What Makes an Emotion Moral?

by

Mara L. Bollard

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in the University of Michigan
2018

Doctoral Committee:

Professor Daniel Jacobson, Chair
Professor Sarah Buss
Professor Peter A. Railton
Associate Professor Chandra Sripada

Mara L. Bollard

mbollard@umich.edu

ORCID iD: 0000-0002-2416-948X

Acknowledgements

I could not have completed this dissertation without the support of many people, and I regret that I cannot properly express my gratitude to everyone who helped shape this project, and my time in graduate school, in these few short pages.

First of all, tremendous thanks are due to my committee members: Daniel Jacobson, Sarah Buss, Peter Railton, and Chandra Sripada, all of whom played no small role in my decision to come to Michigan in the first place, and have continued to intellectually enthrall, challenge, and encourage me ever since. I am especially grateful to my advisor, Dan Jacobson, whose guidance, humor, and unflagging support got me, and this project, across the finish line. Special thanks, too, to Chandra Sripada, who has been a cheerful and constant advocate of my work, my teaching, and the Mind and Moral Psychology Working Group.

The research and writing of this dissertation was supported by a Mellon Recruitment Award, a Rackham One-Term Dissertation Fellowship, a Sweetland Dissertation Writing Institute Fellowship, and numerous Rackham Conference Travel Grants. I am grateful for incisive and helpful feedback on these chapters from members of the University of Michigan Mind and Moral Psychology Working Group, the University of Michigan Graduate Student Working Group, the 2016 University of Michigan Candidacy Seminar, and the 2017 Sweetland Dissertation Writing Institute. I am also grateful to audiences at the 2016 Omaha Workshop in Philosophy of Emotion, the 2016 Princeton-Michigan Graduate Student Workshop on Meta-Normativity, the 2017 International Society for Research on Emotion Biennial Conference, the 2018 Central American Philosophical Association divisional meeting, and the 2018 Society for Philosophy and Psychology annual meeting.

I am grateful to the many members of the Michigan philosophy department for teaching me so much, helping me navigate the various mysteries of grad school, sharpening my ideas, and welcoming me into a rich, warm and wonderful philosophical community. I found my intellectual home at Michigan, and I will miss it immensely. Thanks especially to Chloe Armstrong (not least for her calm formatting wizardry in the final stages), Judith Beck, Gordon Belot, Paul Boswell,

Mercy Corredor, Daniel Drucker, Anna Edmonds, Allan Gibbard, Reza Hadisi, Johann Hariman, Jim Joyce, Meena Krishnamurthy, Molly Mahony, Ishani Maitra, David Manley, Eduardo Martinez, Carson Maynard, Jean McKee, Filipa Melo Lopes, Nick Moore, Jen Nguyen, Laura Ruetsche, Cat Saint Croix, Tad Schmaltz, Patrick Shirreff, Linda Shultes, Nils-Hennes Stear, Rohan Sud, Brian Weatherson, and Robin Zheng. Special thanks to my cohort-mates, Sara Aronowitz and Sydney Keough, for being two of the staunchest and most generous supporters of my work, and of me. I am very happy to have had them as my colleagues, and even happier to call them my friends.

Thank you to Luke Russell, my first philosophy mentor, without whom I almost certainly would not have become a moral psychologist. I am also grateful to Victor Kumar for encouraging me to make something of my moral disgust doubts, and to Tom Dougherty for timely market-mentorship. Thank you to my colleagues at the University of Michigan's Center for Research on Learning and Teaching and to fellow members of the American Association for Philosophy Teachers for widening my academic world and profoundly shaping my teaching development. I am especially grateful to my students for showing me every time I step into the classroom how fun doing philosophy can be, and why it matters. I thank them for keeping me on my toes, revitalizing my philosophical interests – and, often, my spirits – and teaching me countless lessons about how to be a better teacher.

To my family, Mum, Lisa, and Adam (and Domino): thank you for reminding me who I am and where I come from, for cheering me on, and for being my favourite Cypriots, co-chefs, and co-eaters. To my dear friends in Ann Arbor, Sydney, and various places in between: thank you for being the bearers of fun, solace, sustenance and perspective, and for reminding me of what matters most. Finally, to Brent and Jasper: thank you for believing in me, for filling even the toughest times with love and laughter, for being the best beings to come home to, and for wanting to embark on this next chapter with me.

Table of Contents

Abstract

From the standpoint of both philosophers and psychologists, the study of moral psychology has undergone an affective revolution over the last three decades. This revolution has generated substantial interest in the role of the emotions in moral talk, thought, and behavior. Further, it has been claimed that some emotions are distinctively moral in nature. However, what it means for an emotion to count as moral and which emotions count as the moral ones are issues in need of further elucidation. My dissertation addresses these questions in three connected chapters, with a particular focus on two emotions often but obscurely referred to as "moral": disgust and anger.

In Chapter 1, "Is There Such a Thing as Genuinely Moral Disgust", I defend a novel, skeptical view about moral disgust. In so doing, I reject a widely-held, albeit largely implicit, assumption in the moral disgust literature that there exists a distinctive psychological state of moral disgust. To give a positive answer to what I call the *ontological* question about moral disgust, thereby vindicating its existence, I propose that a given psychological state must be shown to bear sufficient resemblance to the familiar, generic version of disgust, yet be distinguishable from it in virtue of its distinctively moral nature. I argue that existing accounts of moral disgust fail to satisfy these conditions. Further, I contend that we should be skeptical about the general prospect of giving a positive answer to the ontological question, because the empirical evidence that can be invoked in favor of moral disgust's existence is too equivocal to properly distinguish (putatively) moral disgust from other psychological states, particularly anger.

In Chapter 2, "What Makes an Emotion Moral?", I develop a novel, empirically-informed answer to the general version of the ontological question that was raised in Chapter 1 with respect to moral disgust: how can we vindicate the existence of a distinctively moral emotion? I examine two contemporary, representative accounts of the "moral" emotions, one that type-identifies the moral emotions based on their effects, and another that defines the moral emotions as those that are constituted by specifically moral judgments. I argue that the former defines the moral emotions too broadly, and thus fails to draw a substantive distinction between the moral emotions and the non-moral ones, whereas the latter defines the moral emotions too narrowly. Informed by the

problems with these accounts, I introduce a *motivational* theory of moral emotion, which defines the moral emotions as those with distinctively moral action tendencies and goals.

Finally, in Chapter 3, "In Defense of Genuinely Moral Anger", I defend the claim that there is a distinctively moral subtype of anger. I argue that moral anger is a genuine form of anger that is differentiable from generic anger primarily in virtue of its action tendencies, which are typically triggered by perceived injustice and aim to satisfy two moral goals: a communicative goal, and a retributive goal. With this account, I offer an empirically-supported account that constitutes a positive answer to the ontological question about moral anger, thereby demonstrating that it is possible to vindicate the existence of a genuinely moral emotion while making sense of the idea that the moral emotions should be understood as a recognizable subset within the general class of the emotions.

Chapter 1 – Is There Such a Thing as Genuinely Moral Disgust?

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

How precisely to interpret the disputes that have arisen 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. It seems to me 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 generic disgust, and from other psychological states? Second, there's a normative question about fittingness:[1] 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 & 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

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

D'Arms and Daniel Jacobson (2000) rightly warn against.[2] As they put it, that "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?

What does *not* seem to be at issue in debates between moral disgust advocates and skeptics, at least not by the lights of the scholars involved, is the first, ontological question. The idea that there is such a thing as a distinctive psychological state of moral disgust has mostly been implicitly assumed, but not explicitly argued for in the moral disgust literature. I find this surprising, not least because proper interpretation and investigation of the other questions 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 chapter, I investigate the previously neglected ontological question. I defend the novel skeptical view that we have strong reason to doubt there is such a thing as genuinely moral disgust.

First, though, let me briefly explain why the position I am defending in this chapter is different from existing forms of skepticism about moral disgust. The *ontological* skepticism I am advancing is distinct from the two most prominent skeptical accounts in the moral disgust literature, which are 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 generic 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. Nussbaum (2004), on the other hand, 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

[2] To give a recent example, 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.

appropriate to feel disgust in the moral domain. Though it's true that Kelly and Nussbaum may well agree with my negative answer to the ontological question and accept that there is no such thing as genuinely moral disgust, it is worth taking note at the outset of the differences in each of our skeptical targets. Arguing that there is no such thing as genuinely moral disgust, as I shall do in the present chapter, is importantly different from arguing that disgust is unreliable, or that it is not morally appropriate.

This chapter proceeds as follows. In section 2, I put forward two conditions that any 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 by the specific problems found in Kekes' and Kumar's accounts of moral disgust, I conclude by discussing why we ought to be skeptical about the general prospect of providing a positive answer to the ontological question about moral disgust.

2. What is genuinely *moral* disgust?

The ontological question at issue in this chapter is whether there is a distinctive psychological state of moral disgust. I propose that if a given emotional response is to count as genuinely moral disgust, it must satisfy the following conditions, which are individually necessary and jointly sufficient:

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

In the next two subsections, I spell out these conditions in more detail.

2.1 Interlude: Some assumptions about emotions

It is beyond the scope of the present chapter to comprehensively defend a specific theory of what emotions are.[3] However, a few words are in order about the theoretical assumptions I am making in this chapter. I assume, following a number of emotion theorists, including Nico Frijda (1986; 2007), Andrea Scarantino (2014), and Justin D'Arms & Daniel Jacobson (forthcoming),

[3] I will spell out my preferred theory of emotion in more detail in chapter 2.

that emotions[4] 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 set of action tendencies, a goal that constitutes satisfaction of the emotion, and the focusing of attention on the emotional goal. This means that emotions tend to seek precedence over and control behavior, thought, and experience.

I also assume, uncontroversially, that emotions are elicited by and directed at objects (which can include actions, agents, states of affairs, thoughts, etc.), and carry with them 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 – e.g. 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 or hunger 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 is fair to say that emotions have paradigm elicitors; that is, objects that typically elicit the relevant emotional response in most people. It is also fair to say that *if* two people perceive a given object in the same way (e.g. as fearsome, angersome, or sad, etc.), then we should expect to see them both in the grip of the same emotional syndrome.

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 is generic disgust.[5] 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, phenomenology, its motivational and behavioral profile, and perhaps also its pattern of neural activation.

[4] D'Arms & Jacobson (forthcoming; see also D'Arms & Jacobson 2003) restrict their focus to a class of emotions that they term the *natural* emotions, or "paradigmatic emotion kinds." The class of natural emotion kinds includes, but is not limited to, anger, contempt, disgust, envy, fear, joy, pride, shame, and sadness. It is worth noting that D'Arms & Jacobson's list of the natural emotions closely resembles the lists of "basic," universal, or pan-cultural emotions put forward by various psychologists "with disparate theoretical approaches," including Paul Ekman, Richard Lazarus, and J. Tooby & L. Cosmides (D'Arms & Jacobson 2003: 138, and n. 23 on the same page).

