria for inclusion. We also requested any unpublished data from the corresponding authors of all of the published

studies and posted a call for unpublished data to several online forums (see Online Supplemental Materials for details). This complete literature search produced a final set of 33 papers, containing 51 relevant studies (31 published, 20 unpublished)<sup>4</sup> with a total  $N$  of 5,102 participants. A full list of all studies included in the meta-analysis, along with their publication statuses, moderator codings, sample sizes, and effect sizes, can be found in Table 1, and a forest plot presenting each study's effect size and 95% confidence interval can be found in Figure 2.

**[Insert Table 1 and Figure 2 about here.]**

Because readers do not have easy access to the unpublished studies included in the meta-analysis, we present brief descriptions of these studies and whether or not they used pre-tests, manipulation checks, or both to ensure that their disgust manipulation did, in fact, induce disgust in participants, in Supplemental Table 1 (viewable online). The vast majority of published and unpublished studies reported successful pre-tests or manipulation checks, so any failure among these studies to replicate the amplification effect cannot be attributed to their having failed to induce disgust successfully; only a handful reported unsuccessful manipulation checks (e.g., Schnall, Haidt et al., 2008, Study 2) or no such checks at all (e.g., Zhong, Strejcek, & Sivanathan, 2010).

### **Analysis Strategy**

#### **Obtaining Effect Sizes**

Because this meta-analysis was concerned with experimental studies, we converted all results in the meta-analysis to standardized mean difference scores (Lipsey & Wilson, 2001), better known as Cohen's  $d$ . In most cases,  $d$  was calculated from reported  $t$ -tests and one- $df$   $F$  tests, or from reported sample  $N$ s, means, and standard deviations or standard errors. When the necessary statistical information was not reported in a paper or manuscript, we contacted the

corresponding author and requested the information or the raw data necessary to calculate  $d$ . All effect sizes were scored such that positive numbers indicated support for the amplification hypothesis. That is, positive  $d$ s indicate higher levels of moral disapproval in the disgust condition(s) of a study than in the control condition(s), and negative  $d$ s indicate the opposite pattern. We selected the between-subjects  $d$  as our effect size primarily because the overwhelming majority of studies in the meta-analysis used between-subjects designs, so it simplified the analysis to calculate the between-subjects  $d$ . For within-subjects designs, we converted the repeated-measures effect size to a between-subjects  $d$  (Morris & DeShon, 2002).

We conducted all of the following analyses using both fixed- and random-effects models. We focus on the results of the random-effects models, as these models allow for variation in effect sizes between studies deriving from sources other than sampling variance. Given the wide range of methods used in the reviewed studies, the expected moderating effects of the variables mentioned above, and the observed heterogeneity of effect sizes (see below), it seemed more reasonable to assume a random-effects model. All reported analyses were conducted using the Comprehensive Meta-Analysis software package.

### **Point Estimate for the Size of the Amplification Effect**

Our first goal was to obtain a point estimate for the size of the amplification effect. To do so, we first computed one effect size for each study that tested this hypothesis ( $k = 50$ ). When studies included multiple types of moral judgment (e.g., Horberg et al., 2009), and the raw data were not available, a single effect size was calculated collapsing across the different moral judgment types, using reported one- $df$   $F$  tests of the effect of disgust. When a study included more than one control condition – for instance, a sadness condition and a no-emotion condition (e.g., Schnall, Haidt et al., 2008, Study 4) – we computed a single effect size comparing the

disgust condition to these two control conditions (Lipsey & Wilson, 2001). However, when a study included an additional condition of theoretical interest, such as a condition inducing another moral emotion like anger that was predicted to produce different results than disgust, this condition was not compared with the disgust condition (and was not collapsed with the control condition). It was generally clear whether a condition was meant to be a control condition or to serve some other theoretical purpose. Similarly, when a study contained multiple levels or variants of its disgust induction (e.g., Schnall, Haidt et al., 2008, Study 1), we computed a single effect size comparing these conditions to the control condition (Lipsey & Wilson, 2001). We did not include judgments of non-moral actions when computing each study's overall effect size. Before computing a mean effect size estimate, we applied the small-sample correction to all *ds* to obtain an unbiased effect size estimator (Hedges, 1981; Lipsey & Wilson, 2001).

We calculated a point estimate for the size of the amplification effect, as well as the 95% confidence interval for this point estimate. Specifically, we calculated a weighted mean of the effect sizes in which each effect size was weighted by the inverse of the sampling variance of the study from which it was calculated (Lipsey & Wilson, 2001). This point estimate for the effect size was statistically significant,  $d = .11$ ,  $p = .002$ , 95% confidence interval (CI): [.04, .19], but suggests a small effect (Cohen, 1992). There is thus evidence for the amplification hypothesis, but the effect appears to be fairly insubstantial. There was also significant heterogeneity among the calculated effect sizes,  $Q(49) = 143.66$ ,  $p < .001$ , so we assessed the moderating influence of two pre-selected variables below.

### **Assessing and Accounting for Publication Bias**

Our second goal was to assess the likelihood that there is publication bias in this literature. As an initial test, we recomputed the above effect size estimate using only studies in

the published literature. This estimate was somewhat larger than the above,  $d = .17$ ,  $p = .001$ , 95% CI: [.07, .27]. This indicates that the published studies tend to show larger effect sizes than the unpublished studies. This is perhaps not surprising. What is striking, however, is how small the effect sizes found by the unpublished studies tend to be. The weighted mean effect size calculated from the obtained unpublished results is effectively zero,  $d = .03$ ,  $p = .59$ , 95% CI: [- .09, .16]. The weighted mean effect sizes of published and unpublished studies were marginally significantly different by a random-effects inverse variance-weighted one-way analysis of variance (ANOVA, Lipsey & Wilson, 2001), the meta-analytic analog of the typical one-way ANOVA,  $Q(1, 48) = 2.89$ ,  $p = .09$ . In short, the published literature suggests a reliable, though small, effect, whereas the unpublished literature suggests no effect.

Figure 3 presents a funnel plot of the published effect sizes. In the absence of publication bias, the number of studies with high sampling error – i.e., studies toward the bottom of the plot – should be symmetrically distributed around the weighted mean effect size. Upon visual inspection, there appears to be a preponderance of low-powered studies on the right side of the graph. Consistent with this, Egger's regression coefficient is significant, Intercept = 1.15,  $t(29) = 3.20$ ,  $p = .003$ , indicating that studies with high sampling variance are associated with larger standardized effect sizes (effect size divided by sample standard error) than would be expected, suggesting bias in the dataset (Egger, et al., 1997).

**[Insert Figure 3 about here.]**

Given that there is suggestive evidence the published literature is unbalanced, the next question is, what becomes of the amplification effect when this is accounted for statistically? Duval and Tweedie's (2000) trim-and-fill procedure imputed 10 missing studies to produce the expected symmetric funnel plot, and estimated a much smaller, and non-significantly negative,

adjusted weighted mean effect size,  $d = -.01$ , 95% CI:  $[-.12, .10]$ .<sup>5</sup> Thus, the published literature appears to overestimate the extent to which incidental disgust amplifies moral judgments, and when this is corrected for, no significant effect is present.

### **Moderator Analyses**

Our third goal was to assess the effects of two theoretically relevant moderators. The first moderator is the sensory modality by which disgust was induced in participants. The oral rejection theory's conception of disgust as having originated as a gustatory response suggests that oral/nasal disgust inductions might be more potent sources of disgust than other sorts of induction, which would in turn exert stronger effects on moral judgments. The pathogen avoidance theory of disgust leads to a similar, though not identical, prediction, as it argues that "certain perceptual modalities (notably smell and touch) may be especially associated with disgust, perhaps because of more privileged access to the insular cortex and anteriomedial temporal lobe structures" (Oaten, et al., 2009, p. 313). Accordingly, we examined sensory modality as a potential moderator. We group olfactory inductions (i.e., ambient odors) and gustatory inductions (i.e., disgusting tastes) together because there is not a sufficient number of studies using such inductions to treat them as individual categories, and because the senses of smell and taste are closely related.

