



Is There Such a Thing as Genuinely Moral Disgust?

Mara Bollard¹ 

Accepted: 8 March 2021/Published online: 19 March 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

In this paper, I defend a novel skeptical view about moral disgust. I argue that much recent discussion of moral disgust neglects an important ontological question: is there a distinctive psychological state of *moral* disgust that is differentiable from generic disgust, and from other psychological states? I investigate the ontological question and propose two conditions that any aspiring account of moral disgust must satisfy: (1) it must be a genuine form of disgust, and (2) it must be genuinely moral. Next, I examine two prominent accounts of moral disgust by John Kekes and Victor Kumar and argue that neither successfully establishes the existence of genuinely moral disgust: Kekes' account does not satisfy condition (2), and Kumar's view does not meet condition (1). I claim that an important general lesson can be drawn from my critiques of Kekes' and Kumar's accounts: to establish the existence of moral disgust, one must provide unequivocal evidence that genuinely moral disgust, not generic disgust or anger, is being elicited in response to relevant moral violations. I conclude by considering why we ought to be skeptical about the general prospect of giving a positive answer to the ontological question, given the available evidence.

Keywords Disgust · Moral psychology · Moral disgust · Moral emotion · Anger

1 Introduction

Two decades ago, Leon Kass touted the 'wisdom of repugnance,' citing the felt experience of disgust as evidence of the moral wrongness of human cloning (1997: 20). Since then, discussions of moral disgust have become increasingly frequent, both in the psychological and philosophical literatures. Advocates of moral disgust (Miller 1997; Kass 1997; Kekes 1998; Kumar 2017) all defend some version of the claim that disgust has a legitimate – or proper, positive, important, or indispensable – role to play

✉ Mara Bollard
mbollard@umich.edu

¹ Weinberg Institute for Cognitive Science, University of Michigan, Ann Arbor, MI, USA

in moral judgment and discourse. Moral disgust skeptics (Kelly and Morar 2014; Bloom 2013; Kelly 2011; Nussbaum 2004), on the other hand, seek to deny such claims.

How precisely to interpret the disputes between moral disgust's advocates and its critics is a surprisingly complex matter. As Alberto Giubilini (2016) has recently observed, 'moral disgust is a lamentably obscure term,' one in need of further elucidation. I suggest that there are (at least) four distinct questions in play in the current literature on moral disgust. First, there's an ontological question: is there a distinctive psychological state of *moral* disgust that is differentiable from (what I will term) generic disgust, and from other psychological states? Second, there's a normative question about fittingness¹: is disgust ever fitting as a response to (some feature of) moral violations? This question about fittingness is best thought of as a question about correctness: when we ask whether an emotion is fitting, in the sense relevant to whether its object has certain features or properties, we are asking about whether the emotion correctly presents its object as having those features (D'Arms and Jacobson 2000). Third, there's another normative question about the moral appropriateness of disgust: is it ever morally right to feel disgust at objects in the moral domain? It's important to note that questions two and three are presented separately in an effort to avoid conflating fittingness and the moral appropriateness (or rightness, or permissibility) of disgust, a fallacy Justin D'Arms and Daniel Jacobson (2000) rightly warn against.² As D'Arms and Jacobson put it, 'there is a crucial distinction between the question of whether some emotion is the morally right way to feel, and whether that feeling gets it right' (2000: 66). Finally, we can distinguish a fourth question: a question of moral epistemology about the reliability of disgust: Does disgust reliably track morally relevant features or properties?

Interestingly, the first, ontological question – whether there is such a thing as a distinctive psychological state of moral disgust – is underexplored, and is often not addressed explicitly in debates (especially among philosophers) about moral disgust.³ This is surprising, not least because proper interpretation and investigation of the other questions I identified above may depend on our having properly identified the psychological state we're talking about – or not talking about – when we try to answer them.

In this paper, I investigate the ontological question. This paper proceeds as follows. In section 2, I put forward two conditions that, I claim, any aspiring account of genuinely moral disgust must satisfy. In section 3, I apply these conditions to two prominent accounts of moral disgust – by John Kekes and Victor Kumar – and argue that neither succeeds in vindicating the existence of genuinely moral disgust. Informed

¹ Giubilini (2016) has recently convincingly argued that this very conceptual confusion plagues discussions of moral disgust.

² Plakias (2017) defends an account of moral disgust that provides a positive answer to the question about fittingness, where the emotional response in question is *generic* disgust. So, she does not take a stand on the ontological question, as I will be delineating it in this paper.

³ For notable exceptions see, e.g. Giner-Sorolla et al. (2018) and Russell and Giner-Sorolla (2013), which both address the importance – and difficulty – of distinguishing moral disgust both from other kinds of disgust, such as physical (generic, in my terms) disgust, and from other emotions, such as anger; Giubilini (2016), who questions whether moral disgust is a genuine form of disgust or some other kind of psychological state; Chapman et al. (2009), who argue that moral disgust just is the same psychological state as physical disgust, and Royzman and Kurzban (2011), who argue (contra Chapman et al., 2009) that the term 'moral disgust' is often used metaphorically and thus does not imply the existence of a special emotion kind.

by the specific problems found in Kekes' and Kumar's accounts of moral disgust, each of which fails to satisfy one of my proposed conditions, I conclude by discussing why we ought to be skeptical about the general prospect of successfully establishing the existence of genuinely moral disgust.

First, though, I will briefly explain why the position I am defending is different from existing skeptical views about moral disgust. The *ontological* skepticism I am advancing is distinct from two influential skeptical accounts in the moral disgust literature, defended by Daniel Kelly (2011) and Martha Nussbaum (2004). Though, like me, Kelly and Nussbaum argue in favor of versions of general skepticism about moral disgust, they are primarily concerned with questions other than the ontological question. Kelly, a self-avowed 'disgust skeptic' argues that, given what we know about generic disgust's evolutionary story and its functional role, there is no good reason to think that disgust is a reliable guide to the moral status of its elicitors, because disgust is designed to be hypersensitive and will thus lead to too many false positives (2011: 139). In short, Kelly is skeptical about the fourth question I identified above: the question of moral epistemology. By contrast, Nussbaum (2004) argues that disgust has no place in the moral domain because it leads to harmful consequences overall. More specifically, the effects of disgust on people's thought and behavior bring harm to members of disadvantaged and marginalized groups who are treated as objects of disgust. So, Nussbaum is a skeptic about the third question listed above: i.e. she thinks it is never morally appropriate to feel disgust in the moral domain. Though it's true that Kelly and Nussbaum may well agree with my (eventual, and admittedly provisional) negative answer to the ontological question and accept that there is no such thing as moral disgust, it is worth noting the differences in our skeptical targets at the outset. Arguing that there is no such thing as genuinely moral disgust, as I shall do here, is importantly different from arguing that disgust is unreliable or that it is not morally appropriate.

2 What Is Genuinely *Moral Disgust*?

The ontological question at issue in this paper is whether there is a distinctive psychological state of genuinely moral disgust. I propose that if a given emotional response is to count as genuinely moral disgust, it must satisfy both of the following conditions:

1. It must be a genuine form of disgust.
2. It must be genuinely moral.

2.1 Interlude: Some Assumptions about the Emotions

It is beyond the scope of the present paper to comprehensively defend a specific theory of the emotions. However, a few words are in order about the assumptions I am making about the emotions. I assume, following a number of emotion theorists (e.g. D'Arms and Jacobson [forthcoming](#); D'Arms and Jacobson 2003; Frijda 1986; Frijda 2007), that emotions are best understood as syndromes of thought (where these thoughts are

characteristic rather than constitutive), feeling, and motivation. The motivational component is particularly important when it comes to characterizing and distinguishing emotions. The motivational role includes a distinctive action tendency, a goal that constitutes satisfaction of the emotion, and the focusing of attention on the emotional goal. This means emotions tend to seek precedence and control over behavior, thought, and experience.

I also assume that emotions are elicited by and directed at objects (which can include actions, agents, states of affairs, thoughts, etc.), and carry some kind of representational content about their elicitors. For example, sadness typically represents its elicitor as involving a loss. It cannot be assumed that a particular object will necessarily or universally elicit a given emotion – a tarantula might inspire fear in me today, but not tomorrow; it might frighten you every time you see it, but never scare your mother; it might inspire eager anticipation instead of fear in your Cambodian friend who grew up regarding tarantulas as a delicious delicacy. So, elicitors may vary across people and cultures, for a variety of reasons. Still, it makes sense to say that emotions have paradigm elicitors; that is, objects that typically elicit the relevant emotional response in most people.