[5] What I'm calling generic disgust in this chapter is also often referred to as "physical disgust," "bodily disgust," "pathogen disgust," or "core disgust" in the philosophy and psychology literatures.

So, 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 infections. Ekman & 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"). The gape face may be 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 sense of (perceived) threat of contamination, and motivates us to withdraw from, and possibly contain the spread of, that contamination.

Given disgust's evolutionary function to protect us from contamination it is unsurprising that the paradigm elicitors of disgust are objects that present possible threats of disease 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). However, as D'Arms & Jacobson (2014) argue, it is also possible, and fitting, for stimuli to elicit disgust simply in virtue of how they look, taste, smell, or feel, even when they do not present any threat of contamination. Fudge shaped to look like feces is perfectly germ-free, but it nevertheless disgusting – and is subsequently avoided rather than eaten – in virtue of its visual similarity to feces. For similar reasons, we should not be at all surprised that hygienic stimuli such as vomit-shaped rubber molds, well-embalmed corpses, or non-toxic slime elicit disgust, and that people are thereby averse to those stimuli. The main point here is that many objects that are not contaminating may still merit disgust in virtue of their perceptual properties (D'Arms & Jacobson 2014: 273-74).

With all of this in mind, 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 possibly 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, unsurprisingly, *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. 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, precisely, to 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 attempt to 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 have different concrete elicitors (i.e. they will be directed at different particular objects). 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. Of course, it isn't enough just to stipulate that genuinely moral disgust is genuinely moral in virtue of its being elicited by moral stimuli. After all, generic disgust could well be elicited by moral objects, too. With this in mind, I take this condition to require evidence that moral disgust appropriately responds to some particular subset of moral violations *as such.* That is, the emotional response must properly respond to the

morally relevant properties[6] in virtue of which the moral violations are wrong *by the lights of the agent.* To successfully meet this condition, the moral disgust advocate will need to provide a story about why their version of (putatively) moral disgust is well-suited to responding to the relevant moral properties.

We should also try to differentiate moral disgust from generic disgust by looking at their respective goals, functional roles and corresponding action tendencies.[7] 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 de-contamination behaviors. By contrast, 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 de-contamination behaviors.[8]

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, as I noted above, is no easy task. It bears saying that, in the 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 implicitly) posit the existence of a new, distinctively moral emotion. This burden-shifting move notwithstanding, for current purposes we're 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 particular properties that mark certain violations as genuinely moral, as opposed to merely

[6] 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 these properties on which moral properties supervene as moral.

[7] I develop this method of type-identifying moral emotions in greater detail in chapter 2.

[8] 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.

conventional.[9] If our commonsense judgment agrees that the properties picked out by the moral/conventional literature are good candidates for the distinctively-moral-job, then that should suffice for now.

Here's 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). 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 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 thought to be rule-dependent.

With these considerations from the moral/conventional literature in mind, here are some (overlapping) 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. Second, certain moral violations are wrong in virtue of their infringement upon the rights of victims. Third, the wrongness of moral violations is authority-independent.[10]

To sum up, the 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 distinctive

[9] Most philosophers think that a minimal, necessary condition of the moral is that it can be distinguished from the conventional. It's worth mentioning an alternative account here in order to set it aside: Jonathan Haidt (2012) has a view according to which purity norms are moral norms. Purity norms involve food and bodily objects, and the violation of purity norms typically elicits generic disgust. Thus, on this view, all disgust just is moral disgust. However, what Haidt is doing in his wide (arguably merely descriptive) circumscription of morality is not what philosophers take themselves to be doing: we have in mind something narrower. So, I take myself to be justified in insisting on a distinction between the moral and the merely conventional, and on a distinction between generic disgust and moral disgust – if there is such a thing as genuinely moral disgust.

[10] Strictly speaking, this last point directs us to a property that is not morally relevant; i.e. moral violations are *not* wrong solely in virtue of their being a breach of authority.

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. As I noted above, providing a comprehensive account of what characterizes the genuinely moral domain is a mammoth job. However, I deny that I must fully shoulder that burden, and I am hopeful that I have said enough about how to distinguish the moral from the non-moral to proceed.[11]

3. Assessing the candidates

Now that we've established the conditions that must be met for some emotional response to qualify as genuinely moral disgust, let's 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. I'll examine these accounts in turn, beginning with Kekes'.

3.1 Kekes' account of moral disgust

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

[11] I am also happy to accept that the moral violations Kumar and Kekes are talking about in their accounts of moral disgust are genuinely moral violations, even though we may not have landed on a comprehensive account of what, precisely, makes them *moral* violations as opposed to some other sort of norm violation. 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.

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 a story about what makes Kekes' moral disgust genuinely moral. In other words, we could think of the above two elements – the harm, and the "outrageous" (or gross, flamboyant, etc.) manner in which the action is performed – as candidates for the morally relevant properties that moral disgust is the proper response 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 chainsaw, slaughtering babies and bathing in their blood, or being drowned in excrement" (101). One important thing to note is that all of these examples are, quite obviously, 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 (etc.) fashion. However, it seems as though what makes each violation outrageous, gross, flagrant, etc. just is what's physically disgusting about it; the generous amount of blood, guts, or feces involved, say. 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. In other words, we can't yet say that the emotional response under discussion is a distinctively moral subtype 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 also wrong, but not disgusting. For example, shooting people dead in order to take their 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.

On a more conciliatory note, although I have argued that Kekes' discussion does not successfully establish the existence of genuinely moral disgust, I think it *does* point us towards some other, important moral category that applies to acts and agents. The actions that Kekes classifies as morally disgusting "threaten the possibility of civilized life" (106). Not only that, Kekes describes the agents who perform morally disgusting acts as "not merely wicked, but *depraved*" (my emphasis). I suggest that we should understand 'morally disgusting,' as Kekes conceives of it, as a term that applies to and describes actions and agents that belong in the class of the depraved, [12] instead of taking it as evidence of a distinctively moral subtype of disgust that is elicited by that category. When we use the term 'morally disgusting' in Kekes' sense, we are recognizing acts and/or agents that exhibit depravity, and expressing our moral condemnation of those acts and/or agents. Of course, much more needs to be said to spell out this picture, but in sum, the point I wish to make here is that we should take Kekes' account to be telling us something interesting about the moral category of the depraved. But we should not infer from his account that

[12] Cf. Luke Russell's (2014) discussion of the use of 'evil' as a secular concept. Kekes (1998: 106) states that "moral disgust is the secular equivalent of sacrilege."

he is justified in positing the existence of a distinctively moral kind of disgust. In the next subsection, we'll see how Kumar's account of moral disgust fares.

3.2 Kumar's account of moral disgust

Recently, Victor Kumar (2017) has advanced a compelling account of genuinely moral disgust. Kumar builds upon what we know about the characteristic features and evolutionary story of generic disgust to develop an account of moral disgust. Kumar holds that just as generic disgust tracks the threat of microbial contamination, moral disgust tracks *moral* contamination.[13] Moral disgust is thus a genuinely moral response to particular moral violations because it accurately reflects, and tracks, the nature of those wrongs.

Before I discuss 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 we saw 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 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* perfectly reasonable to interpret Kumar as arguing in favor of the ontological claim when we 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

[13] It's worth noticing that Kumar's view quite closely resembles Plakias' account of moral disgust (2013; 2017), though Kumar's account is unique in its discussion of disgust's polluting aspect as involving subversion of trust. Plakias (2013) argues that "moral" disgust – she uses the term descriptively – tracks social contagion in much the same way that physical disgust tracks physical contagion. In her 2017 paper, Plakias develops this view further to argue for the response model of disgust, according to which disgust is a fitting response to norm violations insofar as norm violations are potentially contagious.

> 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 subtype of disgust. This holds, even if – perhaps even *especially* 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.

According to Kumar, moral disgust is elicited by the class of moral violations that he calls "reciprocity violations" (12).[14] 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 (as Kumar understands them) elicit moral disgust, Kumar cites self-report data from various studies (e.g. Hutcherson & 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 games[15] elicit responses that are characteristic of generic 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

[14] 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.

[15] The economic games Kumar discussed 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 the 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.

face as the expression that befits their experience (see Chapman et al. 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 – i.e. such violations tend to spread, or circulate. Kumar explains that although we are intrinsically motivated to refrain from harming others, people typically follow norms that prohibit cheating and dishonesty only conditionally, i.e. so long as others follow them, too (14).

The 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.).[16][17] 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 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

[16] Kumar doesn't explicitly discuss evidence about how participants in public goods games respond to free riders when the moral threat of defection hasn't *already* spread through the group. But plausibly, on his view people will still be disgusted by defectors and motivated to socially exclude them in order to punish them.

[17] We might also 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.

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 & Fischbacher 2004), as motivated by disgust. I will return to – and dispute – this point below.

At this point, let's examine 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, as well as its behavioral and motivational profile (e.g. the avoidance action tendency, self-reported feelings of disgust). As a quick look at condition (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 very 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.

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

> 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 & Nichols 2010: 130).

Prinz and Nichols say 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.[18]

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.[19] In what follows, I seek 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 to make the case that anger, not disgust, is the primary emotional response to reciprocity violations. This is the task I turn to in the next section.

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. In what follows, I briefly consider each set of evidence in turn and argue that it fails to clearly establish the presence of disgust rather than anger. In the following section, I argue that the punishing behavior seen in economic games is better explained by anger than by disgust.

Disgust language

[18] 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 strikes me as an artificial, and not especially useful, category. People's interests are not so narrowly tied to physical injury; everyone would prefer to experience a small physical harm, like a paper cut, than a large emotional harm, such as being betrayed or taken advantage of.

[19] Kumar does acknowledge 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 (Kumar 2017: 9). But, as I will argue, this concession does not do justice to the evidence.

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 (Haidt et al. 1997; Rozin et al. 1999; cf. Royzman & Sabini 2001) and has been taken by some (e.g. Rozin et al. 1999) as evidence of the existence of a distinctively socio-moral form of disgust. However, 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 & Sabini 2001). Consider, for instance, this claim from Royzman & 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 (putative) moral disgust to be identical to the responses that characterize generic disgust. We should, however, expect 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. But it's not obvious that this is so.

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, Robin 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).[20] This, by itself, doesn't definitively tell us that some people aren't genuinely disgusted by such actions – perhaps some people find them angersome, and others find

[20] 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, than events.

them generically disgusting? (Kumar would surely take this as encouraging evidence in favor of the claim that people experience disgust in response to reciprocity violations.) 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 latter finding is suggestive of disgust terms being used interchangeably with anger terms, but further investigation into what participants mean when they use disgust language is needed to determine whether, or when, self-reports of disgust refer to experiences of generic disgust as opposed to anger.