The second moderator that we examine is whether or not the violation being judged falls into the purity (or "divinity") domain. Rozin, Lowery, Haidt, & Imada. (1999) found that people tended to associate only purity violations (and not other sorts of moral violations) with disgust reactions (though see Royzman, Atanasov, Landy, Parks, & Gepty, 2014 for a counterpoint). This finding provides some reason to think that moral judgments of purity violations might be especially influenced by experimental manipulations of disgust. In such cases, experienced

disgust may be seen as particularly informative with respect to the gravity of the moral transgression (in line with the “affect-as-information” approach, Schwarz & Clore, 1983). And indeed, consistent with this speculation, some more recent work has reported evidence that purity violations are particularly susceptible to manipulations of disgust (see above, e.g., Horberg et al., 2009; Seidel & Prinz, 2013), although there is also some countervailing evidence (Schnall et al., 2008). There are also some theoretical reasons to expect a particularly strong amplification effect on purity violations. For instance, if purity/divinity violations produce more intrinsic (or ‘integral’) disgust than do other sorts of moral violation (see Haidt & Graham, 2007), participants may be more likely to misattribute the disgust produced by an incidental manipulation to the transgression itself, thus amplifying their judgments of its wrongness (see also Cameron et al., 2013). We therefore predicted that purity/divinity violations would show greater amplification effects than would non-purity violations.

We did not examine any individual difference measures that have been proposed as moderators of the amplification effect because such measures have been included in only a handful of studies (e.g., Private Body Consciousness: Johnson et al., 2013; Schnall, Haidt et al., 2008; attentional control: Van Dillen et al., 2012; emotional differentiation: Cameron et al., 2013; mindfulness: Sato & Sugiura, 2014), thereby rendering an analysis of them infeasible.

These moderator analyses required the creation of a new database of effect sizes that computed separate effect sizes within each study according to the *type of moral transgression* subjects judged (rather than computing a single effect size for each study, as in the above analyses), and which coded the effects according to the nature of the *sensory modality* used to induce disgust, which varied across rather than within studies – visual, gustatory/olfactory, imagined experience, or other (see Table 1). Regarding transgression type, each judged violation

was coded as either purity-related or non-purity-related. Calculating these effect sizes typically required obtaining additional statistical information or raw data from a paper's corresponding author, as many studies did not separately analyze purity-related and non-purity-related transgressions. For some studies, these data were unavailable, meaning that such studies could not be included in the moderator analyses (see Table 1). The nature of a transgression was usually obvious, though there were certain ones that were more difficult to categorize. The handful of transgressions that proved impossible to categorize (e.g., driving to work instead of walking [Schnall, Haidt, et al., 2008], which might be seen as either non-moral or as an offense against one's community, by virtue of polluting it, or even as a purity violation in that it destroys the purity of nature) were not included in the database.<sup>6</sup> Overall, this enlarged database contained a total of 73 effect sizes (43 published, 30 unpublished). As in the above analyses, the small-sample correction was applied to all of these effect sizes prior to analysis (Hedges, 1981; Lipsey & Wilson, 2001). The two moderators themselves were unrelated to one another,  $\chi^2(3) = .90, p = .83$ , which provides reassurance that any observed effects are not attributable to a lack of independence among them.

Because the two proposed moderators are categorical variables, analyses of both were conducted using random-effects inverse variance-weighted one-way ANOVAs (Lipsey & Wilson, 2001). The number of studies falling into each moderator category, and effect size estimates with 95% confidence intervals, are presented in Table 2.

**[Insert Table 2 about here.]**

There was a significant effect of the sensory modality of a study's disgust induction on the resulting effect size,  $Q(3, 69) = 8.18, p = .04$ . As predicted by both the oral rejection and disease avoidance theories, gustatory and olfactory disgust inductions produced much larger



amplification effects than did visual disgust inductions. Moreover, inductions that involved imagined experiences, reading about disgusting stimuli, or recalling disgusting events – which arguably constitute a single conceptual category of “imagined” or “mental” inductions – produced essentially no amplification effect. The heterogeneous “other” class of disgust inductions – consisting of hypnotic suggestions, evaluative conditioning, and auditory and tactile inductions – produced a very small weighted mean effect size, comparable to the effect size for imagined inductions.

The domain of the judged transgression (purity versus non-purity), did not explain a significant proportion of the observed heterogeneity in effect sizes,  $Q(1, 71) = 0.05, p = .82$ . Contrary to the predictions above, and also contrary to a handful of published results (e.g., Horberg, et al., 2009; Seidel & Prinz, 2013), the mean effect size for non-purity violations was, if anything, very slightly *larger* than the mean effect size for purity violations (see Table 2), suggesting that the amplification effect is not restricted to moral transgressions involving bodily or sexual purity, or crimes against nature.

### **Point Estimate for the Size of the Moralization Effect**

Our fourth and final goal was to assess the existing evidence for the moralization hypothesis – the idea that experienced disgust is itself sufficient to produce condemnation of typically amoral actions. To do so, we constructed a further database of effect sizes for judgments of non-moral actions after disgust inductions ( $k = 13$ ). We observed a significant moralization effect,  $d = .21, p = .01, 95\% \text{ CI: } [.05, .37]$ . This is again a rather small effect, and this result should be interpreted with caution because it is based on a small number of studies. Moreover, once again, the published studies ( $k = 6$ ) suggest a larger effect size,  $d = .33, p = .02, 95\% \text{ CI: } [.06, .61]$ , than do the unpublished studies ( $k = 7$ ),  $d = .14, p = .18, 95\% \text{ CI: } [-.06, .34]$ .

This difference was not significant,  $Q(1, 11) = 1.32, p = .25$ , probably owing to the small number of data points. Still, the unpublished studies show no significant effect, whereas the published studies show a small but reliable effect. There is therefore some preliminary evidence in support of the moralization hypothesis, but it is far from conclusive. Future research is needed to investigate this further.

### Discussion

What role, if any, does affect play in moral judgment? Researchers working within a neo-sentimentalist framework have recently argued that affect is a major driver, perhaps *the* major driver, of moral judgment. A key piece of evidence for this claim is that moral condemnation becomes harsher – that is, it is *amplified* – by negative affect that is unrelated to the moral issues at hand. Because of its pre-normative character, disgust is an ideal emotion for testing this “amplification hypothesis,” and it has been the emotion most studied by moral psychologists. And indeed, numerous studies have found evidence for the amplification hypothesis, but several other studies have failed to find evidence for it. Our goal was to clarify the current state of the evidence for this effect.

In order to evaluate the extant evidence for the amplification hypothesis, we meta-analyzed all available studies ( $k = 50$ ), published ( $k = 31$ ) and unpublished ( $k = 19$ ), that experimentally manipulated incidental disgust prior to or concurrent with a moral judgment task. The meta-analysis found a small but reliable amplifying effect of disgust,  $d = .11$ . However, we also found evidence suggesting publication bias in this literature. Indeed, even with the several published failures to replicate included, published studies produced a larger effect size than the overall estimated effect size,  $d = .17$ , while unpublished studies showed essentially no effect,  $d = -.03$ , and when missing studies were accounted for mathematically, no effect was observed,  $d = -$

.01. We also found preliminary evidence that the amplification effect is stronger for gustatory or olfactory disgust inductions, weaker for visual disgust inductions and essentially nonexistent for imagined or mental inductions. Lastly, we found evidence for a small but reliable effect of disgust on the severity of judgments of non-moral actions,  $d = .21$ .

### **Relation to the Affect and Morality Literature**

The existence of a reliable amplification effect may be problematic for hardline rationalist theorists who would deny any role for affect in moral judgment. But, it is not clear that such theorists exist. Even Kohlberg afforded some role for affect in his model of moral judgment, channeled through cognitive processes (see Kohlberg, 1971, p. 230-231). However, the relatively small effect size of the amplification effect seems to pose a conflict with some of the stronger claims made by neo-sentimentalist researchers. For instance, it has sometimes been claimed that affective intuition is the primary determinant of moral judgment, and that “for most people, most of the time, most of the action is in the quick, automatic, affective evaluations they make of people and events” (Schnall, Haidt et al., 2008, p. 1097). However, the estimated effect size of  $d = .11$  is equivalent to a correlation coefficient of  $r = .06$  (Rosenthal & DiMatteo, 2001), and explains only 0.3% of the variance in participants’ moral judgments, while the amplification effect disappears entirely when accounting for publication bias. These data therefore undermine what is arguably the best evidence for a causal role of affect in moral judgment.