2.2 What Counts as a Genuine Form of Disgust?

To meet condition (1), a given emotional response must be shown to be a *genuine form of disgust*. More specifically, it must be a subspecies of the psychological kind that I'm calling generic disgust.⁴ This means that the emotional response in question must be sufficiently similar to generic disgust in terms of (some of) its characteristic features, such as its characteristic thought(s) or appraisal(s), phenomenology, motivational and behavioral profile, and perhaps also its pattern of neural activation.

What is the nature of generic disgust? Disgust is a negatively-valenced emotion that is thought to have originated in distaste for bitter, possibly toxic foods in order to protect us from pathogens and infection. Ekman and Friesen (1971) identified disgust as one of six basic, universal emotions with a characteristic facial expression that is recognized across cultures. Disgust's telltale facial expression is characterized by wrinkling of the nose and retraction of the upper lip. The lips may be closed, to prevent the offensive object from entering the mouth, or the lips and mouth may be open ('gaping'), accompanied by extrusion of the tongue and vocal sounds. Disgust also induces a subjective experience of revulsion, which may involve feelings of nausea, sweating, and shivering. The action tendencies typical of disgust are withdrawal or avoidance behaviors; the gape face, for instance, is a reaction that prevents ingestion of and encourages oral expulsion of suspect, possibly contaminated substances. Disgust involves a perceived threat of contamination, and motivates us to withdraw from the contaminant (Tybur et al. 2013; Chapman et al. 2009). Given disgust's evolutionary function to protect us from contamination, the paradigm elicitors of disgust are objects that present possible threats of disease

⁴ What I'm calling generic disgust in this paper is also often referred to as 'physical disgust,' 'bodily disgust,' 'pathogen disgust,' or 'core disgust.'

or infection, such as decaying or rotten foods, body products (e.g. blood, feces, vomit, viscera), body envelope violations, and certain sexual practices (e.g. incest, bestiality).

We now know how to establish whether some emotional response is a genuine subtype of generic disgust. We should expect the emotional response to share similarities with generic disgust when it comes to its phenomenology (e.g. a feeling of revulsion, and corresponding physiological signs), facial expression (e.g. the disgust face, with a wrinkled nose, curled lip, and possibly gaping mouth), and characteristic action tendencies in response to its elicitors (e.g. avoidance and aversion, and a tendency to describe the elicitor using disgust language).

2.3 What Counts As Genuinely Moral?

Condition (2) requires that, for a given emotional response to count as a form of genuinely moral disgust, it must be *genuinely moral*. But determining whether a given emotional emotion is an instance of genuinely moral disgust requires us to confront a difficult problem: how to delineate the boundaries of the moral domain.⁵ As Daniel Kelly notes, the term ‘moral disgust’ at least suggests that the domain of morality can be demarcated clearly, and that ‘disgust [of some kind] sometimes operates comfortably within its purview’ (2011: 126). Yet, given the lack of consensus about how to properly determine what belongs in the moral domain, Kelly concludes, ‘there is not yet any account of how to *precisely* demarcate the domain of morality, so there is not yet any way to separate out instances of genuinely *moral* disgust from others’ (128, emphasis in original).

I agree with Kelly that we don’t yet have a good account of how to *precisely* delineate the boundaries of the moral domain such that we can clearly distinguish between instances of genuinely moral disgust and non-moral generic disgust in *every case*. However, granting this much does not require us to accept the second part of Kelly’s claim: that we can’t make *any* distinctions between instances of genuinely moral disgust and non-moral generic disgust. In this section, I provide some criteria by which we can do just that.

To satisfy condition (2), I propose that instances of genuinely moral disgust must be demonstrably differentiable from generic disgust in (at least) the following ways. First, the two kinds of disgust will typically have different concrete elicitors (i.e. they will be directed at different particular objects).⁶ We already know that generic disgust is typically elicited by non-moral stimuli. By contrast, we should expect moral disgust to be typically elicited by distinctively moral stimuli, i.e. some moral violations with certain morally relevant features. Importantly, it isn’t enough just to stipulate that genuinely moral disgust is moral in virtue of its being elicited by moral stimuli; after all, generic disgust could well be elicited by (some features of) moral objects, too. With this in mind, I take this

⁵ It should be noted that both Kelly and I are conceiving of the moral domain as both narrow and normative, as is common in the philosophical literature. I discuss this point further on p. 8.

⁶ The ‘typically’ qualifier used above does not rule out the possibility that a particular object (or property of that object) could elicit both forms of disgust (assuming that moral disgust does exist). I thank an anonymous reviewer for raising this point.

condition to require evidence that moral disgust responds to some particular subset of moral violations *as such*. That is, the emotional response must properly respond to the morally relevant properties⁷ in virtue of which the moral violations are wrong *by the lights of the agent*.⁸ To successfully meet this condition, the moral disgust advocate will also need to provide a story about why their version of moral disgust is well-suited to responding to the relevant moral properties.

We should also differentiate moral disgust from generic disgust by looking at their respective goals, functional roles and corresponding action tendencies. Generic disgust's functional role is non-moral: its role is to motivate the avoidance of things that represent a threat of infection. The goal of generic disgust is to ensure that one stays free of contamination, so avoidance of the possible contaminant may also be accompanied by cleansing or decontamination behaviors. Contrastingly, the functional role of moral disgust is, plausibly, to motivate the avoidance of or aversion to particular kinds of moral violations as well as the agents who perpetrate those wrongful actions. Given the nature and function of generic disgust, if there is such a thing as moral disgust, it would be reasonable to think that its goal is to avoid and stop the spread of *moral* contamination. Moral disgust's goal will thus be associated with action tendencies that make it more likely that agents will morally evaluate the violation and the perpetrator. Perhaps moral disgust will also give rise to thoughts about moral contamination, and motivate distinctively moral kinds of decontamination behaviors.⁹

Drawing these contrasts between the generic and moral versions of disgust requires that the boundaries between the moral and the non-moral can be clearly demarcated, which is not an easy task. It's worth noting that, in the philosophical literature, philosophers mostly proceed on the assumption that there *is* a solution to this problem rather than offering novel solutions of their own. If anyone must bear the burden of explaining what counts as moral, I suggest that it is defenders of moral disgust, who (explicitly or, more commonly, implicitly) posit the existence of a special, distinctively moral emotion. This burden-shifting move notwithstanding, for current purposes I'm still in need of a working account of what characterizes some violations as genuinely moral. To that end, I will borrow from the literature regarding the moral/conventional distinction, which attempts to identify

⁷ Strictly speaking, these properties won't themselves be moral – rather, they will be non-moral properties on which moral properties supervene, such as harm, loss of property without consent, failure to cooperate, and so on. For the sake of expedience, in what follows I will refer to properties on which moral properties supervene just as moral.

⁸ I include this 'by the lights of the agent' stipulation to allow for the realistic possibility that agents could erroneously perceive some action as morally wrong – assuming that there is a fact of the matter about which actions are morally wrong – and respond to it as such with moral disgust. For example, imagine someone who experiences moral disgust in response to the prospect of two men having sex. Homosexual sex is not morally wrong, but it could nevertheless elicit genuine albeit *unfitting* moral disgust (assuming there is such a thing). My proposed condition is intentionally formulated to leave this possibility open. I discuss this specific case and its implications for my view further below.