To help determine whether disgust terms are being used interchangeably with anger when describing moral violations, Herz & 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 so-called moral disgust. 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." Herz & Hinds take their findings to suggest that (so-called) moral disgust is not visceral (i.e. in my terms, it's not a genuine form of disgust); rather, the use of disgust language in response to moral violations is indicative of anger.[21]

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 the emotional response that is disgust. 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.[22] Further, evidence suggests that people are not particularly good at identifying disgust faces. People tend to confuse disgust faces with anger faces (Widen et al. 2004), a fact that is not surprising when we consider that these facial expressions share common components. The bilateral upper lip raise occurs in both anger and disgust expressions (Rozin et al. 1999; Rozin et al. 1994). The confusion seems to

[21] See also Gutierrez et al. (2012), who found that the use of disgust words in response to (non-physically disgusting) moral violations was largely predicted by anger language.

[22] Joshua Gert has recently given a compelling account along these very lines. Gert (2015) argues that the facial expressions characteristic of disgust can be used to communicate one's attitudes towards violations of moral norms even when one *isn't* in the grip of a genuine disgust response when making the face.

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 help make clear that the facial expression data Kumar relies on do not clearly support the claim that participants are responding to reciprocity violations with disgust.

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 that is 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 easy. 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 regarding the use of disgust language in response to moral violations, facial expressions, and activation of the anterior insula – do not successfully establish that participants are experiencing disgust rather than anger. Given that we now find ourselves at something of an impasse, I suggest we shift our focus to a different set of findings that will, I hope, prove more conclusive: evidence pertaining to action tendencies. 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/free-riding in public goods games, is best interpreted as a characteristic disgust-motivated punishment (e.g. a withdrawal/avoidance response like social exclusion of the wrongdoer) 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. For present

purposes, we need to determine whether punishing behavior in economic games is best explained by disgust, or by anger. In what follows, I make the case for anger.

Punishments in economic games are commonly thought to be driven by negative emotions (Fehr & Gachter 2002; Xiao & Houser 2005). Participants specifically described the punishments they applied to free riders in a public goods game as expressions of *anger* (Fehr & Gachter 2002; Fehr & Fischbacher 2004). To support the claim that anger, as opposed to other negative emotions such as disgust, drives punishing behavior, we can look to the results of a large scale study, run by Pillutla & Murnighan (1996). Pillutla & Murnighan 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. Rejections were most frequent when responders perceived the offer as unfair and also attributed full knowledge – and thus full responsibility – to the person making the offer.

So, it seems as if anger[23] 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 can now 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 this question, let's look to two studies by Gollwitzer and colleagues (2009; 2011), which sought to clarify what goal is motivating 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. One 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 if, and only if, it was followed by a message from the perpetrator that showed that he was holding himself accountable for his earlier wrongdoing. Notably, participants who got no message from the wrongdoer were dissatisfied; interestingly,

[23] I address the question of whether the anger that's elicited as a response to bad behavior in economic games is generic anger or a distinctively moral kind of anger elsewhere in the dissertation. I argue there is a distinctively moral subtype of anger that is differentiable from generic anger, and that it is *moral* anger that drives punishing behavior in response to wrongdoing in economic games.

they were just as dissatisfied as the participants who elected not to punish the wrongdoer at all. Taken together, this suggests 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 total exclusion of the wrongdoer. Rather, punishing behavior is driven by a desire to hold the wrongdoer to account. The functional goal of anger, which is achieved by its distinctive action tendencies, is to redress the perceived injustice that has taken place. I contend that anger, which motivates approach action tendencies, makes better sense of the punishing behavior seen in economic games than disgust.

In light of all the above considerations, we should doubt that Kumar's account has satisfied condition (1) because the evidence does not successfully establish that the emotional response he is picking out as the response to reciprocity violations is a genuine form of disgust after all. Rather, I've argued that the evidence better supports my claim that anger is the primary response to reciprocity violations. The arguments of this section point to a more general lesson which applies to those who seek to argue 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, 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.

For the sake of argument, I want to consider the following possibility: What if it turns out that some people really *do* experience reciprocity violations as morally disgusting? Kumar just needs to show that *some* people are in the grip of genuine generic disgust in response to reciprocity violations in order to meet condition (1), which requires that it's necessary that a given emotional response be a genuine form of disgust for it to count as 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 must be genuinely moral for it to count as a form 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. As we saw above, 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.

By contrast, Kumar's account seems to gives 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 – i.e. the propensity to circulate – 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, Alexandra Plakias notes that heavy drinking, which is a thought of as a bad form of behavior, is socially contagious; research suggests that 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 acknowledges that whether some act really is immoral is a separate question from the question of whether that act is contaminating (2013: 276).[24]

[24] We might also object to Kumar's account and its emphasis on contamination in this way: what should we make of moral violations that are clearly contagious, yet don't seem to belong in the class of reciprocity violations or elicit moral disgust? Take speeding, for example. Insofar as speeding presents a genuine threat to others' well-being, it is best thought of as a moral rather than a conventional transgression. Both commonsense and empirical data tells us that that once people have seen other drivers breaking the speed limit, they themselves are much more likely to speed, too (Connolly & Aberg 1993). But despite speeding's contamination potency, it doesn't obviously inspire a disgust response. Nor does it neatly fit into Kumar's class of reciprocity violations. The same could be said about looting, another contagious moral violation. That there are examples of contagious-but-not-disgusting moral violations is problematic for Kumar's account. If contamination potency is supposed to explain why moral disgust is the proper response to reciprocity violations, then there is something of an explanatory hole to be filled when it comes to cases like speeding and looting.

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 the present chapter has been moral disgust, but what I have said here has clear implications for discussions of the ontological question as applied to other (putatively) 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. For all I've said here, that possibility remains open. However, though I don't have space in the present chapter to defend this claim at length, I think we ought to be skeptical about the general prospect of providing a successful account of moral disgust. 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. I venture that this confusion between disgust and anger infects both the methodological design of empirical work pertaining to moral disgust *and* scholars' interpretations of that work. Such empirical confusion, combined with the conceptual obscurity about what precisely is at issue between moral disgust advocates and skeptics I identified at the start of this chapter, gives us strong reason to doubt that there is a distinctive psychological state of moral disgust. I have tried to show that there are other, better empirically-supported explanations of people's emotional responses to the kinds of wrongs considered in this chapter that do not require positing, thereby taking on the burden of vindicating, the existence of moral disgust.

References

Blair, R. J., Clark, F. L. & Smith, M. (1995). Is the Psychopath 'Morally Insane'? *Personality & Individual Differences, 19*(5), 741-52.

Bloom, P. (2005). *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. Basic Books.

Bloom, P. (2013). *Just babies: the origins of good and evil.* New York: Crown Publishers.

Cannon, P. R.**,** Schnall, S. and White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science 2*(3): 325–331.

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin, 139*(2), 300-27.

Connolly, T., & Aberg, L. (1993). Some contagion models of speeding. *Accident Analysis and Prevention, 25*(1), 57-66.

D'Arms, J. & Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research, 61*(1), 65-90.

D'Arms, J. & Jacobson, D. (2003). "The significance of recalcitrant emotions (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement, 52,* 127-45.

D'Arms, J., & Jacobson, D. (2014). Sentimentalism and scientism. In J. D'Arms & D. Jacobson (Eds.), *Moral Psychology and Human Agency.* Oxford: Oxford University Press, 253-78.

D'Arms, J. & Jacobson, D. (forthcoming). *Rational Sentimentalism.* Oxford: Oxford University Press.

Ekman, P., & V. Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–9.

Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution & Human Behavior, 25,* 63-87.

Fehr, E. & Gachter, S. (2002). Altruistic punishment in humans. *Nature 415:* 137-40.

Fischbacher, U., Gachter, S. & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397-404.

Frijda, N. (1986). *The emotions.* Cambridge: Cambridge University Press.

Gert, J. (2015). Disgust, moral disgust, and morality. *Journal of Moral Philosophy, 12,* 33-54.

Giubilini, A. (2016). What in the world is moral disgust? *Australasian Journal of Philosophy*, *94*(2), 227-242.

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, *45*(4), 840-44.

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology*, *41*(3), 364-374.

Gutierrez, R. Giner-Sorolla, R., & Vasiljevic, M. (2012). Just an anger synonym? Moral context influences predictors of disgust word use. *Cognition & Emotion, 26* (1), 53-64.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* New York: Pantheon Books.

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*(4), 719-37.

Kass, L. R. (1997). The wisdom of repugnance: Why we should ban the cloning of humans. *New Republic, 216,* 17-26.

Kekes, J. (1998). *A case for conservatism.* Cornell University Press.

Kelly, D. (2011). *Yuck!: The nature and moral significance of disgust.* Cambridge, MA: MIT Press.

Kelly, D. & Morar, N. (2014). Against the yuck factor: On the ideal role of disgust in society. *Utilitas*, *26*(2), 153-77.

Kumar, V. (2017). Foul behavior. *Philosopher's Imprint, 17*(15), 1-17.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *The Behavioral and Brain Sciences*, *35*(3), 121–143.

Miller, W. I. (1997). *The anatomy of disgust.* Cambridge, MA: Harvard University Press.

Nichols, S. & Prinz, J. (2010). Moral emotions. In J. Doris and the Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook.* Oxford: Oxford University Press, 111-46.

Nussbaum, M. C. (2004). *Hiding from humanity: Disgust, shame, and the law.* Princeton: Princeton University Press.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, *68*(3), 208-224.

Plakias, A. (2013). The good and the gross. *Ethical Theory and Moral Practice*, *16*(2), 261-278.

Plakias, A. (2017). The response model of moral disgust. *Synthese.* https://doi.org/10.1007/s11229-017-1455-3

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*(4), 574-586.

Russell, L. (2014). *Evil: A Philosophical Investigation.* Oxford: Oxford University Press.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L. & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science,* 300: 1755-8.

Smetana, J. (1993). Understanding of social rules. In M. Bennett (Ed.), *The Development of Social Cognition: The Child as Psychologist.* New York: Guilford Press.

Turiel, E. (1983), *The Development of Social Knowledge: Morality and Convention.* Cambridge: Cambridge University Press.

Tybur, J., Lieberman, D. & Griskevicius, V. (2009). Microbes, mating, and morality: individual differences in the three functional domains of disgust. *Journal of Personality and Social Psychology*, *97*, 103-22.

Vasquez, K., Keltner, D., Ebenbach, D. H., & Banaszynski, T. L. (2001). Cultural Variation and Similarity in Moral Rhetorics: Voices from the Philippines and the United States. *Journal of Cross-Cultural Psychology*, *32*(1), 93–120.

Widen, S. C., Russell, J. A., & Brooks, A. (2004). Anger and disgust: Discrete or overlapping categories. In *2004 APS Annual Convention*, Boston College: Chicago.

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7398–7401.

Chapter 2 – What Makes an Emotion Moral?