A meta-analysis of this sort is not capable of decisively resolving the debate between neo-sentimentalists and rationalists. On the one hand, it does not (and could not) rule out the possibility that affect plays an important role in moral judgment. For instance, it could be that unmeasured affective reactions, separate from the manipulated disgust, are doing the most of the work in driving participants’ moral judgments (e.g., anger, see Russell & Giner-Sorolla, 2011).

Moreover, because we only examined evidence for the neo-sentimentalist position, our results cannot directly argue *for* the rationalist position or any other alternative theory. In particular, the present analysis also does not allow us to compare our estimate of variance explained by disgust with similar estimates for other processes that might play a role in moral judgment, including cognitive processes such as assessments of harm (see e.g., Royzman, Goodwin, & Leeman, 2011; Royzman, Leeman, & Baron, 2009). We therefore do not know how comparatively large or small the amount of variance explained by manipulations of disgust is

We did uncover evidence that gustatory and olfactory disgust inductions exert a reliable, small- to medium-sized effect on moral judgments. The reasons for this are not entirely clear. It could be that these are more potent or more direct ways of eliciting disgust, and therefore more effective in influencing moral judgment, as both the oral rejection and disease avoidance models seem to predict. Alternatively, these inductions might be less intrusive and less salient, and therefore less likely to be discounted by participants. This latter explanation seems unlikely, however. The largest effect size using a gustatory/olfactory induction (and the third largest effect size in the entire dataset) came from a study in which disgust was induced by having participants consume a bitter beverage prior to making their moral judgments (Eskine et al., 2011). The source of participants' disgust in this study is likely to have been quite salient to them ("I just drank that horrible, bitter liquid"), much more so than if their disgust had been produced by, say, an ambient odor. This suggests that the larger effect of gustatory/olfactory disgust inductions on moral judgment is unlikely to result from these inductions being less salient than visual and mental inductions.

One potential limitation of our meta-analysis concerns the size of the reviewed literature. Although our reviewed literature of  $k = 50$  studies might be considered small, we see two reasons

why the present meta-analysis is important notwithstanding this potential issue. First, studies purporting to show an amplification effect of disgust on moral judgment are widely cited and have informed a great deal of theorizing in moral psychology and even philosophy (e.g., Prinz, 2006). Second, widespread confounds throughout this literature call for more rigorous research on the topic of how disgust affects moral judgment, and it seems preferable to examine the extant evidence critically now, rather than wait for more studies with similar confounds to be conducted. We turn now to the issue of these confounds.

### **The Confounding of Disgust and Disapproval**

There is a general methodological problem with most studies in this literature that warrants further attention. Some researchers have noted that disgust inductions that involve observing (or imagining) other people engaging in disgusting behaviors might elicit moral disapproval directed at these people (Case et al., 2012). Other researchers have suggested that certain disgust inductions could evoke moral disapproval directed at experimenters or others in the environment (Baron Royzman, & Goodwin, 2013), a criticism that we think applies quite broadly.

To illustrate the first problem, a commonly used disgust induction is a scene from the film *Trainspotting*, in which a drug addict digs through a filthy public toilet to retrieve his drugs (Horberg et al., 2009, Study 2; Johnson et al., 2014, Study 2; Schnall, Benton, & Harvey, 2008, Study 2; Schnall, Haidt et al., 2008, Study 4; see Case et al., 2012). This scene clearly elicits core disgust through its prominent and graphic depiction of human bodily waste, but it may also elicit moral disapproval directed at the drug addict himself. This moral disapproval arguably confounds the manipulation of disgust, and makes it difficult to establish whether it is the disgust, *per se*, or the moral disapproval that is doing the work in producing subsequent

amplification effects on moral judgment. By comparison, an example of an induction that does not include a person acting, and which is therefore free of this sort of moral “contamination,” is the presentation of images from the International Affective Picture System (IAPS) depicting rotting animal corpses, insects in contact with food, and similar nauseating content (Case et al., 2012).

It may also be that by intentionally exposing their participants to disgusting, inappropriate stimuli, the *experimenters* in these studies are also seen as breaching social norms, thereby inadvertently making themselves the target of participants’ moral disapproval. This activated moral disapproval may then prime greater disapprobation in the subsequent moral judgment task. A clear instance of this problem is the “dirty desk” manipulation (Schnall, Haidt et al., 2008, Study 2). Participants were seated at a greasy and disgusting work space while filling out their questionnaires. They may well have felt angry and disapproving of the fact that the experimenter had been derelict in their duty to provide a clean environment for them (Baron et al., 2013).

This general problem is, we think, much more pervasive than has been previously recognized. Experimenters in disgust induction studies almost invariably expose participants to stimuli that are clearly going to evoke a negative emotional response and that may induce the unpleasant physiological experience of nausea. Moreover, they do so intentionally and without prior warning. Arguably, therefore, the experimenters are knowingly doing some small harm to their participants, who may feel that they have the right not to be exposed to such things, or to at least be given full warning that they will be so exposed. Thus, this potential problem applies in a different and more subtle way to almost all of the disgust inductions that have been used in past research, including those involving disgusting videos and images.

To address it, an important task for future research is to devise new experimental procedures that reliably induce core disgust in ways that do not breach social norms of appropriateness – a task that will require considerable ingenuity. Furthermore, we think that a more thorough approach is needed during the debriefing stages of future experiments, in order to examine the extent to which participants feel irritation or anger in response to the behaviors depicted in or implied by the primes, or towards the experimenter directly. In particular, it would be important to examine the extent to which both the primes and the experimenter's behavior elicit moral disapproval or produce affront of some kind.

Thus, we consider our estimated effect size of  $d = .11$  to be a plausible upper bound on the size of the amplification effect, because confounds may have inflated the effect sizes observed in many of the reviewed studies. However, even if the amplification hypothesis were to be examined in a way that is free of such confounds, the present meta-analysis suggests that it may be infeasible to do so because a study of this sort would require a very large sample to be sufficiently well powered.<sup>7</sup> Furthermore, this suggests that studies purporting to show the amplification effect with small sample sizes should be approached with skepticism, as they are likely underpowered.

### **The Status of the Moralization Hypothesis**

Although we have focused primarily on the amplification hypothesis in this meta-analysis, some of the relevant evidence also speaks to the stronger moralization hypothesis. The moralization hypothesis asserts that actions that are morally irrelevant can be judged to be morally wrong in the presence of disgust. Although only a handful of studies have included judgments of such “non-moral” actions, we found some preliminary evidence for a small moralization effect of disgust on judgments of otherwise neutral actions. It is therefore possible

that incidental disgust, by itself, can lead to a small degree of moralization of otherwise innocuous actions. However, this result is based on a small number of studies ( $k = 13$ ), and the above methodological criticisms obviously apply to these studies as well. More research on this topic is clearly needed. A particularly important question for such research to address is *when* disgust is sufficient for moralization, as there was significant heterogeneity among these effect sizes,  $Q(12) = 22.52, p = .03$ .

### **Conclusion**

The present meta-analysis has established a plausible upper bound on the extent to which incidental disgust amplifies moral judgment (and the extent to which it produces moralization of non-moral actions), though confounds endemic to this literature leave the lower bound yet to be determined. We have also found evidence suggesting publication bias in this literature. The overrepresentation in the published literature of studies finding amplification effects might partially explain many researchers' rapid acceptance of these findings and their enthusiasm for the neo-sentimentalist approach. Our results cannot definitively adjudicate between the rationalist and neo-sentimentalist positions, and doing so was not our aim. Rather, we hope that this review and meta-analysis can further our understanding of the role of disgust in moral judgment and can help to inspire the next wave of rigorous research on how affect and morality intertwine.