⁹ Kumar (2017: 17) suggests something along these lines when he raises the possibility that someone who is morally disgusted by her own actions may be motivated to reform her behavior to 'cleanse' herself. I suggest that other self-cleansing behaviors could include confessing her sins or performing ritual ablutions (if she is religiously-minded), or engaging in restitution.

particular properties that mark certain violations as genuinely moral, as opposed to merely conventional.¹⁰

Just about all philosophers (and some psychologists) think that a minimal necessary condition of the moral is that it can be distinguished from the conventional, a view with which I agree. However, at this point I should mention an alternative, influential view (which is particularly influential in the psychology literature): the view that purity norms *are* moral norms, and not merely conventional norms (Haidt 2012; Haidt and Graham 2007; Rozin et al. 1999; Haidt et al. 1997; Shweder et al. 1997; Shweder et al. 1987). Purity norms involve food and bodily objects, and the violation of purity norms typically elicits generic disgust (Haidt 2012; Rozin et al. 1999). Thus, on this view, most (if not all) disgust just *is* moral disgust. However, what Haidt and others are doing in their wide (and, by their own lights, merely descriptive) circumscription of morality is not what philosophers take themselves to be doing: we have something narrower and explicitly normative in mind. So, for present purposes I take myself to be justified in setting wide/descriptive accounts of morality aside, and in insisting on a distinction between the moral and the merely conventional, and on the related distinction between generic disgust and moral disgust. I suggest that if commonsense judgment agrees that (some of) the properties picked out by the moral/conventional literature are good candidates for the distinctively-moral-job, then that will enable me to proceed even in the absence of a comprehensive theory of what morality is.

Here is what the moral/conventional literature tells us: moral violations are defined by their consequences for the rights and welfare of others and typically involve harm being done to a victim (e.g. hitting someone or pulling their hair). On the other hand, conventional violations are defined as violations of the behavioral norms that typically operate within social systems (e.g. chewing gum at school). Unlike moral transgressions, the wrongness of conventional violations is considered authority-dependent (i.e. it is only bad to chew gum at school because the teacher says so).¹¹ It is commonly thought that the ability to make the moral/conventional distinction is basic to moral competence. The ability to make the moral/conventional

¹⁰ In what follows, I restrict my focus to approaches that seek to define morally wrong *actions*. There is another influential approach relevant to the broader question of defining morality – the tradition of virtue ethics – which focuses on character rather than on action. However, by virtue ethicists' own lights, this tradition does not sit easily with *deontic* concepts like right/wrong action and instead works best with *aretaic* concepts defined in terms of virtue and vice, which are notably broader in scope than deontic notions. So, for the purposes of the current paper, I will mostly be setting this character-based approach aside. For a more extensive discussion of the link between disgust responses, bodily moral violations, and inferences about bad character, see Giner-Sorolla et al. (2018). I thank Roger Giner-Sorolla for raising this point.

¹¹ An anonymous reviewer helpfully pointed out that defining conventional violations as authority-dependent is question-begging, as illustrated by the following example: perhaps the teacher, the relevant authority on chewing gum, says chewing gum is not permitted precisely *because* it is morally wrong (e.g. maybe it is upsetting, unpleasant to see/hear, or otherwise annoying to others, and it's wrong to upset or annoy people when we could easily avoid doing so). I am very sympathetic to this point, and following Shoemaker (2011), note that it may stem from a problematic foundational assumption of the task itself: i.e. that there *is* a distinction between moral and conventional transgressions, and this distinction is based (partly) on the distinction between authority-independence and authority-dependence. Indeed, this gives rise to a related question-begging complaint: suppose one were a divine command theorist who thought that the wrongness (and rightness) of actions depended on God's commands. Given the assumptions of the moral/conventional distinction task, removing God's authority would render any violation of moral rules non-moral, which would in turn imply, implausibly, that the divine command theorist's moral judgments were merely conventional. These are, admittedly, difficult issues, but for present purposes, as long as the reader grants that other candidate moral properties besides authority-independence plausibly help distinguish between moral vs. non-moral/conventional violations, my argument need not hang on settling them.

distinction is present in children as young as 39 months of age in a variety of cultures, and involves recognizing that moral wrongs are more serious than conventional wrongs (Turiel 1983; Smetana 1993; Blair et al. 1995; Nichols 2004). Participants who competently make the distinction characteristically make justifications for why moral transgressions are wrong with reference to harm done to the victim. Subjects also judge moral transgressions to be impermissible even if there is no rule prohibiting the action, whereas the permissibility of conventional violations is typically thought to be (merely) rule-dependent.

Here, then, are some things we can say about the genuinely moral nature of moral violations. First, harm is morally relevant. Many moral violations are characteristically wrong in virtue of the harm such violations inflict on victims.¹² Second, certain moral violations are wrong in virtue of their infringement upon the rights of victims.¹³ With these insights in hand, I can now say that my second proposed condition requires that a given emotional response must be shown to be *genuinely moral*, in the following ways: Moral disgust must have a distinctively moral functional role and it must motivate distinctively moral action tendencies. Further, it must demonstrably and appropriately respond to (some of) the morally relevant properties of the moral violations it is elicited by, such as harm or rights infringements.

In response to my proposed second condition, it might be objected that there are classes of actions, including some violations of sexual morality, that are commonly taken to be morally wrong, yet do not cause immediate harm to anyone.¹⁴ I gladly acknowledge the general point that, while harm must be accounted for by any plausible theory of morality, harm does not fully explain all moral violations and is thus not the only moral property in town (see e.g. Haidt 2012; Haidt and Graham 2007; Rozin et al. 1999; Haidt et al. 1997; Shweder et al. 1997; Shweder et al. 1987). For example, take the case of someone who reneges – for no reason other than laziness – on a past promise to their now-deceased mother that they would scatter her ashes in a particular place. Plausibly, many people would think the person has done something morally wrong, even though the action does not obviously involve any harm. Examples like this illustrate the point that we must not define moral wrongness *solely* in terms of harm, because otherwise we could not explain morally wrong yet harmless actions.¹⁵ This point is entirely compatible with my proposal, which posits harm as one, but not the only, candidate moral property to which (putative) moral disgust might respond.

I turn now to the more specific point about perceived violations of sexual morality, such as homosexual sex. These violations are perceived (by some) as wrong, tend not to involve harm, and, importantly, elicit disgust. If these actions are perceived as moral rather than merely conventional violations, then we may have the makings for a case in favor of the existence of moral disgust, insofar as the disgust that they elicit could be moral in nature.

It should be noted that not all perceived violations of sexual morality elicit disgust to the same extent. Homosexual sex, which often involves sodomy (I note this for a reason that will become clear shortly), is taken to be disgusting far more often than adultery is, for instance –

¹² These harms are generally understood to be *direct* harms, but cf. Schein and Gray (2017), who offer an extensive review of cross-cultural variation with respect to moral values and perceptions of harm.

¹³ Here, given the considerations discussed in note 9, I am refraining from proposing a third feature of moral violations that could be drawn from the moral/conventional distinction literature: that the wrongness of moral violations is authority-independent.

¹⁴ Thanks to Roger Giner-Sorolla for pressing me on this point.

¹⁵ Indeed, one of the motivations for Kumar's view, which I discuss in the next section, is a class of non-harm-based moral violations called reciprocity violations which, Kumar argues, elicit moral disgust. So, it is all the more important that we do not rule out the possibility of such a class.

and is far less amenable to a harm-based explanation for its (putative) wrongness – which casts doubt on there being a unified class of sexual moral violations that elicit *moral* disgust in virtue of some wrong-making moral property. Further, even if all perceived violations of sexual morality were always met with disgust, that still would not establish that a distinctively moral version of disgust has been elicited rather than generic disgust. This is because the violations of sexual morality that most often bring about disgust are the ones with particularly salient physically disgusting (to some) features (Giner-Sorolla et al. 2018); again, compare sodomy (also, bestiality, or incest) which typically elicits disgust (in some), with adultery, which does not. I will take up a more general and detailed version of this argument shortly in section 3.1, but for now, let me sum up some key insights from the present discussion: Certain sexual violations are perceived by some people as both wrong and disgusting. One possible explanation of these cases, which would spell trouble for my view, is that the people are responding to these acts with genuinely moral disgust. I suggest that there is an alternate, more plausible explanation of what is going on these cases (to be spelled out further below): that people are experiencing *generic* disgust in response to some perceived or imagined bodily/physical feature of these sexual acts, and this disgust response accompanies the agent's independent judgment that these sexual acts are wrong.

As discussed above, providing a comprehensive account of what characterizes the genuinely moral domain is a mammoth job. While I maintain that I need not fully shoulder that burden, I grant that the question of how to define morality raises important questions about whether (and in virtue of what) certain actions, like sexual moral violations, should be considered moral. Though brief, I take the preceding discussion about perceived sexual moral violations to help illustrate that my proposed conditions do actually help us make some progress on these hard questions. For one thing, the comparison I drew between homosexual sex, which I take to be neither wrong nor harmful, and adultery, which is plausibly wrong *because* it is harmful, underscores the importance of harm as a morally relevant property that often can help us distinguish between moral and conventional violations, both in the sexual-moral domain, and more broadly.