1. Introduction

From the standpoint of both philosophers and psychologists, the study of moral psychology has undergone an *affective revolution* over the last twenty to thirty years (Cova, Deonna, & Sander 2015; Haidt 2013). This revolution has brought with it renewed interest in the role of emotions in moral talk, thought, and behavior. One of the implications of the affective revolution is that there has been a significant increase in the use of the term "moral emotions," which implies that there is a class of distinctively moral emotions that can be distinguished from the non-moral emotions (Deonna & Teroni 2012: 18). However, what scholars mean by the term "moral emotions" and which emotions should be included in the class denoted by the term remains unclear. C. Daniel Batson observes that the last twenty-five years have brought numerous claims from philosophers and psychologists about the importance of moral emotions, yet "explicit statements about what qualifies as a moral emotion have been rare" (2016: 152). Similarly, Cova and colleagues note that though the expression "moral emotions"

> is now widespread – and this surely reflects the general realization that emotions matter for morality – it is still debated what counts as a "moral emotion" and what emotions should be considered "moral" (Cova et al. 2015: 397).

In this chapter, I attempt to give a novel, empirically-informed answer to the question of what is required for an emotion to count as a distinctively moral emotion. In so doing, I hope to not only provide a general answer to the methodological question of how to type-identify emotions, but also offer an answer to the ontological question that I raised in the first chapter with respect to moral disgust: how can we vindicate the existence of a distinctively *moral* emotion? (Recall that, in that chapter, I argued that for a given emotional response to count as genuinely moral disgust, it must be a genuine form of disgust, and it must be genuinely moral. I claimed that genuinely moral disgust must bear some resemblance to the generic version of disgust, i.e. it must be in the same emotion family, yet be distinguishable from it in virtue of its genuinely moral nature, i.e. it can't be identical to the generic, non-moral version of the emotion.)

The present chapter proceeds as follows. In section 2, I examine two contemporary accounts of the "moral" emotions, one by psychologist Jonathan Haidt, and the other by philosopher Robert Solomon. I argue that Haidt's definition of moral emotions, which focuses solely on their effects, is too broad, and thus fails to draw a substantive distinction between the moral and non-moral emotions. By contrast, Solomon's account, which holds that the distinctively moral emotions are those that are constituted by a suprapersonal moral principle, yields a definition that's too narrow. I contend that Solomon's approach to defining the moral emotions fails to properly distinguish the moral emotions from the non-moral ones by his own lights. Further, the cognitivist theory of emotions that underpins his account commits him to an indefensible method of type-identifying emotions in general. In section 3, informed by the problems with Haidt's and Solomon's accounts, I draw from recent work by both philosophers and psychologists, most notably Nico Frijda (1986; 2007), Andrea Scarantino (2014), and Justin D'Arms & Daniel Jacobson (forthcoming), to develop a more plausible, empirically-informed account of the moral emotions which defines the moral emotions as those with distinctively moral action tendencies.

Before I proceed to section 2, some caveats are in order. As I discussed in the previous chapter, it is beyond the scope of the present project to comprehensively address the difficult question of how to demarcate the boundaries of the moral domain. Fortunately, my task here is much more modest: to provide a satisfying account of the moral emotions that successfully differentiates the moral emotions from the non-moral emotions in a way that does justice to the empirical nature of the emotions and to our moral psychology and practices. As I try to identify the boundary between the moral and the non-moral emotions as it is operating Haidt and Solomon's accounts – and then attempt to draw it for myself – it will be useful for us to keep the following in mind. Allan Gibbard warns in *Wise Choices, Apt Feeling* that "the term 'moral' has no sharp boundaries in normal thought" (1990: 51),[25] which helps explain why it is so hard to articulate what distinguishes the moral from the non-moral. We can begin to zero in on a useful conception of morality by thinking about it in the narrow sense, as opposed to understanding it broadly. According to Gibbard, adopting the broad conception is to understand morality as practical rationality (40). The morally right thing to do just is whatever is the practically rational thing to do (40). To put the point another way, "broadly the moral question is *how to live*" (6,

[25] Bernard Williams (1981: ix) makes a related (albeit stronger) claim in the preface to *Moral Luck* that "there cannot be any very interesting, tidy or self-contained theory of what morality is."

emphasis added). By contrast, if we understand morality more narrowly, it makes sense to ask whether what's morally right is also rational, since moral considerations are just some of the considerations that bear on the question of what to do (Gibbard 1990: 41; D'Arms & Jacobson 2000). On Gibbard's narrow view of morality, morality "concerns what is blameworthy…what is morally wrong and what is morally permissible" (1990: 51).[26] In a slightly different (and much more critical) vein, Bernard Williams (1985; 1993) argues that morality, understood narrowly, is fundamentally about moral obligation and blame.

There are, of course, other ways to (narrowly) define morality, and such conceptions may invoke concepts other than the ones mentioned so far.[27] To see what those terms might be, it is instructive, as Kwame Anthony Appiah suggests, to pay attention to the "messy array" of terms we use when talking and thinking with each other about morality, which include 'responsibilities,' 'duties,' 'rights,' 'fairness' and 'reciprocity' (2008: 159).[28] Though I cannot settle on a precise definition of 'moral' here, for present purposes the key point is that an interesting and satisfying account of the moral emotions should adopt a narrow conception of morality – and may also centrally concern at least some of the narrowly moral concepts suggested in this section – in order to draw a substantive distinction between the moral and non-moral emotions. This would not be possible against the background of a broad conception of morality; insofar as it's plausible to think that all emotions have something to do with how to live, every emotion would count as a moral emotion. So, for the purposes of the present chapter, we need to adopt a narrow conception of morality, in the hope of building towards an account that will enable us to type-identify the narrowly moral emotions.

[26] More specifically, on Gibbard's narrow reading "morality consists in norms for moral sentiments" (1990: 227), where the central moral sentiments are "guilt and resentment and their variants" (6). As will become clear later in the present chapter, I find Gibbard's view very compelling and will draw from his work on these paradigmatic, narrowly moral emotions to motivate my own view of what it means for an emotion to count as moral. For now, rather than beg any questions against Haidt and Solomon about which emotions are moral and why, I just wish to highlight Gibbard's main point that it is most helpful for present purposes to understand morality in a narrow rather than a broad sense.

[27] For example, as I discussed in the previous chapter, we might try to draw upon the moral/conventional distinction to see what, if anything is, distinctive about moral violations as compared to merely conventional transgressions. We could build upon this approach by looking to, e.g. Kurt Baier's (1958) book, *The Moral Point of View*, in which Baier provides an account an account of the distinctively moral point of view. In so doing, Baier not also distinguishes morality in the narrow sense from mere systems of customs, conventions, or laws, but also identifies several characteristics that are distinctive of morality, such as its being impartial, universal, and categorical.

[28] Appiah also mentions 'obligations.'

2. What's wrong with existing accounts of "moral" emotions?

In this section, I will examine two recent accounts of the moral emotions, by Jonathan Haidt and Robert Solomon, which each represent a different approach to defining the moral emotions (and distinguishing them from the non-moral emotions). I argue that there are serious problems with both accounts that render them untenable.

2.1 Moral emotions are beneficial to others and/or the social order

One way of understanding an emotion as distinctively moral is to claim that it is morally beneficial. Though this kind of framework has been discussed by numerous theorists (Ben-Ze'ev 1997; Ben-Ze-ev 2002; Tangney et al. 2007; Deonna and Teroni 2012; Batson 2016; Thomason 2018), for the sake of exposition I will focus on one representative and prominent version of this approach to defining the moral emotions, by the social psychologist Jonathan Haidt (2003).[29]

Haidt defines the moral emotions "as those emotions that are linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent" (2003: 853).[30] On this account, what primarily determines the *moral* nature of the moral emotions, and thus distinguishes them from the non-moral emotions are their effects. Haidt claims that emotions count as moral so long as they motivate actions which "either benefit others or else uphold or benefit the social order" (854). Haidt also notes that moral emotions are typically (but not necessarily) triggered by disinterested elicitors, such as "events that do not directly harm or benefit the self" (854).

[29] Here, I want to briefly address why I'm choosing to begin with a discussion of Haidt. First, a general reason, one I share with Krista Thomason: "philosophy does not have the market cornered on moral emotions" (Thomason 2018: 9). I take it as uncontroversial that philosophers need to engage with empirical work in the moral psychological literature in order to provide an empirically-adequate account of the moral emotions, which is my purpose here. Second, and more specifically, Haidt's work about the emotions and their role in moral psychology has been highly influential. Haidt's account of the moral emotions is often taken as definitive in discussions of moral emotions in the psychology literature (see e.g. Tangney et al. 2007), and insofar as philosophers rely on that literature for their own theorizing, it is worth examining Haidt's account directly.

[30] Beyond noting that emotions are commonly analyzed as having a number of component features, such as eliciting events (i.e. objects), a characteristic facial expression, a phenomenological experience, and a characteristic motivation or action tendency, Haidt (2003) does not say more about the theory of emotion he has in mind. In his 2012 book *The Righteous Mind*, Haidt writes that moral emotions are a type of moral intuition (2012: 45). Intuitions are "cognitions" which entails that emotions are also cognitions. Haidt does not clearly state which theory of the emotions he endorses for the purposes of the book, though he does assume that emotions involve appraisals with cognitive content (2012: 44).

Though Haidt takes himself to be addressing the question "how can we identify *the subset* of emotions that should be called moral emotions?" (853, my emphasis), at least in his initial framing of his account, he has not in fact provided an account of the moral emotions that picks out a subset of emotions, as opposed to the entire set. The definition Haidt provides is so broad that it fails to draw a substantive distinction between the moral and non-moral emotions; indeed, *any* emotion could conceivably count as moral on this view, something that Haidt himself acknowledges when he says, "any emotion that leads people to care about [the human social] world and to support, enforce, or improve its integrity should be considered a moral emotion" (855).

I contend that a definition of the moral emotions that is so capacious as to allow any emotion to count as moral is untenable, and ought to be rejected.[31] One reason for this is that Haidt's account has no good way of marking the distinction between emotions that are distinctively moral, and those that are evaluative in some broader sense. For example, consider the contempt elicited by and directed at people who fail to "eat clean" – perhaps these people who are the objects of contempt are pushing shopping carts that contain more sugary sodas than produce – or who opt for cheeseburgers over salad. In these cases, contempt functions to uphold social norms regarding food, grocery shopping, and health. But it's not clear that either contempt or the content embedded in the operative social norms are distinctively moral as opposed to evaluative in some wider sense – perhaps the soda shopper's or burger-eater's choices are simply being negatively evaluated as imprudent. We can experience negative (or positive) emotions in response to people's grocery shopping and eating habits for a variety of normative reasons, not all of them moral. Daniel Jacobson expresses this point when he writes,

> There seems to be a crucial and obvious difference between the attitude of the militant vegan (who is outraged by meat eating, which he considers tantamount to murder) and the uptown sophisticate (who holds burger eaters in contempt for their bad taste) (2008: 222).