### References

- \*Argo, T. N. (2013). *Dirt, disorder, and deontology*. (Unpublished doctoral dissertation). The New School, New York, NY.
- \*Baron, J., Royzman, E. B., & Goodwin, G. P. (2013). *Is disgust a moralizing emotion? Exploring the sufficiency thesis*. Unpublished manuscript. University of Pennsylvania.
- Bodenhausen, G. V. (1993). Emotions, arousal, and stereotypic judgments: A heuristic model of affect and stereotyping. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 13-37). San Diego, CA: Academic Press.
- \*Cameron, C. D., Payne, B. K., & Doris, J. M. (2013). Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology, 49*, 719-725.
- Cannon, P. R., Schnall, S., & White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science, 2*, 325-331.
- \*Case, T. I., Oaten, M. J., & Stevenson, R. J. (2012). Disgust and moral judgment. In R. Langdon & C. Mackenzie (Eds.), *Emotions, Imagination, and Moral Reasoning* (pp. 195-218). New York, NY: Psychology Press.
- \*Cerban, B. & Helweg-Larsen, M. (2010). *Judging smokers: How anger and disgust shape our moral beliefs*. Unpublished manuscript. Dickinson College.
- Chapman, H. A. & Anderson, A. K. (2011). Response to Royzman and Kurzban. *Emotion Review, 3*, 272-273.
- Chapman, H. A. & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin, 139*, 300-327.

- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*, 1222-1226.
- Cheng, J. S., Ottati, V. C., & Price, E. D. (2013). The arousal model of moral condemnation. *Journal of Experimental Social Psychology*, *49*, 1012-1018
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.
- \*David, B. & Olatunji, B. O. (2011). The effect of disgust conditioning and disgust sensitivity on appraisals of moral transgressions. *Personality and Individual Differences*, *50*, 1142-1146.
- Duval, S. (2006). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & Borenstein, M. (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment, and Adjustments* (pp. 127-146). West Sussex, UK: Jon Wiley & Sons.
- Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-463.
- Egger, M., Smith, G. D., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629-634.
- Ekman, P. & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice-Hall.
- \*Eskine, K. J. (2011). *From perceptual symbols to abstraction and back again: The bitter truth about morality*. (Unpublished doctoral dissertation). City University of New York, New York, NY.
- \*Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgments. *Psychological Science*, *22*, 295-299.

- Fine, C. (2006). Is the emotional dog wagging its tail, or chasing it? *Philosophical Explorations*, 9, 83-98.
- Gromet, D. M. & Darley, J. M. (2011). Political ideology and reactions to crime victims: Preferences for restorative and punitive responses. *Journal of Empirical Legal Studies*, 8, 830-855.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116.
- 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, 613-628.
- Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship of disgust to morality. *Psychology and Developing Societies*, 9, 107-131.
- \*Harlé, K. M. & Sanfe, A. G. (2010). Effects of approach and withdrawal motivation on interactive economic decisions. *Cognition and Emotion*, 24, 1456-1465.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- Helzer, E. G. & Pizarro, D. A. (2011). Dirty liberals! Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science*, 22, 517-522.
- \*Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, 97, 963-976.
- Hume, D. (2003). A treatise of human nature. Courier Dover Publications. (Original work published 1739).

- Hutcherson, C. A. & Gross, J. J. (2011). The moral emotions: A social-functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*, 719-737.
- Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition and Emotion, 23*, 714-725.
- \*Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology, 45*, 209-215.
- Jones, A. & Fitness, J. (2008). Moral hypervigilance: The influence of disgust sensitivity in the moral domain. *Emotion, 8*, 613-627.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). New York: Academic Press.
- Landy, J. F. & Piazza, J. (2014). *More than just morality: Disgust sensitivity is associated with a wide array of evaluative judgments*. Manuscript submitted for publication.
- \*Lindeman, L. M. (2010). *Embodied concept learning as a developmental pathway to social disgust*. (Unpublished doctoral dissertation). University of Wisconsin-Madison, Madison, WI.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.
- Loewenstein, G. & Lerner, J. (2003). The role of affect in decision-making. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of affective sciences* (pp.619-642). Oxford, UK: Oxford University Press.

- Miller, L. C., Murphy, R., & Buss, A. H. (1981). Consciousness of body: Public and Private. *Journal of Personality and Social Psychology, 41*, 397-406.
- Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 111*, 99-111.
- \*Moretti, L. & Di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion, 10*, 169-180.
- Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125.
- Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition and Emotion, 16*, 695-703.
- Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin, 135*, 303-321.
- \*O'Connor, C. & Helweg-Larsen, M. (2011). *The role of disgust in morality and perceptions of smoking*. Unpublished manuscript. Dickinson College.
- \*Ong, H. H., Mullette-Gillman, O. A., Kwok, K., & Lim, J. (2014). Moral judgment modulation by disgust is bi-directionally moderated by individual sensitivity. *Frontiers in Psychology, 5*, 194.
- \*Piazza, J., Loughnan, S., & Vauclair, C. M. (2010). [Effects of auditory and visual disgust cues on moral and non-moral judgments]. Unpublished raw data.
- Pizarro, D. A. & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review, 110*, 193-196.
- Pizarro, D., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review, 3*, 267-268.

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, 9, 29-43.

Rosenthal, R. & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.

Royzman, E., Atanasov, P., Landy, J.F., Parks, A., & Gepty, A. (2014). CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the divinity code. *Emotion*, 14, 892-907.

Royzman, E. B., Goodwin, G. P., & Leeman, R. F. (2011). When sentimental rules collide: “Norms with feelings” in the dilemmatic context. *Cognition*, 121, 104-114.

Royzman, E. & Kurzban, R. (2011). Minding the metaphor: The elusive character of moral disgust. *Emotion Review*, 3, 269-271.

Royzman, E. & Kurzban, R. (2011). Facial movements are not goosebumps: A response to Chapman and Anderson. *Emotion Review*, 3, 274-275.

Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, 112, 159-174.

Royzman, E. B., Leeman, R. F., & Sabini, J. (2008). “You make me sick”: Moral dyspepsia as a reaction to third-party sibling incest. *Motivation and Emotion*, 32, 100-108.

Rozin, P., Haidt, J., & Fincher, K. (2009). From oral to moral. *Science*, 323, 1179-1180.

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of Emotions* (3rd ed.) (pp. 757-776). New York: Guilford Press.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (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, 574-586.
- Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology*, 67, 186-205.
- \*Russell, P. S. (2009). *Why is disgust a more unreasoned emotion than anger? An examination of the cognitive processes that accompany moral anger and disgust*. (Unpublished doctoral dissertation). University of Kent, Canterbury, Kent, UK.
- Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion*, 11, 233-240.
- Russell, P. S., Piazza, J., & Giner-Sorolla, R. (2013). CAD revisited: Effects of the word *moral* on the moral relevance of disgust (and other emotions). *Social Psychological and Personality Science*, 4, 62-68.
- Saltzstein, H. D. & Kasachkoff, T. (2004). Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology*, 8, 273-282.
- \*Sato, A. & Sugiura, Y. (2014). Dispositional mindfulness modulates automatic transference of disgust into moral judgment. *Japanese Journal of Psychology*, 84, 605-611.
- \*Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19, 1219-1222.
- \*Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34, 1096-1109.
- Schwarz, N. & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.

- Schwarz, N. & Clore, G. L. (1988). How do I feel about it? The informative function of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior* (pp. 44-62). Toronto, Canada: Hogrefe.
- \*Seidel, A. & Prinz, J. (2013). Sound morality: Irritating and icky noises amplify judgments in divergent moral domains. *Cognition*, *127*, 1-5.
- Simonsohn, U., Simmons, J., & Nelson, L. (2014, December 3). [30] Trim-and-fill is full of it (bias). [Web log post]. Retrieved from <http://datacolada.org/2014/12/03/30-trim-and-fill-is-full-of-it-bias-2>
- Simpson, J., Carter, S., Anthony, S. H., & Overton, P. G. (2006). Is disgust a homogeneous emotion? *Motivation and Emotion*, *30*, 360-365,
- \*Singh, G. (2010). *Effect of core and sociomoral disgust on moral judgments*. Unpublished manuscript. California State University, Fresno.
- \*Strohming, N. (2012). *The hedonics of disgust*. Unpublished doctoral dissertation. University of Michigan.
- \*Strohming, N., Lewis, R. L., & Schwarz, N. (2010). [Wrongness and humor of “ridiculous” moral violations]. Unpublished raw data.
- Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., ... Rescober, P. (2007). People as intuitive prosecutors: The impact of social control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, *43*, 195-209.
- \*Tobias, M. & Helweg-Larsen, M. (2010). *Judging fatness: The effects of emotion on moral judgment and moral judgment on emotion*. Unpublished manuscript. Dickinson College.
- Travers, J. B., & Norgren, R. (1986). Electromyographic analysis of the ingestion and rejection of sapid stimuli in the rat. *Behavioral Neuroscience*, *100*, 544-555.



- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: Individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology, 97*, 103-122.
- Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli (2013). Disgust: Evolved function and structure. *Psychological Review, 120*, 65-84.
- Tybur, J. M., Merriman, L. A., Caldwell Hooper, A. E., McDonald, M. M., & Navarrete, C. D. (2010). Extending the behavioral immune system to political psychology: Are political conservatism and disgust sensitivity really related? *Evolutionary Psychology, 8*, 599-616.
- \*Ugazio, G., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion, 12*, 579-590.
- \*Van Dillen, L. F., van der Wal, R. C., & van den Bos, K. (2012). On the role of attention and emotion in morality: Attentional control modulates unrelated disgust in moral judgments. *Personality and Social Psychology Bulletin, 38*, 1222-1231.
- \*Vernon, D., D'Amico, C., Glazer, K., & Helweg-Larsen, M. (2012). *The effect of disgust on moral judgment in Americans and Indians: Assessing the CAD hypothesis*. Unpublished manuscript. Dickinson College.
- \*Wagner, B. C. (2012). *The effects of disgust on social judgment: A thought-validation perspective*. (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.
- \*Wheatley, T. & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*, 780-784.

\*Zhong, C. B., Strejcek, B., & Sivanathan, N. (2010). A clean self can render harsh moral judgment. *Journal of Experimental Social Psychology, 46*, 859-862.

\*Zickgraf, H., Kim, Y., & Rozin, P. (2012). *Physical cleanliness and moral decision-making: The effect of Private Body Consciousness*. Unpublished manuscript. University of Pennsylvania.

\*Indicates that a study or studies from the paper were included in the meta-analysis.

### Footnotes

<sup>1</sup> In describing his Social Intuitionist Model, Haidt uses the terms “intuition” and “emotion” more-or-less interchangeably. It is not necessary to the model that the “intuitions” be rooted in disgust, only that they involve some form of automatic affective or evaluative response. However, far more research has been conducted on disgust than on other putative moral emotions. As we describe in the main text, presumably one reason for this is that disgust as opposed to other negative emotions, like anger, does not involve an intrinsically normative component, thus in principle allowing a clearer test of the effect of “pure” emotion on moral judgment. The disgust induction studies that are meta-analyzed in this paper therefore provide the best current evidence for the social intuitionist hypothesis that emotion causally influences moral judgment.

<sup>2</sup> This argument could, potentially, be made about almost any study that successfully induces disgust but fails to show the amplification effect. The problem with this, of course, is that in the absence of any independent criterion for what the ideal level of induction potency is, it renders the amplification hypothesis unfalsifiable. Therefore, until such a criterion is articulated in the literature, we can only assess whether studies induced “enough” disgust – i.e., showed successful manipulation checks – but not whether they induced “too much” disgust, as Schnall, Haidt et al. speculate.

<sup>3</sup> It might be argued that punishment in economic games depends on factors other than moral judgment, and therefore including in our meta-analysis studies in which this is the dependent variable might diminish the observed effect sizes. In fact, quite the opposite is true. There are three studies of punishment in economic games in the meta-analysis, two of which constitute the two largest effect sizes in the entire dataset. When all three are excluded, the observed effect size becomes smaller than what is reported in the main text,  $d = .08$ ,  $p = .02$ . The difference in

effect sizes between published,  $d = .12$ , and unpublished,  $d = .03$ , studies becomes non-significant,  $Q(1, 45) = 1.24$ ,  $p = .27$ , but Egger's regression remains significant, Intercept = .90,  $t(26) = 2.45$ ,  $p = .02$ , and trim-and-fill imputes eight studies and estimates a revised effect size of  $d = -.02$ , almost identical to the estimate obtained when the studies of economic games are included.

<sup>4</sup> One of the unpublished studies involved only judgments of non-moral actions (Lindeman, 2010), and therefore speaks only to the moralization hypothesis, rather the amplification hypothesis. It was therefore only included the fourth section of the meta-analysis, in which we examine this hypothesis. It is also excluded from Figure 1, because no effect size for the amplification effect was calculated from this study.

<sup>5</sup> Strictly speaking, we did not use Duval and Tweedie's (2000) original procedure, which imputes missing studies using a random effects model. More recently, Duval (2006) recommended using the more conservative fixed effects version of trim-and-fill, and that is the approach that we have taken here. This approach generally imputes more missing studies, and therefore adjusts the estimated effect size to a greater degree than the random-effects procedure. It is therefore less susceptible to over-estimating the true effect size, a bias in trim-and-fill that was pointed out in a recent blog post (Simonsohn, Simmons, & Nelson, 2014). The method for imputing missing studies should not be confused with the model used to combine the effect sizes after the new effect sizes were imputed: as in all reported analyses, a random effects model was assumed for that purpose.

<sup>6</sup> Two statistical issues with this database bear mentioning. First, several of the data points are non-independent because they are drawn from the same studies. Relatedly, when raw data were not available and it was possible to do so, we estimated effect sizes from means and standard

deviations of multiple judgments. Pooling the standard deviations treats the judgments as independent, but they would actually be expected to correlate within-subjects. Thus, the standard deviations used in calculating some effect sizes in this analysis were slightly too large, producing effect size estimates that are slightly too small. However, there is reason to think that the statistical shortcomings of this database did not much affect the obtained results. We re-ran the moderator analysis for sensory modality (which varied only between studies) using only one overall effect size per study, and the pattern of results was largely consistent with those reported in the main text. The only differences were that the effect was reduced to marginal significance,  $Q(3, 46) = 6.55, p = .09$ , probably due to the smaller number of data points in this analysis, and the effect size estimates tended to be slightly larger overall (Gustatory/Olfactory:  $d = .50$ , Visual:  $d = .15$ , Imagined:  $d = .08$ , Other:  $d = .03$ ).

<sup>7</sup>To power a study using an independent-groups design at  $\beta = .20$  (i.e., with statistical power of .80) and  $\alpha = .05$ , two-tailed, with an assumed effect size of  $d = .11$ , one would need 1299 participants in each condition, for a total  $N$  of 2598. To power the same study at the more stringent  $\beta = .10$ , one would need 1738 participants in each condition, for a total  $N$  of 3474. These numbers are for a simple two-group design; they do not account for factorial designs in which a hypothesized moderator is also manipulated or measured. It may be impractical to run such studies, regardless of how theoretically interesting they might be.

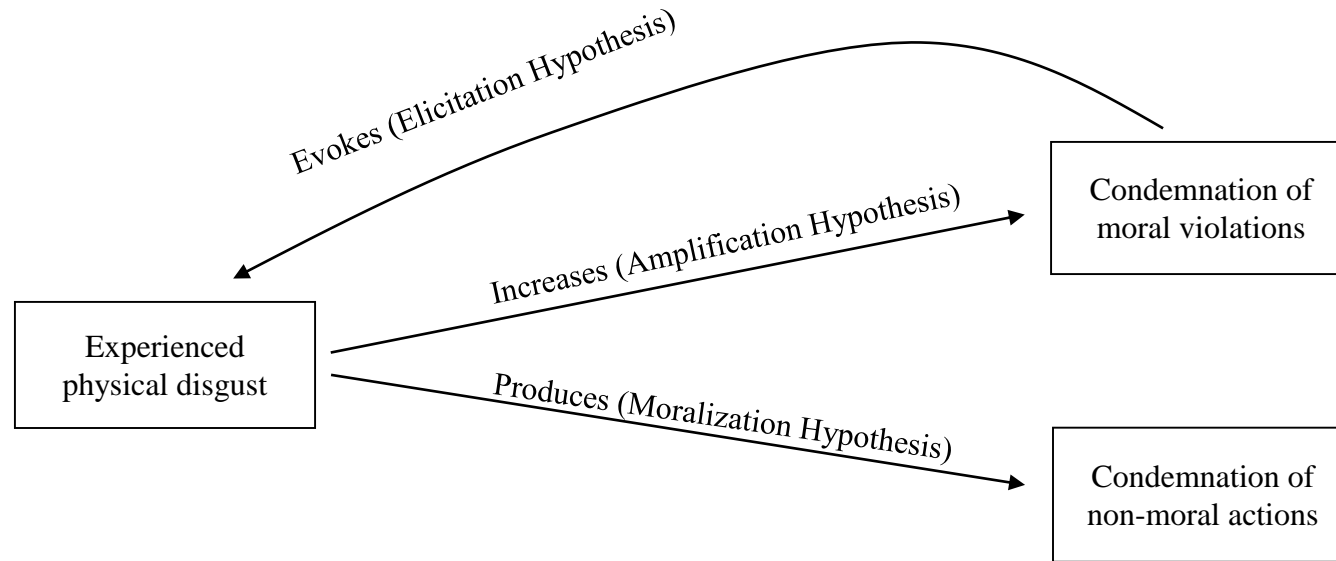


Figure 1. Schematic representation of three possible relationships between disgust and moral judgment.