3 Assessing the Candidates

Now that I've established the conditions that must be met for some emotional response to qualify as genuinely moral disgust, I will apply them to two existing accounts of moral disgust by John Kekes and Victor Kumar. It will be seen that both accounts fall afoul of the conditions, but in different ways: Kekes' view fails to meet condition (2): there is insufficient evidence that the disgust response he examines is genuinely moral. Contrastingly, Kumar's view cannot satisfy condition (1): we have reason to doubt that he has identified an emotional response that's best described as a genuine form of disgust.

3.1 Kekes' Account of Moral Disgust

Kekes claims that moral disgust is a 'reasonable reaction' to moral violations that are performed in a 'gross, flamboyant, flagrant and contemptuous manner' (1998: 105).¹⁶

¹⁶ I am happy to accept that the moral violations Kumar and Kekes are talking about in their accounts of moral disgust are genuinely moral violations. Kekes (1998: 107), for one, explicitly accepts that we should make a distinction between violations of moral norms – what he calls 'required conventions' – and non-moral conventions.

Like generic disgust, moral disgust is an emotion that is elicited by revolting stimuli (101). Although generic disgust ‘is often a matter of taste’ when it comes to things like food, smells, insects, sexual practices and jokes, certain experiences of disgust are so ‘profound’ and ‘instinctive’ that they would be uniformly experienced by anyone confronted with the relevant elicitors. For Kekes, it is these experiences of disgust that we should call ‘moral disgust’ (102). Implicit in this discussion is the idea that moral disgust is differentiable from run-of-the-mill (generic) disgust, so it is reasonable to interpret Kekes as offering a positive answer to the ontological question about the existence of genuinely moral disgust. To explain what he takes moral disgust to be, Kekes writes,

What makes some experiences of disgust *moral* is the combination of two elements. One element is the grievous and unjustified harm inflicted on a human being. The other is that it is done in a manner that outrages the sensibility of morally committed witnesses (102-103, emphasis added).

We can interpret this as an account of what makes Kekes’ moral disgust genuinely moral. In other words, we can think of the above two elements – the harm, and the ‘outrageous’ (gross, flamboyant, etc.) manner in which the action is performed – as candidates for the morally relevant properties that moral disgust responds to.

To see whether Kekes’ account meets my proposed conditions, it will be helpful to consider some specific moral violations that Kekes takes to be morally disgusting. The examples include ‘slowly disemboweling a person, dismembering someone with a chain-saw, slaughtering babies and bathing in their blood, or being drowned in excrement’ (101). It’s important to note that all of these examples are extremely physically disgusting. A large volume of blood, guts, gore, dead bodies, and feces are all clearly within the purview of generic disgust. The examples also include the morally relevant property of harm, and they do seem to be performed in an outrageous fashion. However, it seems that what makes each violation outrageous, gross, flagrant, and so on just is what’s physically disgusting about it; the generous amount of blood, guts, or feces involved, say (Giubilini 2016). So it’s not clear from these examples that what Kekes is calling moral disgust is being elicited by the distinctively moral properties of the wrongful actions rather than by the sheer volume of generically disgusting stimuli.¹⁷ In other words, we can’t (yet) say that the emotional response under discussion is a distinctively moral kind of generic disgust. So, condition (2) hasn’t been satisfied.

Unfortunately, no further clarity is forthcoming. Kekes provides some further examples in order to compare morally disgusting violations with other acts that are wrong but not disgusting. For example, shooting people dead in order to take their

¹⁷ Giubilini (2016: 243) makes a similar point, arguing that Kekes’ account, and its reliance on the (putatively) morally disgusting cases discussed below, fails to provide us with a reason for thinking that the disgust that’s elicited by moral violations is elicited by morally relevant aspects of those violations. He asks, ‘[W]hat makes the presence of excrement morally relevant, such that the disgust it elicits is not only physical disgust that happens to accompany an independently immoral act like killing, but also *moral* disgust’ (235: emphasis in original). Giubilini goes on to develop his critique of Kekes by arguing that Kekes’ account is circular because it “fails to distinguish disgust as a consequent from disgust as a moralizing emotion” (234) – in other words, he argues that Kekes fails to distinguish between fittingness and appropriateness as D’Arms & Jacobson (2000) counsel we must. Giner-Sorolla et al. (2018) also address the possibility that (putative) moral disgust only responds to physically disgusting elements of moral violations and note that much of the current literature does not adequately control for it.

money is morally wrong but not morally disgusting. The same applies to torturing captured enemy soldiers by depriving them of sleep and food. By comparison, disemboweling children and ‘[watching] them writhe as they die’ is both wrong and morally disgusting, as is mutilating someone by ‘cutting off their limbs inch by inch with a chain saw’ (105–106). All of these examples include grievous harm, yet only the latter two are morally disgusting, so harm cannot be the morally relevant property.

What about the other candidate property i.e., the ‘outrageous’ (or gross, etc.) manner in which the action is performed? Perhaps what this means is that the morally disgusting examples, but not the examples of merely wrong acts, are performed in a way that expresses a particularly objectionable attitude that is especially cruel, evil, or callous. Though this does seem like a genuinely moral way of spelling out Kekes’ second element, the worry is that it doesn’t properly distinguish the morally disgusting cases from the other examples. Plausibly, torturing people by depriving them of sleep and food also expresses extreme callousness and cruelty. I think, to echo a point made above, that the gross flamboyance, flagrance, etc. that supposedly characterizes the morally disgusting examples is, in fact, best explained by the presence of (non-moral) generically disgusting elicitors. This means, then, that Kekes’ account does not meet condition (2), and thus does not qualify as a form of genuinely moral disgust. What Kekes has given us is a set of moral violations that also happen to be very non-morally disgusting. But the compresence of moral violations and generic disgust does not suffice to make that disgust genuinely moral, and this is, I suggest, an important, general lesson for anyone who attempts to argue in favor of the existence of moral disgust.

3.2 Kumar’s Account of Moral Disgust

Victor Kumar (2017) has recently built upon what we know about the characteristic features and evolutionary story of generic disgust to develop a compelling account of moral disgust. Kumar argues that just as generic disgust tracks the threat of microbial contamination, moral disgust tracks *moral* contamination.¹⁸ Moral disgust is thus a genuinely moral response to particular moral violations because it accurately reflects, and tracks, the nature of those wrongs.

Before discussing Kumar’s view in more detail, it’s worth flagging a possible source of confusion, and defusing an objection that may stem from it. Kumar introduces his central claim in this way:

I will explain how moral disgust can be a *fitting* moral attitude. Disgust is fitting when it is evoked by moral wrongs that pollute social relationships by eroding shared expectations of trust (1).

As I argued earlier, the ontological question of whether there is in fact such a thing as genuinely moral disgust needs to be distinguished from the question of whether generic disgust can be a fitting response to moral violations. If Kumar’s view is best understood

¹⁸ Kumar’s view resembles Plakias’ account of moral disgust (2013, 2017) in some respects, though Kumar’s account is unique in its discussion of disgust’s polluting aspect as involving subversion of trust. Plakias argues that moral disgust tracks social contagion in much the same way that physical disgust tracks physical contagion.

as an account of when generic disgust is fitting as a response to certain moral violations, and he is not in fact arguing for a positive answer to the ontological question, then my criticisms here are unfounded.

However, it *is* reasonable to interpret Kumar as arguing in favor of the ontological claim. Consider this excerpt:

Moral disgust plays a causal role in our psychology that is similar to pathogen disgust, but it exhibits several differences. Most obviously, disgust has become attuned to new abstract cues: norm violations... More importantly, moral disgust motivates distancing that is social as well as physical (4).