Further, Haidt's definition of moral emotions allows what seem like substantively *immoral* emotions to count as moral. Though Haidt refers to (the promotion of) other people's welfare – a good candidate for a right-making property of action – when giving his definition of moral

[31] C. Daniel Batson (2016) criticizes Haidt's conception of the moral emotions as being too broad and imprecise. Batson thinks Haidt's definition is better thought of as picking out a class of prosocial emotions, which is larger than the class of moral emotions. I agree with Batson's general point that Haidt's conception of the moral emotions is too broad, but as will be seen below, my criticisms of Haidt's account go beyond Batson's.

emotions (853), we can still question whether upholding the social order amounts to promoting others' welfare, or has moral value in and of itself. For example, consider a cruel joke, shared at an unpopular kid's expense, that brings amusement to a large group of high schoolers. Assuming that the amusement brought about by the joke yields a net benefit overall – there's a *lot* of laughter, say, and perhaps the joke helps solidify long-lasting friendships among those who find it funny – Haidt's definition commits him to calling that amusement moral, despite the fact that amusement at someone else's expense, which very likely means humiliating them and causing them pain, is much more plausibly thought of as immoral.[32] We need not look far for other examples of emotions that bring about effects that function to uphold the social order that are deeply morally problematic: just consider the widespread disgust directed at Jews by many Germans during the Holocaust, which facilitated (and was also taken to partially justify) their extermination (Nussbaum 2004).

The key lesson to be taken from the shortcomings of Haidt's account is that an interesting and useful account of the moral emotions must allow us to draw a substantive distinction between the moral and the non-moral emotions. Attempting to distinguish the moral emotions from the non-moral ones just in virtue of their effects on other people or on the social order as a whole, as Haidt does, fails as a solution to the problem of how to type-identify the moral emotions because it relies upon a conception of 'moral' that's far too broad.

2.2 Moral emotions are constituted by moral judgments

As I have just argued, Haidt's attempt to define the moral emotions by focusing on their effects is untenable because such an approach makes it possible for any emotion to count as moral. A more promising way to identify the moral emotions is to focus on their *content*. One such method, which promises to carve out a narrower set of the emotions than a method that focuses on emotions' effects, is to ask which emotions centrally, or perhaps even constitutively, involve moral content.

There are multiple ways for an emotion to involve moral content. One way is, to borrow John Rawls' terms, to say that a moral emotion necessarily "invokes a moral concept" (1999: 421). In his discussion of what distinguishes moral emotions from the "natural" non-moral emotions and attitudes in *A Theory of Justice*, Rawls says that "it is a necessary feature of moral feelings…that

[32] Jacobson (2008: 223) makes a similar point about amusement.

the person's explanation of his experience invokes a moral concept and its associated [moral] principles." The emotions of guilt, resentment, and indignation, for instance, generally invoke the concept of right (423).

We can think of Rawls' suggestion as consistent with a cognitivist view of *moral* emotions. To get clear on what this involves, let me first introduce cognitivism about the emotions more generally. Proponents of cognitivist theories of emotion typically claim that emotions involve propositional attitudes. Cognitivists attempt to define and differentiate emotions in terms of the content contained in these attitudes, as William Lyons does when he describes cognitivism as a theory of emotion "that makes some aspect of thought, usually a belief, central to the concept of emotion and… essential to distinguishing different emotions from each other" (1980: 33). So, an emotion type is differentiable from other emotion types by the content of the propositional attitude it involves. For example, fear involves the belief that danger is present, whereas sadness involves the belief that one has suffered a loss. Rawls seems to have something like this method of individuation in mind for distinctively moral emotions when he says, "what distinguishes the moral feelings from one another are the [moral] principles…which their explanations typically invoke" (Rawls 1999: 422).

While advocates of cognitivist theories of the emotions disagree about which propositional attitudes are necessary, and/or sufficient for emotion – judgmentalists hold that the propositional attitude in question is a belief or judgment, whereas quasijudgmentalists have in mind a propositional attitude like construal, which falls short of consciously-endorsed full belief (Roberts 1988; 2007; 2013)[33] – several prominent cognitivists, including Robert Solomon (1988; 2007) and Martha Nussbaum (2001; 2004) follow the Stoic tradition of identifying emotions with judgments alone.[34] Cognitivism of this sort holds that "the beliefs involved in emotion are *constitutive*" (Nussbaum 2004: 28-29, my emphasis). There may be multiple thoughts and associated concepts that constitute a given emotion; Nussbaum claims that "emotions involve a complex family of thought" (2004: 28), Solomon thinks that most emotions are constituted by at least a dozen

[33] Some cognitivists introduce other elements into their analyses and claim that emotions are constituted by a propositional attitude cluster made up of beliefs *and* desires. See Griffiths (1997) for a helpful overview.
[34] Cognitivists need not follow Rawls in insisting that the content of the propositional attitude must be verbally expressed in an explanation of the emotion, though it is common for cognitivists who take beliefs to be necessary features of emotions to have *reportable* beliefs in mind (Griffiths 1997: 27).

judgments (2007: 213), and Robert Roberts says that each emotion contains a "package of concepts" (2013: 47).

With a version of cognitivism in mind that identifies emotions with constitutive judgments, it follows that specifically moral emotions will necessarily involve *moral* judgments. In this vein, Patricia Greenspan conceives of distinctively moral emotions as the subset of emotions "with…specifically moral content" (1998: 107 n. 7). Greenspan holds that guilt, for example, requires a constitutive thought about moral responsibility. To give another example of this sort of cognitivism applied to moral emotions, Roberts suggests that the defining proposition for indignation (i.e. moral anger) is:

> *S has very culpably offended in the important matter of X (action or omission), and is bad; I am very confident of being in a moral position to condemn; S deserves (ought) to be hurt for X; may S be hurt for X* (Roberts 2003: 215).

In what follows, I will focus specifically on Robert Solomon's (2007) cognitivist account of the emotions, with particular attention to what he has to say about the moral emotions. Solomon adopts a narrow view of morality, and provides an explicit, perspicuous statement of how to type-identify the moral emotions. Since my primary aim is to investigate the question of how to define the moral emotions *given a narrow conception of morality*, Solomon is an excellent cognitivist representative for my purposes.[35]

Solomon claims, in line with the central cognitivist commitment, that "emotions are constituted or structured by evaluative judgments" (2007: 209). Further, he holds that "an understanding of emotions… involves an understanding of the judgments that structure them… [and] it is the nature of these judgments that determines the type of emotion" (209). These constitutive judgments can differ radically between emotions, can be very precise and fine-grained, and involve quite sophisticated concepts. For example, pride and jealousy involve judgments of entitlement, and contempt involves judgments of "relative merit and status" (167).

On Solomon's view, then, identifying and distinguishing between emotion types involves identifying and differentiating between the evaluative judgments that constitute them. In line with

[35] In contrast to Solomon, both Roberts (2007; 2013) and Nussbaum (2001; 2004) seem to be operating with a broader conception of morality than I'm interested in exploring here. Roberts explicitly states that he is using "moral" in a wider sense, akin to Bernard William's (1985) notion of "ethical." Similarly, Nussbaum's brand of cognitivism identifies emotions with *eudaimonistic* judgments. On this view, all emotions are moral in a broad ("ethical") sense because all emotions are constituted by judgments about human flourishing.

the suggestion I made earlier about a cognitivist theory of specifically moral emotions, Solomon claims that distinctively moral emotions are constituted by *moral* judgments (167). More specifically, the moral emotions – a class Solomon takes to include shame, guilt, regret, envy, remorse, pride, grief, and moral indignation, some members of which I will discuss in more detail below – are the subset of the emotions that are structured by moral principles. Solomon suggests that moral principles are suprapersonal, in the sense that they tend to be impersonal and generalizable (207).

By way of illustration, Solomon discusses the paradigm case of moral indignation, which he describes as "a special kind of evaluative judgment embodying a moral principle" (207). Moral indignation is, unsurprisingly, a form of anger. It is, however, distinguishable from non-moral (what I call generic) anger in virtue of its moral content. Non-moral anger is a personal judgment that I, or someone close to me, have been offended (208). By contrast, moral indignation is "anger with suprapersonal moral clout" (40); in other words, it has an "essential moral component" (254) that generic anger lacks. More specifically, moral indignation is constituted by the judgment, "*This is wrong!"* Importantly, the "wrongness need not have anything in particular to do with me, my tastes, or my personal values" (207) – this is what makes it suprapersonal, for Solomon. For my current purposes, the invocation of wrongness is good evidence that Solomon has a narrow conception of morality in mind.

Moral indignation is also distinct from irritation and annoyance. The difference is not about the comparative severity or intensity of the affective state; Solomon rightly points out that one can be greatly annoyed, even "to the point of distraction," and that it's possible for one to experience a mild, even "cool," bout of moral indignation (207). Rather, consistent with Solomon's general theory of the emotions, the difference lies in the evaluative structure of these states, where the structure is determined by the judgments that constitute them.

Assuming that it is possible to generalize from the specific case of moral indignation, Solomon might be thought to have provided a clear method for distinguishing the moral emotions from the non-moral emotions. The difference between the two classes of emotions lies in the content embedded in the emotions' conceptual structure. Moral emotions are constituted by moral judgments, which invoke suprapersonal moral principles, whereas non-moral emotions are constituted by non-moral (yet still evaluative) judgments. To work out whether an emotion is moral, we just need to work out which moral principle its constitutive judgment contains.

However, it's not at all clear that Solomon's method for defining the moral emotions picks out all the emotions he wishes to deem moral, which is a surprisingly large collection. Recall that Solomon says that, in addition to moral indignation, the following emotions belong in the class of the moral emotions: shame, guilt, regret, remorse, envy, pride, and grief. However, besides moral indignation, which Solomon uses as the illustrative, paradigm example of a moral emotion, Solomon does not explicitly identify the moral principles that the other (putatively) moral emotions are structured by. In what follows, I will examine Solomon's analyses of three moral emotions – pride, grief, and envy – to assess whether his approach to defining the moral emotions succeeds. I argue that it does not: Solomon does not obviously have a coherent notion of which principles are moral, and in the case of grief is invoking a different sense of 'moral' than the one his cognitivist methodology for individuating emotions prescribes. It will also be seen that there are more general problems with the cognitivist method of type-identification of emotions that Solomon relies upon.

Let's begin with pride. Solomon considers pride a "positive, self-evaluative… emotion of self-praise" (90). Solomon considers pride a member of the class of emotions that "are not only emotions about the self but emotions about responsibility, including moral responsibility" (90). The evaluations that are constitutive of pride are about one having done something, or possessing some trait, that is meritorious, such as winning a race (100). The evaluations are self-ascribed, yet are what Solomon describes as "social" or "tribal," in the sense that these ascriptions depend on one's membership or status in the community. This can be seen in the example Solomon gives of out-of-shape sports fans who are "entitled to feel proud of their team insofar as they are part of the community" (100).