Table 1. Studies included in the meta-analysis, moderators, sample sizes, and effect sizes. For ease of interpretation, effect sizes are reported as uncorrected *ds*. Sample sizes and effect sizes are reported for each study overall (these effect sizes were used in the first and second sections of the meta-analysis), and separately for non-purity violations and purity violations (used in the third section), and non-moral judgments (used in the fourth section). Dashes indicate that a study did not include a judgment of that type.

Published Studies	Sensory Modality of Disgust Induction	Overall		Non-Purity		Purity		Non-Moral	
		<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>
Cameron et al. (2013), Study 1	Visual	130	-.10	Data by transgression type unavailable					
Cameron et al. (2013), Study 2	Visual	119	-.04	Data by transgression type unavailable					
Case et al. (2012), Study 1	Visual	63	.05	63	.05	--	--	--	--
Case et al. (2012), Study 2	Visual	82	.17	82	.17	--	--	--	--
Cheng et al. (2013), Study 1	Visual	121	.04	121	.04	--	--	--	--
Cheng et al. (2013), Study 2	Imagined	33	.07	33	.06	33	.38	--	--
David & Olatunji (2011)	Other (Evaluative Conditioning)	61	.11	Data by transgression type unavailable					
Eskine et al. (2011)	Gustatory/Olfactory	54	1.16	54	1.01	54	1.18	--	--
Harlé & Sanfey (2010)	Visual	74	.46	74	.46	--	--	--	--

Horberg et al. (2009), Study 2	Visual	122	.29	122	.15	122	.37	--	--
Johnson et al. (2014), Study 2	Visual	126	-.02	126	.05	126	-.04	--	--
Moretti & di Pellegrino (2010), Study 1	Visual	45	1.34	45	1.34	--	--	45	1.10
Moretti & di Pellegrino (2010), Study 2, Human condition	Visual	30	1.44	30	1.44	--	--	30	.76
Ong et al. (2014), Study 1, Sample A	Other (Primed with Facial Expression)	21	-.37	21	-.37	--	--	--	--
Ong et al. (2014), Study 1, Sample B	Other (Primed with Facial Expression)	27	-.38	27	-.38	--	--	--	--
Ong et al. (2014), Study 2	Other (Primed with Facial Expression)	26	-.02	26	-.02	--	--	--	--
Sato & Sugiura (2014)	Imagined	78	.73	78	.73	--	--	--	--
Schnall, Benton, & Harvey (2008), Study 2	Visual	43	.85	43	.70	43	.34	--	--
Schnall, Haidt et al. (2008), Study 1	Gustatory/Olfactory	120	.75	120	.42	120	.45	--	--
Schnall, Haidt et al. (2008), Study 2	Visual <sup>a</sup>	43	.13	43	.21	43	.07	--	--
Schnall, Haidt et al. (2008), Study 3	Imagined	69	.04	69	.11	69	-.03	--	--
Schnall, Haidt et al. (2008), Study 4	Visual	133	.27	133	.14	133	.38	--	--
Schnall, Haidt et al. (2008), Pilot Study summarized in General Discussion	Other (Tactile)	43	-.12	43	-.17	43	.02	42 <sup>b</sup>	-.30
Seidel & Prinz (2013)	Other (Auditory)	113	.43	59	-.06	54	.92	--	--
Ugazio et al. (2012), Study 1a	Gustatory/Olfactory	58	.10 <sup>c</sup>	58	.10 <sup>c</sup>	58	.03 <sup>c</sup>	58	.08 <sup>c</sup>



Ugazio et al. (2012), Study 1b	Visual	109	.10 <sup>c</sup>	109	.09 <sup>c</sup>	109	.07 <sup>c</sup>	109	.01 <sup>c</sup>
Van Dillen et al. (2012), Study 1	Imagined	74	.97	74	.97	--	--	--	--
Wheatley & Haidt (2005), Study 1	Other (Hypnotic Suggestion)	45	.53	33	.35	21	.44	--	--
Wheatley & Haidt (2005), Study 2	Other (Hypnotic Suggestion)	63	.25	61	.05	58	.22	63	.81
Zhong et al. (2010), Study 2	Imagined	323	-.20	--	--	323	-.20	--	--
Zhong et al. (2010), Study 3	Imagined	136	-.37	--	--	136	-.37	--	--
		<b>Overall</b>		<b>Non-Purity</b>		<b>Purity</b>		<b>Non-Moral</b>	
		<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>	<i>N</i>	<i>d</i>
<b>Unpublished Studies</b>	<b>Sensory Modality of Disgust Induction</b>								
Argo (2013), Study 1	Visual <sup>a</sup>	92	.47	92	.47	--	--	--	--
Argo (2013), Study 2	Visual <sup>a</sup>	76	-.22	76	-.22	--	--	--	--
Baron et al. (2013), Study 1	Visual	297	-.15	97	-.27	101	-.08	99	-.09
Baron et al. (2013), Study 2	Gustatory/Olfactory	323	-.18	323	-.11	323	-.20	323	.07
Baron et al. (2013), Study 3	Visual	375	.03	375	-.03	375	.09	375	.12
Cerban & Helweg-Larsen (2010), Study 1	Visual	60	.01	60	.09	60	-.08	--	--
Eskine (2011), Study 3	Gustatory/Olfactory	27	.11	Data by transgression type unavailable					

Lindeman (2010), Study 1 <sup>d</sup>	Gustatory/Olfactory	218	.41	--	--	--	--	218	.41
O'Connor & Helweg-Larsen (2011)	Visual	45	.07	45	-.01	45	.04	45	.12
Piazza et al. (2010), Study 1	Visual	107	-.06	107	.01	107	-.12	107	.11
Piazza et al. (2010), Study 2	Other (Auditory)	53	-.11	53	-.26	53	-.02	53	.28
Russell (2009), Study 4	Gustatory/Olfactory	54	-.04	54	-.28	54	.13	--	--
Singh (2010)	Visual	80	.35	80	.35	--	--	--	--
Strohminger (2012), Study 2	Imagined	131	.09	131	.09	--	--	--	--
Strohminger et al. (2010), Study 1	Imagined	20	.09	20	.10	20	.10	--	--
Strohminger et al. (2010), Study 2	Imagined	85	-.10	85	-.21	85	-.05	--	--
Tobias & Helweg-Larsen (2010), Study 1	Visual	65	-.19	63	.13	63	-.31	--	--
Vernon et al. (2010)	Visual	88	.05	88	-.04	88	.16	--	--
Wagner (2012), Study 3	Imagined	64	.34	Data by transgression type unavailable <sup>d</sup>					
Zickgraf et al. (2012), Study 2	Imagined	258	.19	258	-.10	258	.23	--	--

<sup>a</sup> The disgust induction in this study was a filthy work area. It was mostly visual, but included olfactory and tactile elements as well.

<sup>b</sup> Data from one participant were missing.

<sup>c</sup> Four of the "Thoughts" scenarios contained both moral and non-moral elements (e.g., bad intent with no bad consequences, or vice versa). Because these could not be easily classified as moral or non-moral, they were not included in our analyses.

<sup>d</sup>In this study, participants self-generated transgressions that they had witnessed, then made character judgments of the perpetrators.

Because the transgressions were not controlled and likely varied greatly, this study was not used in the moderator analyses.