It's clear from this passage that Kumar takes moral disgust to be a different kind of psychological state from generic (pathogen) disgust. So, I am on solid ground in taking him to be positing the existence of a distinctively moral kind of disgust. This holds, even if it turns out that Kumar, like others in the moral disgust debate, has failed to clearly distinguish the ontological question from the separate normative question about whether generic disgust is ever a fitting response to certain moral violations.¹⁹

According to Kumar, moral disgust is elicited by the class of moral violations that he calls 'reciprocity violations' (12).²⁰ Reciprocity violations include instances of *cheating* (being treated unfairly, line-cutting); *dishonesty* (hypocrisy, betrayal, disloyalty, lying); and *exploitation* (fraud, embezzlement, taking advantage). Kumar does not explicitly define reciprocity violations; rather, he takes the class of wrongs that contains cheating, dishonesty, and exploitation to have enough of a family resemblance to warrant a common label. Kumar distinguishes reciprocity violations from 'violations of norms related to harm, theft, autonomy, or special obligations' (9). Whereas reciprocity violations typically elicit disgust, these other kinds of moral violations usually elicit moral anger.

To support the claim that reciprocity violations elicit moral disgust, Kumar cites self-report data from various studies (e.g. Hutcherson and Gross 2011; Rozin et al. 1999; Nabi 2002; Tybur et al. 2009, discussed in Kumar 2017: 12) that shows that people affirm experiencing disgust in response to a wide range of moral violations actions like embezzling money from a bank and stealing from the blind. Additionally, Kumar points to research that does not rely on self-report measures that suggests that reciprocity violations in economic games²¹ elicit responses that are characteristic of generic

¹⁹ In drawing attention to this distinction between fittingness and appropriateness (and developing the list of separable questions about moral disgust in the introduction), I am indebted to D'Arms and Jacobson (2000) for their original work on the distinction between fittingness vs. appropriateness of moral emotions, and to Giubilini (2016) for his application of the fittingness vs. appropriateness distinction specifically to other extant accounts of moral disgust.

²⁰ Kumar notes that moral disgust is also elicited by in-group and purity violations, but he focuses primarily on moral disgust as a fitting response to reciprocity violations – reciprocity violations, unlike purity violations, are a purer kind of example, since they are not physically disgusting. I follow him in this regard.

²¹ The economic games Kumar discusses are the ultimatum game and the public goods game. The ultimatum game is a two-player game in which the first player is given a sum of money and instructed to offer some portion of it, however small, to the other player, who decides to either accept or reject the proposal. If they reject the proposal, neither player receives any money. In a public goods game, multiple players contribute money to a pot that is then multiplied and redistributed among all players. Though it is in everyone's best interests, collectively, to contribute all their money to the common pot, players may still defect by choosing not to contribute their fair share to the pot and benefitting from others' contributions to the pot when the money is redistributed.

disgust. For example, participants who receive low ball offers in the ultimatum game make the disgust (gape) face (Cannon et al. 2011), select the gape face as the expression that befits their experience (see Chapman and Anderson 2013 for a review), and exhibit increased activity in the anterior insula, an area of the brain correlated with disgust (Sanfey et al. 2003). People who receive low ball offers in the ultimatum game typically reject those offers because they expected a fair(er) distribution of the money, and thus perceive that they have been cheated (Kumar 2017: 12).

Kumar argues that the distinctive feature of reciprocity violations that makes moral disgust particularly well-suited as a response to them is that they tend to pollute and contaminate (Kumar also uses the term ‘circulate’). People who commit reciprocity violations subvert shared expectations of trust in benign social interactions, thereby ‘spoiling these interactions’ (13). Reciprocity violations are contaminating in the sense that once they have been committed, other people are more likely to start cheating or acting dishonestly too. Kumar explains that although we are intrinsically motivated to refrain from harming others, people typically follow norms that prohibit cheating and dishonesty only so long as others follow them, too (14).

A key piece of evidence Kumar cites to support the claim that reciprocity failures are contaminating is the observation that defection – that is, free riding by hoarding one’s own money rather than contributing to the shared pot – in public goods games tends to spread rapidly across the pool of participants (Fischbacher et al. 2001; discussed in Kumar 2017: 14–15.).²² Importantly, the distinctive polluting and contaminating features of reciprocity violations help explain the kind of punishment that is motivated by moral disgust. Unlike more direct forms of punishment such as blame and confrontation, which are more likely to be motivated by anger, disgust motivates an avoidance action tendency: physical withdrawal from and social exclusion of the wrongdoer (Kumar 2017: 13–14). Kumar argues that such punishment (which is usually collectively meted out in cases where third parties are disgusted by reciprocity violators) is an appropriate response to wrongdoers whose ‘actions tend to pollute and spoil otherwise benign or positive social interaction’ (14). The punishment’s purpose is to deprive the wrongdoer of social contact and continued access to shared resources, and to contain the spread of reciprocity failure so that the wrongdoer’s bad behavior won’t ‘infect’ others. On this view, we should interpret participants’ behavior in economic games, such as rejection of low ball offers in ultimatum games and second or third-party punishment of defectors in public goods games (Fehr and Fischbacher 2004), as motivated by disgust. I will return to this point below.

Let’s consider whether Kumar’s account satisfies the first condition. To satisfy condition (1), Kumar must show that what he is calling moral disgust is a genuine subtype of generic disgust. Here he seems, at least initially, to be on safe ground; in light of the evidence he cites, the emotional response Kumar is describing appears to be sufficiently similar to generic disgust when it comes to its characteristic facial expression, brain activation, and its behavioral and motivational profile (e.g. the avoidance action tendency, self-reported feelings of disgust). As a preliminary look at condition

²² We should consider this alternative explanation of the cases where defection has already spread throughout the group: people respond to growing defection by hoarding their money, too, *not* primarily because defection is itself polluting, but because they just want to salvage what they can of their own pot of money once it becomes clear that everyone else is defecting.

(2), Kumar's account of moral disgust successfully differentiates it from generic disgust when it comes to its elicitors and its functional role. Generic disgust is not a punishing attitude, and so punishment does not show up in its functional role. By contrast, the functional role of Kumar's moral disgust – to motivate avoidance behavior in order to stop the contaminating spread of reciprocity failures – does explain why the avoidance action tendency may take the form of punishment.

Here, though, I want to challenge an assumption present in Kumar's account: that moral violations can be cleanly classified as a certain kind of violation or another. It's not clear that the kinds of wrongs Kumar has in mind are obviously distinct from harm or autonomy violations. Indeed, you might think that the violation of expectations of reciprocity is what constitutes harm in some cases. For example, consider the case of spousal infidelity. Cheating, understood in this sense, is a paradigm example of dishonesty *and* a violation of special obligations to one's partner. It's plausible to say that people who have been the victims of such cheating experience it as harmful, largely in virtue of the betrayal of trust that cheating involves.

How does this apply to the cheating that takes place in economic games, which Kumar takes to be a paradigm elicitor of moral disgust? Consider this claim from Jesse Prinz and Shaun Nichols:

When people fail to cooperate or take more than is just for themselves, they treat others as inferior or less deserving. In our society, where presumptions of equality are strongly emphasized, that is seen as a harm (Prinz and Nichols 2010: 130).

Prinz and Nichols claim that this observation rings particularly true when it comes to cases of free riding: acts of free riding are harms, and are perceived as such, because the victims are taken advantage of. If this is right, then we should question Kumar's classification of the particular wrongs he labels reciprocity violations as distinct from harm violations, autonomy violations, and violations of special obligations. This is important, because, by Kumar's own lights, violations that involve harm are more likely to elicit anger, not disgust.^{23,24}

We now have reason to doubt that the kinds of wrongs that Kumar dubs reciprocity violations are distinct from harm violations. This raises the possibility that anger, not disgust, is the primary emotional response to reciprocity violations precisely because the victims of such violations take themselves to have experienced a harm.²⁵ In what follows, I defend this claim in order to show that Kumar's account fails to satisfy condition (1). To do so, I need to more closely examine the relevant empirical evidence.

²³ Kumar has a narrower reading of harm in mind, where harm means inflicting physical (as opposed to psychological, emotional, financial, etc.) injury on a victim. However, this narrow interpretation of harm is, arguably, picking out an artificial category. People's interests are not so narrowly tied to physical injury; many would prefer to experience a small physical harm, like a paper cut, than a large emotional harm, such as being betrayed. And anger is a common response to emotional harm as well as physical harm.

²⁴ An alternative explanation is that, when one is cheated, one experiences such treatment as a form of disrespect, without necessarily taking oneself to have been harmed. If this explanation is true, it does not weaken the force of my argument, since disrespect is just as likely to elicit anger (and not disgust) as a setback to one's interests.