However, nothing in Solomon's analysis of pride suggests that a judgment embodying a principle about *moral* responsibility is constitutive of pride. And even if Solomon's discussion was faithful to his own claim about the nature of pride, the claim that pride is constituted by judgments of moral responsibility is deeply implausible. Solomon, however, does not stick to his own definition. Solomon deems the pride of the unfit sports fans in their team's victory a genuine instance of pride, despite spelling out the example in such a way that the fans have "contributed nothing to the team's victory" (100). With this detail of the case in mind, we might even doubt whether a judgment about responsibility of *any* sort is necessary.

Leaving this last objection aside, let us grant, for the sake of argument, that pride centrally involves judgments of responsibility. Consider, for example, the pride Brent feels in running a

marathon in under 3 hours, a time fast enough to gain entry to the Boston marathon in his age group. Unlike the unfit sports fans, it's clear that Brent is responsible for his (genuinely impressive) achievement, and in feeling proud of his performance in the marathon, he is positively evaluating himself. But what *moral* principle could possibly be structuring Brent's experience of pride? We can feel proud of all kinds of things we're (at least partially) responsible for; an academic can be proud of the book she wrote, a 4-year-old can be proud of having learned to tie his shoes, the individual members of the Golden State Warriors basketball team can be proud of having won back-to-back championships, and one's parents can be proud of the adult their child has grown up to become. We may even feel genuine (and fitting) pride when we haven't been directly responsible for, yet have been proximate enough to feel personally invested in, someone else's achievement, such as when one's dear friend graduates from their doctoral program, gets a great job, or overcomes a long and difficult illness with strength and grace. Though it is possible for pride to take objects with moral content – take, for example, a middle schooler who sees their classmate being cruelly made fun of and bravely steps in to stand up to the bully – it's fair to say that pride doesn't typically take moral objects or involve judgments with specifically moral content. As such, we should reject Solomon's analysis of pride and deny that pride is constituted by judgments that contain moral principles about responsibility. This is not to say that pride can't ever have anything to do with morality. But Solomon's analysis does not establish that pride is a good candidate for a distinctively moral emotion, because many instances of pride do not use moral concepts or concern anything particularly relevant to morality, understood in a narrow sense.

Let's move on to Solomon's analysis of grief. Solomon writes that grief is a *moral* emotion, in the following sense:

> Grief is not only expected as the *appropriate* reaction to the loss of a loved one, but it is in a strong sense *obligatory.* We are not just surprised when a person shows no signs of grief after a very personal loss. We are morally outraged and condemn such a person (75, emphasis in original).

The sense of "moral" Solomon is invoking in this passage is that of moral appropriateness. The experience of grief in response to the death of a loved one is not just permissible, but is morally required, and anyone who fails to have this reaction is failing to do something that is expected of them, morally speaking. However, as I discussed at length in the previous chapter, we need to take care to differentiate between the different ways an emotion can be described as "moral." To call an emotion moral, in the sense that we are morally required to feel it, is different to calling an

emotion moral in the ontological sense, which amounts to offering an answer to the question of how to vindicate the existence of a distinctively moral emotional kind. In claiming that *moral* emotions are constituted by moral judgments, Solomon is plausibly read as offering an answer to the ontological question, which requires an answer that's distinct from the positive answer given in response to the question about grief's moral appropriateness.

It might be that Solomon thinks that grief is moral in the appropriateness sense *and* also moral in the sense that it is constituted by a moral judgment. Solomon holds that the judgments that constitute grief are, roughly, that one has lost someone or something and it is a terrible loss that erodes one's sense of self (76). These judgments are evaluative, to be sure, but as was the case with Solomon's discussion of pride, they do not obviously centrally contain moral content. The loss of a loved one, perhaps the paradigmatic elicitor of grief, is in many cases terrible, as Solomon rightly says, but it is difficult to see how the life- and even identity-altering terribleness of the loss that characterizes one's experience of grief (as opposed to, say, anger at the perceived cosmic injustice of a loved one being taken too soon, even if they died naturally and peacefully) can be made sense of using moral notions like wrongness or responsibility (etc.).

Finally, let's consider Solomon's analysis of envy. Solomon claims that envy involves a negative self-evaluation, where the defining judgment is: "I do not have anything like a right or a claim to the possession or talent or honor in question" [and I want it] (102). Central to envy is the idea that one doesn't just want what someone else possesses; "it is wanting it without merit, without any intelligible claim of a right to it, without any real hope of getting it" (101).

Solomon discusses a wide variety of objects that can elicit envy: someone else's looks, their birthright, or their skills or abilities, like the ability to learn languages quickly (102). We can also envy others for the deserved honors they have received – Solomon speaks of envying Russell Crowe his Academy Award for Best Actor, even though Solomon says he himself is "not an actor and [has] never displayed any acting talent at all" (102).

The constitutive judgment(s) for envy include broadly evaluative notions of 'desert' or 'rights,' but there's nothing to suggest that these concepts can be understood in a narrowly moral way. Based on the collection of envy elicitors he suggests, it's doubtful that Solomon even intends to invoke a moral notion of rights. In speaking of my right (or lack thereof) to be as beautiful or as good a singer as Beyoncé, I'm just not talking about the same kind of thing as, say, the rights of immigrants to be treated with dignity and kindness at the US-Mexico border.

We can also question Solomon's stipulation that envy requires the thought that one has no claim to the thing one wishes to possess. Solomon is essentially making the empirical claim that someone can only be in a state of envy if they make the judgment that they want the possession, talent, honor, or trait (etc.) in question, yet do not have a right or a claim to it. But it seems very plausible for us to envy someone when, or even because, we *do* have a claim to the thing they have that we want. For example, suppose that two graduate students are equally qualified for, and thus deserving of, a coveted and lucrative fellowship. Since the candidates can't be separated on the basis of merit, the recipient of the fellowship is determined by the outcome of a coin toss. In this case, it seems as if the unlucky loser would be just as envious of the winner as the other graduate students who weren't deserving of the fellowship, if not more so, yet the "without merit" part of Solomon's definition commits him to denying that this is a real instance of envy.

Given the cognitivist theory Solomon endorses, it is his prerogative to stipulate which evaluative judgment(s) define emotion types. But this methodology is vulnerable to counterexamples, as I just demonstrated. In response to the counterexample I raised about genuine cases of envy seeming possible in the absence of a constitutive judgment that includes the part about wanting what someone else has without merit, Solomon might say that I have identified another emotion-type, differentiable from envy on the basis of its slightly-different constitutive thought, which doesn't include the "without merit" clause. This move, which involves the drawing of very fine-grained distinctions between emotions, brings with it two main problems. First, it threatens to give way to a proliferation of emotion-types, each defined by increasingly specific constitutive judgments, with each new emotion-type adding weight to the cognitivist's ontological burden. Second, and relatedly, it "threatens to turn seemingly genuine disputes over the nature of an emotion into merely terminological quarrels" (D'Arms & Jacobson 2003: 133). We should also recall that, in the previous subsection, a genuine instance of pride failed to meet Solomon's stipulated definition – i.e. pride is constituted by judgments of moral responsibility, yet the unfit sports fans who bore no responsibility, let alone moral responsibility, for their team's victory count as being in a state of pride – *by Solomon's own lights*. I take this to suggest that Solomon himself sometimes has trouble sticking to the dictates of the cognitivist methodology to which he is antecedently committed.

These considerations give us reason to object not just to Solomon's account of the moral emotions in particular, but also to the cognitivist methodology for type-identifying emotions in

general. As I will explain in more detail below, a theory of the emotions which seeks to define and individuate emotions primarily on the basis of their distinctive action tendencies – that is, how people are typically impelled to act when in the grip of a given emotion – can avoid the problems that come with the cognitivist's approach to defining emotions. Whereas cognitivists seek to determine what is (and isn't) a genuine token of an emotion type according to whether the stipulated constitutive judgment is made, the motivational theory allows us to look at the world and observe how people actually behave in order to work out whether they are experiencing a given emotion. On the motivational view – properly spelled out with appropriate details about which action tendencies and goals are characteristic of which emotion types – claims about what counts as envy (or pride, or grief, or moral indignation, and so on) can be vindicated by empirical observation. Importantly, they can also be falsified. This approach allows for genuine disputes about the nature of emotions to be settled without resorting to the questionable cognitivist move of drawing additional distinctions between emotion-types when faced with counterexamples. The motivational theory generates predictions about people's behavior that can be tested empirically, which means that we need not rely only on (possibly idiosyncratic) conceptual analysis or linguistic intuitions about the nature of emotions. As I argued in chapter 1 with respect to moral disgust, folk language about the emotions shouldn't always be taken at face value.

In light of the various problems we've encountered with Solomon's analyses of three of his (so-called) moral emotions, I claim that the distinction Solomon draws between moral emotions and non-moral emotions is deficient. As I discussed at the outset of this chapter, there are several ways to draw the moral/non-moral distinction, such as by relying on the idea that morality is centrally about wrongness, obligation, moral responsibility, duty, or fairness, etc. But only Solomon's definition of moral indignation plausibly invokes one of these notions (wrongness). Further, my discussion of Solomon's account of envy helped bring out some more general problems with his underlying cognitivist methodology for type-individuating emotions.

It might be suggested at this point that my rejection of the cognitivist approach to defining the moral emotions is too brisk. My arguments against cognitivism have centered primarily on my rejection of Solomon's analyses of particular emotions, which, one might argue, is too thin a base on which to advance the general conclusion that the cognitivist approach should be rejected. It is true that my general rejection of cognitivism has leaned heavily on my specific rejection of Solomon's cognitivist account, but as I alluded to above, I think there are good reasons to worry

about the cognitivist methodology for type-identifying the emotions in general, and the moral emotions in particular. Essentially, if the cognitivist is right, and the moral emotions are those that are constituted by a thought or proposition containing moral content, it would be possible to generate a moral variant of every single emotion type simply by adding a moral concept to every generic emotion's constitutive thought. Though cognitivism initially seemed to offer a more promising method for carving out a narrower set of moral emotions than Haidt's effects-based approach, cognitivism provides us with what turns out to be a very cheap way of defining the moral emotions that has the unwelcome implication of allowing every emotion type to be made a moral emotion by stipulation. For example, consider two cases of pride: someone who is proud at having generously donated to a worthwhile charitable cause, and another who is proud at having reached a personal fitness goal, such as lifting more weight than they've ever lifted before. The cognitivist approach to type-identifying emotions suggests that, just by building content about the moral worth of the donation into the first person's constitutive thought, they are thereby in the grip of a distinctively moral kind of pride. Moral pride is differentiable from the state of pride the weight lifter is in, which we might think just is regular pride, but risks being type-identified by the cognitivist as another kind of pride: strength pride, say. This example shows how unilluminating it is to differentiate pride at one's generosity from pride at one's strength as distinct emotion kinds according to which concepts are contained in emotions' constitutive thoughts. Instead, it is more plausible to think there is one emotion type – namely, pride – which can take different objects and generate a variety of pride tokens that involve different concepts, depending on the particular eliciting conditions.