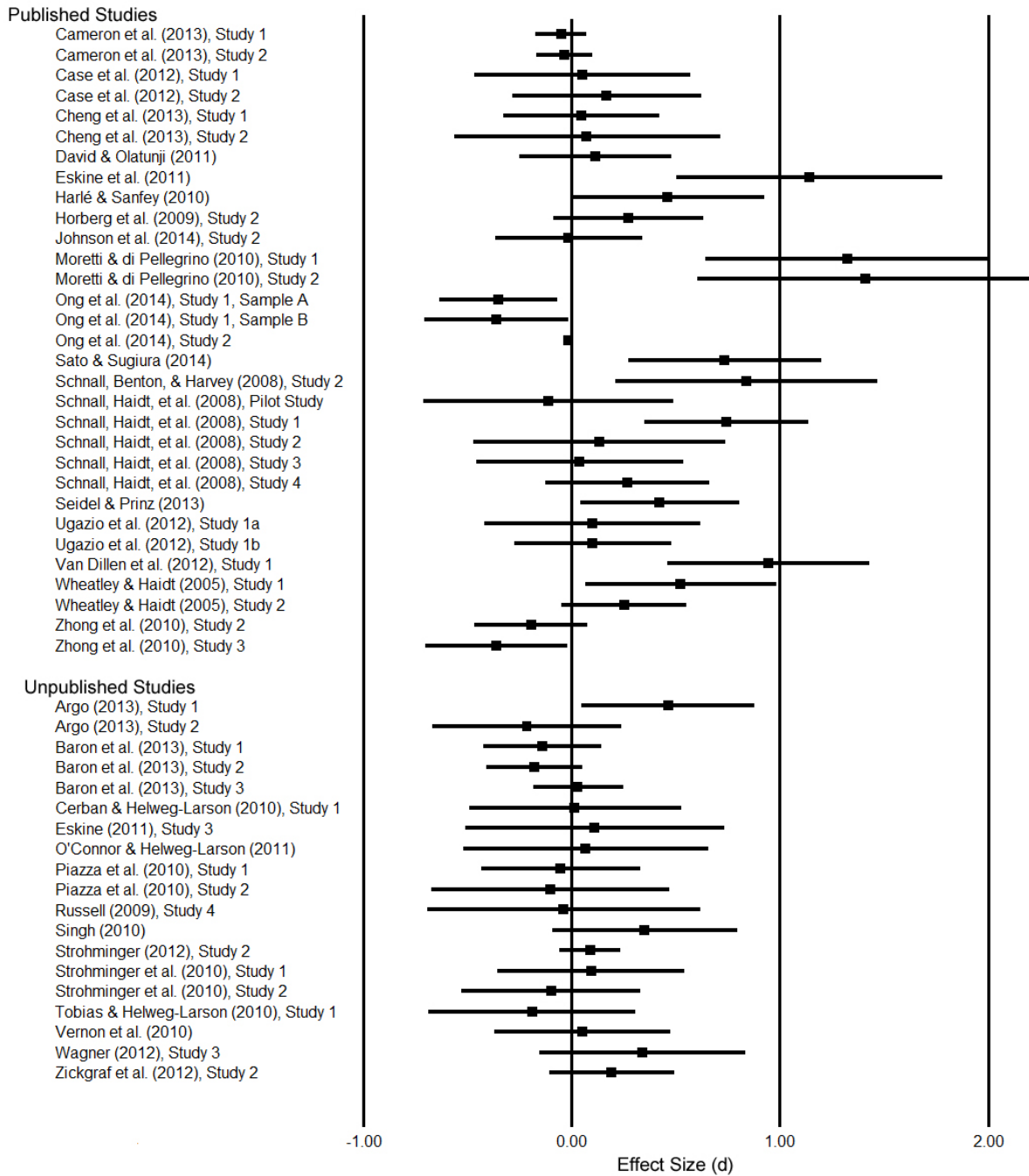


Figure 2. Forest plot showing effect sizes and confidence intervals for all published and unpublished studies of the amplification effect.

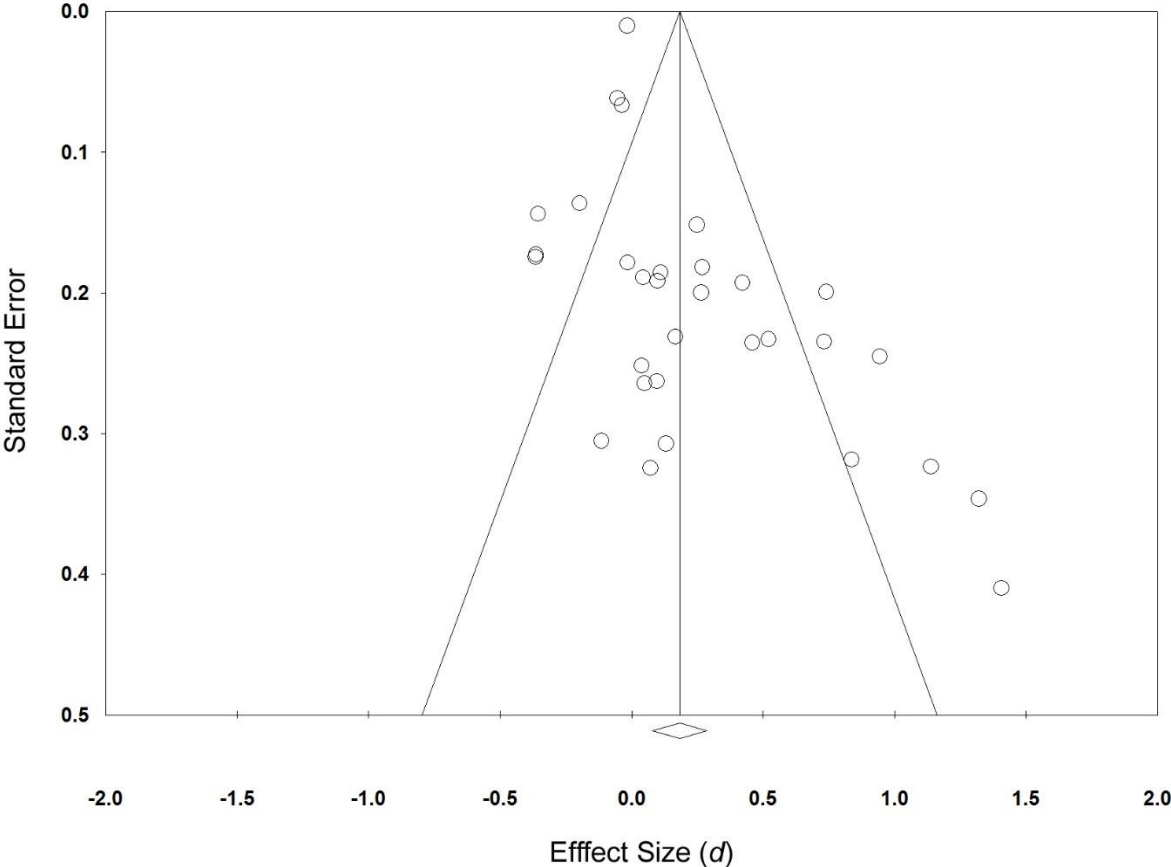


Figure 3. Funnel plot of published incidental disgust induction effect sizes.

Table 2. Number of studies in each moderator category (*k*), estimated effect sizes for each category (*d*), and 95% confidence intervals for effect size estimates (95% CI).

	<i>k</i>	<i>d</i>	95% CI
Gustatory/Olfactory Disgust Inductions	8	.37	[.15, .60]
Visual Disgust Inductions	35	.13	[.03, .22]
Imagined Disgust Inductions	17	.04	[-.09, .17]
Other Disgust Inductions	13	.01	[-.15, .16]
Non-Purity Violations	43	.11	[.02, .19]
Purity Violations	30	.10	[-.01, .19]

## Online Supplemental Materials

### Details of Literature Search

Several types of papers from the initial online search did not meet our two criteria and were not included in the meta-analysis. These included studies in which subjective disgust was a dependent variable rather than an independent variable, neurological studies in which emotion centers of the brain were studied during moral judgment tasks, review papers, and correlational studies in which the independent variable was measured trait disgust sensitivity rather than manipulated state disgust. Moreover, studies in which the dependent variable was liking toward outgroups (e.g., Inbar et al., 2012) were excluded because we did not consider this type of judgment to be purely moral.