²⁵ Kumar acknowledges that reciprocity violations in the context of economic games like the ultimatum game, and public goods games, sometimes elicit anger as well as disgust, though he does say anger is elicited to 'a lesser degree' than disgust (9). But, as I will argue below, this concession does not do justice to the evidence.

3.3 Disgust Versus Anger: What Does the Evidence Show?

In this section, I examine the empirical evidence Kumar relies on to argue that disgust is elicited in response to reciprocity violations. Recall that Kumar points to self-report data (i.e. people say they are ‘disgusted’ by reciprocity violations), evidence about the facial expressions people make in response to or associate with reciprocity violations, and data about the brain activity of victims of reciprocity violations. Below, I briefly consider each of these sets of evidence in turn and argue that they fail to clearly establish the presence of disgust rather than anger. Next, I argue that the punishing behavior seen in economic games is better explained by anger than by disgust.

3.3.1 Disgust Language

People often employ disgust terminology when talking about moral violations. This isn’t just a quirk of English; the heavily moralized use of disgust language is seen in numerous languages besides English (Haidt et al. 1997; Rozin et al. 1999; cf. Royzman and Sabini 2001) and has been taken as evidence of the existence of a distinctively socio-moral form of disgust (Rozin et al. 1999). However, I claim that when people describe moral violations as ‘disgusting’ or use the term ‘disgust’ to refer to their emotional response to those violations, we are not licensed in inferring, as Kumar does, that they are literally disgusted.

It has been argued that disgust language is used metaphorically when talking about (non-generically disgusting) moral violations (Bloom 2005; Royzman and Sabini 2001). Consider this claim from Royzman and Sabini:

The common usage of the word ‘disgust’ in moral contexts should not seduce us into believing that, in the cases like these, the feelings on the response-side are *anything* like the feelings one is surely plunged into through an encounter with feces, slime, and severed limbs (2001: 53, my emphasis).

As I argued when spelling out the conditions for genuinely moral disgust above, we should not expect the feelings – and, I would add, behavioral responses, including the tendency to use disgust language to describe moral violations and one’s own response to them – characteristic of moral disgust to be identical to the responses that characterize generic disgust. We should, however, expect – in line with my first proposed condition – that people who use disgust language in response to moral violations are having a genuine disgust response if their language is to be taken as evidence that they are in the grip of genuinely moral disgust. As we learned from the earlier critique of Kekes, it is not sufficient for the experience of genuinely moral disgust for someone to respond with generic disgust to physically disgusting features of moral violations. But, as I will now argue, it’s not obvious that even the weaker claim that people feel generic disgust in response to reciprocity violations is adequately supported by the empirical evidence.

Research that relies on participants’ self-reports of emotional experiences is often based on the assumption that scholars and laypeople use emotion terms in a way that reflects a shared meaning. However, Nabi (2002) argues that this assumption is

unjustified when it comes to disgust, because disgust terms can be used to refer to experiences of generic disgust as well as experiences of anger. Nabi found that the emotion terms ‘angry,’ ‘disgust’ and ‘disgusted’ prompted participants to describe events like being treated unfairly, being offended, being lied to, or being cheated on (698–99).²⁶ This, by itself, doesn’t tell us that some people aren’t genuinely disgusted by such actions – perhaps some people find them angersome, and others find them generically disgusting? Yet, Nabi also found that ‘angry,’ ‘disgust’ and ‘disgusted’ were all associated with action tendencies characteristic of anger, such as the desire to retaliate or to ‘lash out’ (701). This suggests that, without more investigation into what participants mean when they use disgust language, we can’t be sure whether self-reports of disgust refer to generic disgust or anger.

To determine how disgust terms are being used when describing moral violations, Herz and Hinds (2013) examined whether participants endorsed the words ‘disgusted,’ ‘angry,’ and ‘grossed out’ when applied to moral violations. In their study, they distinguished between ‘visceral’ (i.e. generic) disgust and moral disgust. They found that the former is associated with visceral disgust language like ‘gross’ or ‘grossed out,’ but these visceral terms are not applied to moral violations, even though they are described as ‘disgusting.’ Similarly, Gutierrez et al. (2012) found that the use of disgust words in response to (non-physically disgusting) moral violations was largely predicted by anger language. These findings provide preliminary support for the claim that (putative) moral disgust is not visceral (i.e. in my terms, it’s not a genuine form of disgust); instead, the use of disgust language in response to moral violations is primarily indicative of anger.

3.3.2 Facial Expressions

As we saw earlier, Kumar infers from the fact that some participants make or select facial expressions characteristic of disgust in response to reciprocity violations as evidence that they are experiencing a genuine disgust response. However, we can question whether making a face characteristic of disgust is good evidence that someone is experiencing a disgust response rather than simply making the face for communication or signaling purposes (Royzman and Kurzban 2011; Gert 2015; Giner-Sorolla et al. 2018). Further, evidence suggests that people are not very good at identifying disgust faces. People tend to confuse disgust faces with anger faces (Widen et al. 2004) – indeed, anger and disgust are the most commonly confused pair of facial expressions (Russell 1994) – a fact that is not surprising when we consider that these facial expressions share common components, such as the bilateral upper lip raise (Rozin et al. 1999; Rozin et al. 1994). The confusion seems to extend to contempt, too: the unilateral lip raise is also associated with the characteristic contempt expression (Rozin et al. 1999), and people often mislabel photos displaying contempt faces as disgust faces (Vasquez et al. 2001). Though I do not have space here to discuss the implications of these findings in more detail, for now I hope they demonstrate that the facial expression data Kumar relies on do not establish the claim that participants are

²⁶ By contrast, the trigger terms ‘grossed out’ and ‘repulsed’ were much more likely to prompt descriptions of events or objects that are generically disgusting, such as bodily products, sexual acts, insects and rodents.

responding to reciprocity violations with disgust rather than some other emotional response.

3.3.3 Brain Activity

Kumar refers to evidence that people who have received low ball offers in the ultimatum game exhibit increased activity in the anterior insula, an area of the brain correlated with disgust (Sanfey et al. 2003, discussed in Kumar 2017: 2). But the inference from insula activation to the experience of disgust isn't so straightforward. Previous research has found that the anterior insula is involved in a range of negative or arousing emotional states besides disgust, such as fear and anger (Phan et al. 2004). Indeed, a recent meta-analysis of imaging studies found the anterior insula to be no more active during experiences of disgust than experiences of other negative emotions (Lindquist et al. 2012). This suggests that activation of the anterior insula is not uniquely associated with disgust, and thus cannot be taken as clear evidence of a disgust response.

So far in this section, I have argued that these kinds of evidence about the experience of so-called moral disgust in response to moral violations – i.e. evidence pertaining to the use of disgust language in response to moral violations, facial expressions, and activation of the anterior insula – do not conclusively establish that participants are experiencing disgust rather than anger. That the data are equivocal, though, also means that my preferred hypothesis (about anger) has not yet been established. So, since we are now at something of an impasse, I now turn to an additional set of findings that may prove more conclusive: evidence pertaining to action tendencies.

3.3.4 Behavioral Data

Following Kumar's lead, I will primarily focus on people's behavior in economic games. If disgust really is the primary emotion being elicited by bad behavior in economic games, we should expect, given the details of Kumar's account, that the kind of punishment being motivated in response to these violations, such as rejecting low ball offers in ultimatum games, or punishing defection in public goods games, is best interpreted as a characteristic disgust-motivated punishment (e.g. a withdrawal response, like social exclusion of the wrongdoer) rather than a characteristic anger-motivated punishment (e.g. retaliation towards or confrontation of the wrongdoer). Rejecting offers in ultimatum games is typically understood as a form of punishment. In what follows, I argue that the punishing behavior in economic games is best explained by anger, not disgust.

Punishments in economic games are commonly thought to be driven by negative emotions (Fehr and Gächter 2002; Xiao and Houser 2005). Participants specifically describe the punishments they applied to free riders in a public goods game as expressions of *anger* (Fehr and Gächter 2002; Fehr and Fischbacher 2004). To support the claim that anger, as opposed to disgust, drives punishing behavior, we can look to the results of a large-scale study, run by Pillutla and Murnighan (1996). They found that anger was the strongest predictor of rejections of low ball offers in ultimatum games. Anger almost always occurred in

conjunction with perceptions of unfairness, and rejections were most frequent when responders perceived the offer as unfair and also attributed full knowledge – and thus full responsibility – to the offerer.