Insofar as we care about marking a meaningful distinction between the subset of emotions that are distinctively moral and the non-moral emotions, I maintain that cognitivism should be rejected, because its methodology for type-identifying emotions allows for there to be a moral version of every emotion, and thus obscures that distinction. Since I am committed to avoiding that outcome, it had better be the case for my purposes that cognitivism is false. However, it must be acknowledged that if someone were to come along and offer a compelling response to the arguments I've made against cognitivism, I had better have something more to say, or else risk the remainder of my arguments in this chapter, and the next – which depend upon cognitivism's being false – falling through.

In the next section, I will discuss my preferred theory of emotion in more detail. One of the aims of the next section will be to establish that the motivational theory can avoid the problems faced by cognitivism and provide us with a viable way of type-identifying emotions – including the distinctively moral emotions.

3. An alternative account of the moral emotions

In section 2, I examined two accounts of the moral emotions by Jonathan Haidt and Robert Solomon and rejected both as untenable. But my discussion of these accounts has given rise to two helpful insights, which I will elaborate on in this section in order to motivate and spell out an alternative account of the moral emotions.

In section 2.1, I argued that Haidt's attempt to define the moral emotions in virtue of their effects yielded an account that failed to draw a substantive or interesting distinction between the moral emotions and the non-moral ones. Yet, Haidt's view does seem to get something important right: recall that Haidt says that emotions count as moral so long as they generate actions which benefit others, and/or uphold or benefit the social order (Haidt 2003: 854). Haidt is right to the extent that he acknowledges that the action tendencies of emotions matter when it comes to figuring out which emotions are the moral ones. However, unlike Haidt, we need to pay more attention to the character of the action tendencies that are associated with different emotions, and what they are *for*, which requires a focus that's substantially narrower than Haidt's unconstrained focus on the various effects that follow from these action tendencies. Below, I aim to show that an emotion can be characterized as a moral emotion in virtue of its having distinctively moral action tendencies. On the picture I will defend, an emotion can count as moral even if it does not bring about a net benefit to others and/or the social order.

Solomon's account (and, more broadly, the general cognitivist approach to defining moral emotions) also gets something right, in that it reflects the plausible idea that we can define the moral emotions as the subset of emotions that involve, or invoke, moral content. Where Solomon goes wrong is its adoption of the view that moral emotions are *constituted* by moral judgments. Instead, we should take seriously the possibility that moral emotions involve moral content without necessarily being constituted by moral judgments or thoughts. D'Arms and Jacobson, for example, defend the idea that emotions (in general) are quasi-perceptual, in the sense that "emotions present

things to us as having evaluative features" (2000: 72).[36] If we apply this insight to specifically moral emotions, the idea is that moral emotions perceive and represent their objects as having distinctively moral features (e.g. wrongness, unfairness, or some other moral notion(s) that should be familiar from some of our earlier discussion of narrowly moral concepts).

In the remainder of this section, I defend a motivational theory of emotion[37] according to which emotions are syndromes of feeling, appraisal, and motivation. Crucially, motivation is central to the syndrome, in the sense that emotions are primarily defined and individuated according to their motivational profile. Once a general version of the motivational view – so-named to highlight that I take the motivational component to be particularly central to the syndrome – is on the table, I will then apply it to the paradigmatic moral emotion, guilt, to demonstrate that a motivational theory of moral emotions is a viable possibility. I suggest that the application to guilt helps support my earlier claim that a motivational theory of moral emotion can capture the Haidt-inspired insight that what the emotions impel us to *do* is important for classifying them as moral (or not). Additionally, I suggest that the motivational view's appraisal component can make sense of the Solomon/cognitivist-inspired insight that moral emotions involve narrowly moral content.

At this point, I must flag that throughout this chapter and the next, I sometimes talk about the emotions as syndromes, and sometimes speak more specifically in terms of the motivational theory of emotion. It is beyond the scope of this chapter, or indeed this entire project, to comprehensively defend a novel theory of the emotions. I should admit that, in my thinking about which theory of the emotions to defend, I am pulled in two not-entirely-consistent directions. The first, weaker view of the emotions I draw on is the view that emotions are syndromes which are made up of an appraisal component, a phenomenological component, and a motivational component. The second, stronger view builds upon the first to insist that the motivational component is the most important part of the syndrome, which means that other elements are less

[36] Cf. Christine Tappolet (2016), who defends an all-out (as opposed to quasi-) perceptual theory of the emotions according to which emotions are perceptions of value.

[37] Some scholars have expressed doubt over whether the wide, varied assortment of psychological states we typically think of as emotions form a unified class that lends itself to proper definition (e.g. Rorty 1980; Griffiths 1997). Yet, most contemporary emotion theorists tend to assume that there are some shared features in virtue of which (so-called) emotions all count as emotions. As Jesse Prinz observes, "the very idea that there can be *a* theory of emotions, as opposed to several different theories, carries this presupposition" (Prinz 2004: 79, emphasis in original). Here, for the sake of simplicity I am operating under this presupposition. In so doing, I follow the lead of several emotion theorists (e.g. Frijda & Scherer 2009; Cova & Deonna 2014; Deonna & Scherer 2010).

important to establishing the existence of an emotion. A given psychological state might lack any a characteristic thought, or not be felt very strongly and still count as an emotion if the motivational element is there. By contrast, if the motivational component is missing – a possibility suggested in cases where it appears that a given emotion hasn't generated any actions – I must be able to provide a compelling explanation for why the motivation really is present, despite appearances, that isn't worryingly post hoc. If I'm unable to do that, then I must instead give up the part of the theory which accords primacy to the motivational element, and revert to the weaker view of the emotions according to which the motivational element is on equal footing with the appraisal and phenomenological components. In the remainder of this chapter, I will declare my allegiance to the stronger motivational view and say some things in its favor, while acknowledging that I am unable to mount a full defense of it. It is my hope that, if the reader finds this sketch of the motivational theory unconvincing, they will allow me to fall back on the weaker, syndrome view of emotions, which I wager can still do the work I need it to do in the present project, as long as the following anti-cognitivist commitment holds: that any thoughts involved in the emotional syndrome are *characteristic*, not constitutive.

3.1 The motivational theory of (all) emotions

Emotions are typically thought to involve "appraising a stimulus [or object] a particular way, feeling a particular way, and being motivated to act a particular way" (Scarantino 2014: 156). As Andrea Scarantino observes, emotion theorists of various stripes generally accept that emotions have these three components, while zeroing in on one particular aspect as an entry point to the question of how to explain what emotions are. For example, as we've seen, cognitivists focus on the cognitive element of the appraisal, insisting that it takes the form of a constitutive judgment, whereas perceptualists (e.g. Prinz 2004) focus on the feelings aspect of emotions. What's distinctive about the motivational theory is that it gives primacy to the motivational dimension of emotions, while still making sense of their appraisal and feeling aspects.

In endorsing a motivational theory of emotion, I am following the lead of psychologist Nico Frijda (1986; 2007), who gives the most extensive and influential treatment of the motivational theory of emotion, Andrea Scarantino (2014), who develops his own motivational account of the emotions inspired by Frijda, and Justin D'Arms and Daniel Jacobson (forthcoming),

who also follow Frijda in spelling out their motivational theory of the natural emotions.[38] I argue that the motivational view of the emotions, properly spelled out and applied to the subset of the moral emotions, enables us to identify the moral emotions as those emotions with distinctively moral action tendencies that aim at a distinctively moral goal. We can tell emotions apart, thereby distinguishing between the moral and the non-moral emotions, primarily by attending to emotions' specific goal-directed actions.

Before I get to the most distinctive aspects of the motivational theory, I shall first spell out some other features of emotions, about which there is broad consensus among emotion theorists (see, e.g. Deonna & Scherer 2010; Frijda & Scherer 2009; Roberts 1988; Gibbard 1990; de Sousa 1987; Frijda & Scherer 2009; Ben-Ze'ev 2010; Deonna & Scherer 2010; Cova & Deonna 2014; Deonna & Teroni 2012; Teroni 2007). Given that I ultimately want to provide an account that can make sense of the idea that the moral emotions are recognizable as a subset of the general class of emotions, it is worth first spending some time unpacking the key elements of a general theory of emotions.

The first important characteristic of emotions is that they are *intentional* states that are about, or directed at objects.[39] Objects can be events, actions, people, states, etc. To get at the idea that emotions are always about or directed at objects, consider that "anger is anger *about* something or *at* someone, jealousy is *directed* at a rival *over* someone, shame is shame *of* oneself *because of* some trait or act" (Deonna & Scherer, 2010: 44, emphasis in original).

Emotions can be said to have or take objects in different ways. The first way of thinking about this is to identify the *particular* object, sometimes also called the elicitor, that the emotion is about. For example, I may be sad that my favorite dumpling restaurant closed down, or you may be angry that a stranger entered your office without permission. The second way of thinking about emotions' objects is to identify the appraisal, also sometimes referred to as the *formal* object (Teroni 2007) or core relational theme (Lazarus 1991) of an emotion type.

Importantly, emotions are elicited by objects that are *relevant* to the subject, in the sense that they are linked to the subject's needs, goals, values, or well-being (Deonna & Scherer 2010). The object is appraised by the subject as having this this relevance (or value or importance). So,

[38] D'Arms and Jacobson (forthcoming) hold that the natural emotions are a subset of the emotions that are psychological kinds.

[39] I am again following Scarantino in adding the element of intentionality to the motivational theory of emotion. Intentionality is "an ingredient sorely missing from Frijda's [original] account" (Scarantino 2014: 169).

emotions represent their objects evaluatively. Different instances of an emotion-type may well be elicited by a wide variety of particular objects, but the appraisal will be the same across those instances. For example, sadness can be elicited by a break up, the death of a loved one, having to move out of a favorite apartment, or a cherished possession stolen from one's car. As different as these objects are from one another, in each case sadness appraises the situation as involving the loss of something valuable.[40]

Next, emotions are felt, and these feelings tend to reflect bodily changes or reactions (Roberts 1988; Prinz 2004; Deonna & Scherer 2010; Cova & Deonna 2014). These bodily changes include things like changes to one's breathing, heart rate, the prickling of tears, feelings of heat or warmth, goose bumps, a lump in the throat, and so on. It is notoriously tricky to give comprehensive descriptions of the phenomenology of emotions – and, indeed, for affective states in general (Cova & Deonna 2014), but for now it should suffice to note that there is something "that it is like" to be in the grip of an emotion. It also bears noting that emotions are typically experienced as unified states of mind rather than as sets of components (Roberts 1988; Gibbard 1990; Deonna & Scherer 2010.).

We're now in a position to focus on the *motivational* part of the motivational theory. According to the motivational view, emotional episodes – understood as transitory bouts elicited by an object – are goal-directed states of action readiness. Every emotion has a specific goal or aim, and every emotion involves action tendencies that aim at that goal. Emotions are better thought of as involving readiness for its goal rather than involving readiness for a specific behavior, since a number of quite different actions could be appropriate means to reaching the emotion's overall goal. The goal *unites* the disparate actions someone might take while in the grip of an emotional episode; for example, someone in the grip of fear might run, hide, or freeze. Further, the goal identifies the outcome(s) that would satisfy the emotion. For example, if running away allows the frightened person to avoid the danger they are in, the goal of that episode of fear has been achieved, and the emotion should subside.