We also excluded studies that manipulated only subjective cleanliness rather than disgust. This approach contrasts with that taken in a recent systematic review of research on moral disgust (Chapman & Anderson, 2013) in which it was argued that cleanliness is the “conceptual opposite of disgust.” While cleanliness and disgust are certainly related concepts, we chose not to include cleanliness studies in this meta-analysis because the theoretical underpinning of these studies and the psychological mechanisms that they engage are very different from typical studies of incidental disgust. The latter are, at their core, misattribution studies; disgust is induced, and then presumably at least some of the felt disgust is attributed to the judged moral transgression, leading to an increased judgment of its severity (Schnall, Haidt et al., 2008). Cleanliness studies, on the other hand, typically involve changes in participants’ own self-concepts, and are based on the idea that when one feels physically – and by extension, morally – clean, one will seek to protect that status by rejecting moral transgressors (Zhong, Strejcek, &

Sivanathan, 2010; Helzer & Pizarro, 2011). In essence, incidental disgust studies manipulate how “disgusted” participants feel, whereas cleanliness studies manipulate how “(not) disgusting” they feel. Thus, cleanliness studies originate from a different theoretical approach and rely on different psychological mechanisms than do incidental disgust studies. Accordingly, while we included studies that compared cleanliness conditions to explicit inductions of disgust (e.g., Schnall, Benton, & Harvey, 2008, Study 2; Zhong et al., 2010, Studies 2 and 3), we excluded those studies that included no induction of disgust and only compared cleanliness and control conditions (i.e., Schnall, Benton, & Harvey, 2008, Study 1; Zhong, et al., 2010, Study 1; Helzer & Pizarro, 2011). Beyond these theoretical reasons for excluding these cleanliness studies, we also think that this approach gives the amplification hypothesis the best possible chance at being supported. On the Chapman and Anderson (2013) approach, cleanliness inductions are taken to be the “conceptual opposite” of disgust inductions, that is, they are thought to represent a “lack of disgust,” or perhaps “negative disgust.” However, cleanliness inductions have typically produced harsher moral judgments compared with control conditions – in these studies, therefore, it is the apparently *less disgusting* condition that amplifies moral judgment. Thus, including these results would weaken the overall evidence for the amplification hypothesis, in a way that we do not see as fully justified.

Our initial literature search through the online databases yielded 16 published papers and one published book chapter containing a total of 30 relevant studies. This search was then supplemented by a forward and backward citation search from each of these 17 papers, which uncovered one additional published paper containing one relevant study. Additionally, we contacted the corresponding authors of all 18 of these published reports and requested any unpublished studies on this topic and information about any colleagues who may have conducted



such studies. We also posted announcements to the email listservs of the Society for Social and Personality Psychology, the Society for Judgment and Decision Making, the Society for Experimental Social Psychology, and the Society for Philosophy and Psychology, and on the *Perspectives* blog, describing the meta-analysis and asking for any results, published or unpublished, on the topic of interest.

**Supplemental Table 1. Methodological summary of unpublished studies included in the meta-analysis.**

<b>Study</b>	<b>Disgust Induction</b>	<b>Dependent Variable</b>	<b>Validation of Disgust Induction</b>
Argo (2013), Study 1	Filthy work area, similar to Schnall, Haidt et al. (2008, Study 2)	Dichotomous rating of killing one person to save more as “acceptable” or “unacceptable” in several hypothetical trolley-type dilemmas	None reported
Argo (2013), Study 2	Filthy work area, similar to Schnall, Haidt et al. (2008, Study 2)	Dichotomous rating of killing one person to save more as “acceptable” or “unacceptable” in several hypothetical trolley-type dilemmas	None reported
Cerban & Helweg-Larsen (2010), Study 1	Film clip depicting a woman vomiting in a dirty bathroom, retrieving gum from the toilet, and chewing it	Likert ratings of wrongness of actions and immorality of person; vignettes were original, but correspond to the three moral domains of the CAD hypothesis	Successful pre-test and manipulation check
Eskine (2011), Study 3	Bitter beverage (see Eskine et al., 2011 for details); manipulation was within-subjects and concurrent with moral judgment task	Continuous rating of wrongness of actions; vignettes were drawn from Wheatley & Haidt (2005)	Successful manipulation check
Lindeman (2010), Study 1	Disgusting ambient smell	Likert ratings of moral wrongness and likeability of characters; vignettes were	Manipulation check was conducted;

		original and largely amoral in content	detailed results not reported; seems to have been largely unsuccessful
O'Connor & Helweg-Larsen (2011)	Film clip depicting a woman vomiting in a dirty bathroom, retrieving gum from the toilet, and chewing it	Likert ratings of wrongness of actions and immorality of person; vignettes were original, but correspond to the three moral domains of the CAD hypothesis and were pilot tested	Successful manipulation check
Piazza, et al. (2010), Study 1	Images of core disgust elicitors (maggots, moldy fruit, etc.)	Likert ratings of "approval" of character's behavior; vignettes were adapted from Schnall, Haidt et al. (2008)	Successful manipulation check
Piazza et al. (2010), Study 2	Audio recordings of a person burping and vomiting	Likert ratings of "approval" of character's behavior; vignettes were mostly adapted from Schnall, Haidt et al. (2008) and Haidt et al. (1999), some were original	Successful pre-test; unsuccessful manipulation check, but trend was as predicted
Baron et al. (2013), Study 1	Article about the Bristol Stool Scale, with descriptions and illustrations of different consistencies of feces	Likert ratings of wrongness of actions; vignettes were adapted from Wheatley & Haidt (2005)	Successful pre-test
Baron et al. (2013), Study 2	Instructions to vividly imagine ingesting a food or medicine that the participants personally find disgusting and aversive	Continuous measures of wrongness; vignettes were adapted from several prior published studies	Successful manipulation check

Baron et al. (2013), Study 3	Short vignettes about disgusting experiences with photographic illustrations	Continuous measures of wrongness; vignettes were adapted from several prior published studies	Successful manipulation check
Russell (2009) Study 4	Disgusting ambient odor, modeled after Schnall, Haidt et al. (2008, Study 1)	Multi-item scale measuring moral disapproval of either pedophiles or Islamic fundamentalists	Successful manipulation check
Singh (2010)	Disgusting images, adapted from Simpson, Carter, Anthony, & Overton, 2006	Likert ratings of wrongness of actions; vignettes were drawn from Schnall, Haidt et al. (2008)	Successful manipulation check
Strohminger (2012), Study 2	Unscrambled disgusting sentences (e.g., “noxious was sweat his”)	Likert ratings of wrongness of humorous purity violations; vignettes were mostly original, some were adapted from prior studies	Successful pilot test
Strohminger et al. (2010), Study 1	Unscrambled disgusting sentences (e.g., “noxious was sweat his”)	Likert ratings of wrongness of humorous purity violations; vignettes were mostly original, some were adapted from prior studies	Successful pilot test
Strohminger et al. (2010), Study 2	Unscrambled disgusting sentences (e.g., “noxious was sweat his”)	Likert ratings of wrongness of humorous purity violations; vignettes were mostly original, some were adapted from prior studies	Successful pilot test

Tobias & Helweg-Larsen (2010), Study 1	Film clip depicting a woman vomiting in a dirty bathroom, retrieving gum from the toilet, and chewing it	Likert ratings of wrongness of actions and immorality of person; vignettes were original, but correspond to the three moral domains of the CAD hypothesis	Successful pre-test and manipulation check
Vernon, et al. (2010)	Film clip depicting a woman vomiting in a dirty bathroom, retrieving gum from the toilet, and chewing it	Likert ratings of wrongness of actions and immorality of person; vignettes were adapted from Rozin et al. (1999)	Successful pre-test and manipulation check
Wagner (2012), Study 3	Participants wrote about a recent personal experience that made them feel disgusted	Likert ratings of character of friend or acquaintance who had performed prior bad behaviors	Successful pre-test
Zickgraf, Kim, & Rozin (2012), Study 2	Participants read and reproduced a paragraph describing being dirty; manipulation was taken from Zhong et al. (2010, Study 2)	Likert ratings of immorality of actions; actions were taken from Zhong et al. (2010, Study 2)	None reported

---