So, it seems as if anger does motivate punishment behavior in economic games. But it's not yet obvious why. Recall that, on the view of emotions I favor, emotions have a goal, the satisfaction of which is facilitated by the emotion's characteristic action tendencies. So, we should ask: what is the goal of anger in this context, and how do the action tendencies expressed by the punishing behavior in economic games help satisfy that goal? To help answer these questions, let us look to two studies (Gollwitzer and Denzler 2009; Gollwitzer et al. 2011) which sought to clarify what goal motivates punishment behavior in economic games. In both studies, participants played a two-person public goods game with an unseen partner. After receiving an unfair offer, some of the participants had the opportunity to punish their exploitative partner by signing them up for an unpleasant follow-up task. A group of participants who did elect to punish then received a message from the wrongdoer communicating his understanding that he deserved the punishment he received, while other subjects did not receive any such message. When participants were asked how satisfied they felt after punishing the wrongdoer, Gollwitzer et al. (2011) found that punishment was satisfying for participants only if it was followed by a message from the perpetrator that showed that he was holding himself accountable for his earlier wrongdoing. Participants who got no message from the wrongdoer were dissatisfied; interestingly, they were just as dissatisfied as the participants who elected not to punish the wrongdoer at all. Taken together, these findings suggest that what people are seeking when they punish people who wrong them is not mere retribution – i.e. people are not content simply to have an outlet for their anger – nor the mere exclusion of the wrongdoer. Rather, punishing behavior is driven by a desire to hold the wrongdoer to account. The functional goal of anger in these contexts, which is achieved by its distinctive action tendencies, is to redress the perceived injustice that has taken place. I claim that anger, with its approach-oriented action tendencies, makes better sense of the punishing behavior seen in economic games than disgust.

In light of the foregoing discussion in this section, I contend that we should doubt that Kumar's account has satisfied condition (1) because the evidence does not establish that the emotional response he picks out as the response to reciprocity violations is a genuine form of disgust. Rather, I've argued that the evidence better supports my claim that *anger* responds to reciprocity violations. The arguments of this section cohere with a key insight from Giner-Sorolla et al. (2018): i.e. that many existing studies of putative moral disgust in response to non-bodily violations do not adequately control for anger (see also Russell and Giner-Sorolla 2013).

I also take the arguments of this section to point to a more general lesson which applies to anyone who argues in favor of the existence of genuinely moral disgust: to successfully shoulder the burden posed by the ontological challenge, the moral disgust proponent must provide *unequivocal* evidence that genuinely moral disgust, and not some other psychological state (notably, anger or generic disgust), is being elicited by (features of) relevant moral violations. Indeed, this lesson can be

generalized even further to anyone who seeks to defend the existence of *any* genuinely moral emotion.²⁷

4 Conclusion

In this paper, I addressed the ontological question of whether there is such a thing as genuinely moral disgust. I spelled out two conditions that any aspiring account of moral disgust must satisfy, then applied these conditions to two leading accounts of moral disgust, by John Kekes and Victor Kumar. I concluded that neither account successfully vindicated the existence of genuinely moral disgust. The failures of both views are instructive, not least because they provide a useful illustration of just how challenging it is to provide a positive answer to the ontological question that satisfies the conditions I have set forth. My focus in this paper has been moral disgust, but my arguments may generalize to discussions of the ontological question as applied to other (putative) moral emotions.

Of course, it must be acknowledged that I can't infer from the failure of just two accounts of moral disgust that no account could *ever* succeed at providing a positive answer to the ontological question. Giner-Sorolla et al. (2018), for instance, review emerging evidence from the recent literature that moral disgust is elicited by acts that are suggestive of bad character, even in the absence of any bodily/physical elicitors, *and* that it is most clearly distinguished from anger in these instances. They write,

²⁷ For the sake of argument, I want to consider the following possibility: What if some people really *do* experience reciprocity violations as morally disgusting? Kumar just needs to show that *some* people are in the grip of generic disgust in response to reciprocity violations in order to meet condition (1), which requires that a given emotional response be a genuine form of disgust for it to be an instance of genuinely moral disgust. After all, not everyone will necessarily have the same emotional response to the same elicitors. Let's grant, then, that some people are in fact responding with genuine disgust to reciprocity violations, and examine how Kumar's account fares with respect to condition (2), which requires that a given emotional response be genuinely moral for it to count as an instance of genuinely moral disgust. To successfully meet this second condition, Kumar must show that his moral disgust is appropriately tracking (what the agent takes to be) a genuinely moral property that helps explain in virtue of what reciprocity violations are wrong. Recall that what's distinctive about reciprocity violations on Kumar's account is that they are *polluting* and *contaminating*. I'll address each of these features in turn. Kumar claims that reciprocity violations 'pollute' by subverting shared expectations of reciprocity and trust during social interactions. But, as I argued earlier, reciprocity violations may well overlap with harm violations and/or violations of special obligations in this very respect. If this is right, the polluting aspect is not unique to reciprocity violations and thus cannot explain why moral disgust as opposed to anger is best suited to respond to pollution. Leaving aside pollution, Kumar's account seems to give a compelling explanation for why moral disgust is a uniquely appropriate response to contamination potency – its nature as a subspecies of generic disgust, which tracks non-moral contamination, makes it well-suited to track contamination in the moral domain. But why think that contamination potency is a morally relevant property? The mere propensity to spread is morally neutral: contagiousness is a property that both good and bad things can possess. In her account of descriptive moral disgust, which also identifies contamination potency as the key property that disgust responds to, Plakias claims that heavy drinking, a bad form of behavior, is socially contagious; people are 50% more likely to drink heavily if some of the people they are with do so (2013: 275). On the other hand, research also suggests that if people think other people are already performing some right action, they themselves are more likely to do it, too (Giubilini 2016: 238). If morally right actions can also possess the property of contamination potency, then it's hard to see how morally wrong actions can be wrong in virtue of their being potentially contaminating. Plakias herself points out that whether some act really is immoral is a separate question from the question of whether it is contaminating (2013: 276).

This evidence *indirectly* suggests that bodily moral violations elicit both bad character inferences and disgust. The link also suggests that disgust toward non-purity violations, such as harm and unfairness, might be explained by negative character inferences about the person enacting them. If disgust can be evoked from nonpurity violations, then it is a great departure from disgust's original function as a mechanism of disease avoidance... to a sociomoral mechanism that helps us avoid not only disgusting objects but disgusting people as well (Giner-Sorolla et al. 2018: 256, emphasis added).

This is indeed a promising line of research, particularly if it can ultimately vindicate the conditional claim made in the final sentence of the quote just cited. I think, though, that it does not yet establish the existence of genuinely moral disgust. My hesitation stems from the indirect nature of the evidence, which relies on a connection between character judgments and bodily/purity violations. Perhaps generic disgust is elicited by some feature(s) of bodily/moral violations and these features, which are physically disgusting, become part of the content of the appraisal of the person who performed that action. While this suggestion is still mere speculation, such an interpretation would fit with Giner-Sorolla et al.'s own suggestion that disgust operates in part to help us avoid people *whom we appraise as disgusting* without positing the existence of a special moral kind of disgust that is distinct from generic disgust.

While nothing I say here rules out the possibility that future research will provide more direct support for the existence of a distinctively moral kind of disgust, for now I think we ought to remain skeptical about this prospect. One reason for this, which was made salient by the examination of relevant empirical evidence in section 3.3, is the widespread confusion between disgust and anger. Much of the available empirical work pertaining to moral disgust does not control for anger, which in turn casts doubt on the evidential value of such empirical findings when used to argue for the existence of genuinely moral disgust.

When combined with the lesson drawn from my critique of Kekes' account – i.e. that generic disgust plus an independent moral judgment does not suffice to establish the existence of moral disgust – these considerations give us strong reason to doubt that there is a distinctive psychological state of moral disgust. I have argued that there are other, better empirically supported explanations of emotional responses to the kinds of wrongs considered in this paper that do not require positing, thereby taking on the burden of vindicating, the existence of genuinely moral disgust: such explanations rely instead on generic disgust (coupled with an independent moral judgment), or anger. If it turns out that there is no such thing as genuinely moral disgust, then discussions about its normative relevance such as those between moral disgust advocates and skeptics about its fittingness or appropriateness will at best need to be reinterpreted in terms of generic disgust or anger (or both), or, at worst, be significantly undermined.

Acknowledgements For helpful feedback on this paper, I am grateful to many people, including Sara Aronowitz, Sarah Buss, Justin D'Arms, Iskra Fileva, Joshua Gert, Dan Jacobson, Sydney Keough, Victor Kumar, Alice MacLachlan, William Melanson, Hichem Naar, Peter Railton, Nate Smith, Chandra Sripada, and two anonymous referees for this journal. Previous versions of the paper were presented to audiences at the University of Michigan, the 2018 Rocky Mountain Ethics Congress, the 2018 meeting of the Central division of the American Philosophical Association, the 2017 Biennial Conference for the International Society on

Research in Emotion, and the 2017 Omaha Workshop in the Philosophy of Emotion. Many thanks to all these audiences for valuable discussion.

References

- Blair, R.J., F.L. Clark, and M. Smith. 1995. Is the psychopath 'morally insane'? *Personality and Individual Differences* 19: 741–752.
- Bloom, P. 2005. *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.
- Bloom, P. 2013. *Just babies: The origins of good and evil*. New York: Crown Publishers.
- Cannon, P.R., S. Schnall, and M. White. 2011. Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science* 2: 325–331.
- Chapman, H.A., and A.K. Anderson. 2013. Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin* 139: 300–327.
- Chapman, H., D. Kim, J. Susskind, and A. Anderson. 2009. In bad taste: Evidence for the oral origins of moral disgust. *Science* 323: 1222–1226.
- D'Arms, J., and D. Jacobson. 2000. The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research* 61: 65–90.
- D'Arms, J., and D. Jacobson. 2003. VIII. The significance of recalcitrant emotions (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement* 52: 127–145.
- D'Arms, J. and Jacobson, D. (Forthcoming, m.s.). *Rational sentimentalism*. Oxford: Oxford University Press.
- Ekman, P., and W. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17: 124–129.
- Fehr, E., and U. Fischbacher. 2004. Third-party punishment and social norms. *Evolution and Human Behavior* 25: 63–87.
- Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. *Nature* 415: 137–140.
- Fischbacher, U., S. Gächter, and E. Fehr. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71: 397–404.
- Frijda, N. 1986. *The emotions*. Cambridge, UK: Cambridge University Press.
- Frijda, N. 2007. *The laws of emotion*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Gert, J. 2015. Disgust, moral disgust, and morality. *Journal of Moral Philosophy* 12: 33–54.
- Giner-Sorolla, R., T. Kupfer, and J. Sabo. 2018. What makes moral disgust special? An integrative functional review. In *Advances in experimental social psychology: Vol. 57. Advances in experimental social psychology*, ed. J. Olson, 223–289. Cambridge, MA: Elsevier Academic Press.
- Giubilini, A. 2016. What in the world is moral disgust? *Australasian Journal of Philosophy* 94: 227–242.
- Gollwitzer, M., and M. Denzler. 2009. What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology* 45: 840–844.
- Gollwitzer, M., M. Meder, and M. Schmitt. 2011. What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology* 41: 364–374.
- Gutierrez, R., R. Giner-Sorolla, and M. Vasiljevic. 2012. Just an anger synonym? Moral context influences predictors of disgust word use. *Cognition and Emotion* 26: 53–64.
- Haidt, J. 2012. *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon Books.
- Haidt, J., and J. Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research* 20: 98–116.
- Haidt, J., P. Rozin, C. McCauley, and S. Imada. 1997. Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies* 9: 107–131.
- Herz, R. S., and Hinds, A. 2013. Stealing is not gross: Language distinguishes visceral disgust from moral violations. *The American Journal of Psychology* 126 (3): 275–286.
- Hutcherson, C. A., and J. J. Gross. 2011. The moral emotions: A social-functional account of anger, disgust, and contempt. *Journal of Personality and Social Psychology* 100: 719–737.
- Kass, L.R. 1997. The wisdom of repugnance: Why we should ban the cloning of humans. *New Republic* 216: 17–26.
- Keke, J. 1998. *A case for conservatism*. Ithaca: Cornell University Press.
- Kelly, D. 2011. *Yuck!: The nature and moral significance of disgust*. Cambridge, MA: MIT Press.
- Kelly, D., and N. Morar. 2014. Against the yuck factor: On the ideal role of disgust in society. *Utilitas* 26: 153–177.
- Kumar, V. 2017. Foul behavior. *Philosopher's Imprint* 17 (15): 1–17.
- Lindquist, K. A., T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett. 2012. The brain basis of emotion: A meta-analytic review. *The Behavioral and Brain Sciences* 35: 121–143.

- Miller, W.I. 1997. *The anatomy of disgust*. Cambridge, MA: Harvard University Press.
- Nabi, R. L. 2002. The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition & Emotion* 16 (5): 695–703.
- Nichols, S. 2004. *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.
- Nussbaum, M.C. 2004. *Hiding from humanity: Disgust, shame, and the law*. Princeton: Princeton University Press.
- Phan, K. L. T. D. Wager, S. F. Taylor, and I. Liberzon. 2004. Functional neuroimaging studies of human emotions. *CNS Spectrums* 9 (4): 258–266.
- Pillutla, M. M., and J. K. Murnighan. 1996. Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes* 68: 208–224.
- Plakias, A. 2013. The good and the gross. *Ethical Theory and Moral Practice* 16: 261–278.
- Plakias, A. 2017. The response model of moral disgust. *Synthese* 195 (2): 1–20.
- Prinz, J. and Nichols, S. (2010). Moral emotions. In J. Doris and the moral psychology research group (Eds.), *The moral psychology handbook*. Oxford, UK: Oxford University Press, 111–146.
- Royzman, E., and R. Kurzban. 2011. Minding the metaphor: The elusive character of moral disgust. *Emotion Review* 3 (3): 269–271.
- Royzman, E., and J. Sabini. 2001. Something it takes to be an emotion: The interesting case of disgust. *Journal for the Theory of Social Behaviour* 31 (1): 29–59.
- Rozin, P., L. Lowery, and R. Ebert. 1994. Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology* 66 (5): 870–881.
- Rozin, P., L. Lowery, S. Imada, and J. Haidt. 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. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin* 115: 102–141.
- Russell, P.S., and R. Giner-Sorolla. 2013. Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological Bulletin* 139 (2): 328–351.
- Sanfey, A., J. Rilling, J. Aronson, L. Nystrom, and J. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300: 1755–1758.
- Schein, C., and K. Gray. 2017. The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review* 22: 32–70.
- Shoemaker, D. 2011. Psychopathy, responsibility, and the moral/conventional distinction. *Southern Journal of Philosophy* 49 (1): 99–124.
- Shweder, R.A., M. Mahapatra, and J. Miller. 1987. Culture and moral development. In *The emergence of morality in young children*, ed. J. Kagan and S. Lamb, 1–83. Chicago: University of Chicago Press.
- Shweder, R., N. Much, M. Mahapatra, and L. Park. 1997. The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In *Morality and health*, ed. A. Brandt and P. Rozin, 119–169. New York: Routledge.
- Smetana, J. 1993. Understanding of social rules. In *The development of social cognition: The child as psychologist*, ed. M. Bennett. New York: Guilford Press.
- Turiel, E. 1983. *The development of social knowledge: Morality and convention*. New York: Cambridge University Press.
- Tybur, J., D. Lieberman, and V. Griskevicius. 2009. Microbes, mating, and morality: Individual differences in the three functional domains of disgust. *Journal of Personality and Social Psychology* 97: 103–122. <https://doi.org/10.1037/a0015474>.
- Tybur, J., R. Kurzban, D. Lieberman, and P. DeScioli. 2013. Disgust: Evolved function and structure. *Psychological Review* 120: 65–84.
- Vasquez, K., D. Keltner, D. H. Ebenbach, and T. L. Banaszynski. 2001. Cultural variation and similarity in moral rhetorics: Voices from the Philippines and the United States. *Journal of Cross-Cultural Psychology* 32: 93–120.
- Widen, S. C., Russell, J. A., and Brooks, A. (2004). Anger and disgust: Discrete or overlapping categories. In 2004 APS Annual Convention, Boston College: Chicago.
- Xiao, E., and D. Houser. 2005. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America* 102: 7398–7401.