The action readiness of emotions involves a felt urgency to act as quickly as possible to achieve the emotion's specific goal, and is characterized by (what Frijda calls) *control precedence*

[40] It's worth noting that what constitutes a loss or instance of perceived contamination (and so on) may well be in part socially constructed. As such, we should expect the particular elicitation conditions for emotions to depend on social relations and factors and to vary somewhat cross-culturally. Even so, there is substantial agreement across people and cultures when it comes to which particular objects will evoke certain emotions.

(Frijda 1986; 2007; Scarantino 2014; D'Arms & Jacobson forthcoming). This means that emotions tend to seek precedence over and control behavior, planning, thought, attention, and experience, and may interfere with or interrupt ongoing thoughts and behaviors. Emotions tend to narrow one's attentional focus, restricting access to information one already has, or could otherwise more easily get, were it not for the influence of the emotion. Emotions also prioritize their characteristic goals, which means that they persist while the goal remains unsatisfied, even in the face of interruptions or obstacles, and are insensitive "to considerations of propriety and… unwanted consequences of instigated actions" (Deonna & Scherer 2010: 51, discussing Frijda 2007). Emotions' goal prioritization also means that emotions typically motivate one to seek the most direct means to goal satisfaction, which may or may not be the most effective course of action. Lastly, an emotion's control precedence underlines the appraised relevance, value, or importance of whatever it is that elicited the emotion.

3.2 The motivational theory of *moral* emotions: the case of guilt

In the preceding section, I outlined the key features of the motivational theory of emotion. Here, to illustrate how the motivational theory works in the case of specifically moral emotions, I apply the motivational theory to what is widely considered a quintessential moral emotion: guilt (see e.g. Williams 1973; Gibbard 1990). In their survey article about moral emotions, Jesse Prinz and Shaun Nichols write that "no other emotion is more directly associated with morality" [than guilt]. Though I do not have space in the present chapter to present the many references in the philosophy and moral psychological literatures to guilt as a moral emotion, I take it that assessing how, or whether, the motivational theory can make sense of the empirical nature of a paradigmatic moral emotion – and guilt is the best emotion for this job – is a useful strategy for testing the motivational theory's adequacy. Both of my foils from the previous section, Haidt (2003) and Solomon (2007) agree that guilt counts as a moral emotion, so I am hopeful that they (and others) will allow me to assume guilt *does* count as a moral emotion for the purposes of assessing whether the motivational theory of moral emotions can capture this important datum.

Guilt typically arises when a person feels as if they've done something wrong, thereby violating a moral norm (Baumeister et al. 1994; Tangney et al. 2007; Prinz 2009; Prinz & Nichols 2010). Importantly, guilt is especially likely to arise when we take ourselves to have caused harm

to another person, and is felt most intensely when the victim is a loved one. While thoughts of responsibility for the harm are a common feature of guilt experiences, it's important to note that people can feel genuinely guilty even when they are not in fact responsible for harming anyone. For example, people who keep their jobs when others are laid off feel guilty, as do those who receive greater recognition or rewards than a peer who is equally deserving. A more extreme example is that of survivor guilt: people who survived the Holocaust, the AIDS epidemic among gay men, and other catastrophes or atrocities often report feeling guilty for surviving, especially if people they cared about were killed (Prinz & Nichols 2010; Baumeister et al. 1994).

Assuming that the self-reported guilt in these cases reflects genuine guilt experiences, these examples might be taken to suggest that guilt does not, in fact, have any special connection to moral violations or to bearing moral responsibility for causing harm. Prinz & Nichols suggest that the best way to make sense of survivor guilt cases, and other instances of guilt without (actual) moral violations, is to think of these cases of over-extensions of core guilt cases that do involve transgressions. In these cases, the violations are conceived of as omissions – i.e. failures to act that bring about harm, rather than active commission of harm. They write,

> People who experience survivor guilt feel responsible for the misery of others. They sometimes say that they should have done more to help. It seems that survivors erroneously believe that they have violated a norm; they think they were obligated to protect [the other people who didn't survive, especially] their loved ones and were in a position to do so (Prinz & Nichols 2010: 134).

We might also think about the guilt in survivor guilt cases and cases of inequity as essentially involving the ideas of desert or unfairness, both of which are narrowly more notions that have a connection to responsibility. If this is right, Prinz and Nichols suggest we should think of guilt's appraisal, or core relational theme, as something like: "someone I am concerned about has been harmed and I have responsibility for that in virtue of what I have done or failed to do" (134). There is a lot of content in this appraisal, but the motivational theory need not hold that it's necessary for people to explicitly form this thought (or something close to it) to be in a state of guilt, as cognitivist theories of guilt would. Guilt appraises its object as having certain (narrowly) moral features – i.e. harm, wrongness, responsibility – and evaluatively presents it as such to its subject. I contend that this analysis of guilt's appraisal both does justice to the empirical character of guilt and supports the claim that moral emotions involve moral content.

Further, this formulation of guilt's appraisal helps unify the varied circumstances that elicit guilt as those in which people *conceive* of themselves as somehow responsible for causing harm, even though they may not in fact have been causally responsible. The question of whether guilt is *fittingly* or correctly elicited in all these cases – we might question whether Holocaust survivors' guilt is warranted, given that they do not bear responsibility for the deaths of those they cared about[41] – is a separate question of whether guilt is genuinely elicited in these cases. The appraisal I've put forward here as characteristic of guilt helps explain the latter question, but not the former, which is as it should be, given that my target of inquiry is ontological. The issue of when and whether we're dealing with a genuine instance of an emotion is distinguishable from the issue of whether that emotion is fitting. To address the latter, we could ask whether guilt's appraisal accurately or correctly represents its object, given the specific circumstance and eliciting conditions, but, crucially, we should recognize that such a question presupposes that the emotional state in question *is* an instance of guilt.

Let's now consider what guilt's characteristic action tendencies are, and what goal those action tendencies aim at. Guilt may involve characteristic thoughts of having done wrong (Lewis 1971) and how the norm violation has affected others (Tangney et al. 2007). This captures the idea that people in the grip of guilt are more likely to focus their attention on others, e.g. by focusing on how their wrongdoing has affected the victim, rather than focusing on themselves (Tangney et al. 1994, cited in Tangney et al. 2007). This last point connects to why guilt characteristically motivates reparative actions.[42] Bernard Williams captures this when he writes, "[h]e who thinks he has done wrong may not just torment himself, he may seek to put things together again" (Williams 1973: 222). The kinds of reparative actions characteristic of guilt include confessions, apologies, forgiveness-seeking (Riek et al. 2014), and attempts to make up for or mitigate the negative consequence of the wrong or harm one has perpetrated. People may also actively seek punishment (Baumeister et al. 1994). Guilt, then, is best thought of as an other-oriented emotion, with action tendencies that motivate approach, not avoidance. It's not hard to see that these action-

[41] Along similar lines, the lorry driver who was driving safely and with care may well feel tremendous guilt for hitting the child who suddenly runs in front of the lorry, even though he has not committed a culpable wrong (Williams 1981) and may judge his guilt unfitting. See also Williams (1993).

[42] It also helps distinguish guilt from shame. Shame is associated with self-focused thoughts and attentional patterns, and tends to motivate withdrawal behaviors like hiding and avoiding others (Tangney et al. 2007).

tendencies have a distinctively moral flavor: punishment and forgiveness-seeking, apologies, and attempts to make up for harm caused to others are all elements central to our moral practices.

Tangney et al. (2007: 350) describe the action tendencies promoted by guilt as "constructive" and "proactive," in the sense that guilt motivates people to take actions aimed at repairing relationships that may have been harmed by the negative behavior. Thus, guilt aims at repairing relationships that have been damaged by the wrongdoer's harmful conduct, and we should think of the various action tendencies characteristic of guilt as attempts to satisfy that goal.

If the foregoing claims about guilt's characteristic aim are right, we should expect that guilt's distinctive aim is satisfied when the guilty party feels as if they have properly made up for the wrong they caused, and has received assurances from the victim to that effect. In line with this suggestion, research suggests that the victim's acceptance of the wrongdoer's apology leads to guilt-reduction (Prinz 2009; Baumeister et al. 1994). These points explain why survivor guilt, or guilt over a harm done to a loved who has since died, can linger, largely unabated, for a long time: if the (perceived) victim is gone, direct means to repairing the relationship by offering apologies and so on are unavailable, so the goal of guilt remains unsatisfied, and the feeling of guilt persists.[43]

4. Conclusion

In this chapter, I attempted to give a novel, empirically-adequate answer to the question of what makes an emotion distinctively moral. Informed and inspired by what I take to be two contemporary illuminating yet fatally flawed accounts of the "moral" emotions by Jonathan Haidt and Robert Solomon, I introduced a motivational theory of moral emotions, derived from a general motivational theory, which defines the moral emotions as those with distinctively moral action tendencies and goals. By applying the motivational view to the specific case of guilt, widely agreed to count as a moral emotion, I argued that the motivational theory can accommodate key insights, suitably adjusted, from Haidt's and Solomon's accounts: first, that the action tendencies of emotions matter for determining whether they are moral or non-moral, and second, that moral emotions involve (though are not constituted by) narrowly moral content. By introducing this

[43] This is not to say that guilt won't fade over time, or couldn't be ameliorated or extinguished in other ways. (Memory-erasure would do the trick, but the extinguishment of an emotion is not the same as proper satisfaction of the emotion.) The present point is that reparative actions are the most direct means to satisfying guilt's goal (Prinz & Nichols 2010), and should thus be considered guilt's distinctive action tendencies.

novel account of the moral emotions – to be applied and explored further in the next chapter with respect to moral anger – I sought to provide a plausible answer to the ontological question of how to vindicate the existence of a distinctively *moral* emotion in a way that has the resources to successfully distinguish the moral emotions from the non-moral emotions while staying true to an empirically-adequate methodology for type-identifying emotions in general.

References

Appiah, A. (2008). *Experiments in ethics*. Cambridge: Harvard University Press.

Baier, K. (1958). *The Moral Point of View.* Ithaca: Cornell University Press.

Batson, C. D. (2016). *What's wrong with morality?: A social-psychological perspective*. New York: Oxford University Press.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin, 115*: 243-267.

Ben-Ze'ev, A. (1997). Emotions and Morality. *Journal of Value Inquiry, 31*(2), 195–212.

Ben-Ze'ev, A. (2002). Are Envy, Anger, and Resentment Moral Emotions? *Philosophical Explorations, 5*(2), 148–154.

Ben-Ze'ev, A. (2010). The Thing Called Emotion. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion*. New York: Oxford University Press, 41-61.

Cova, F., & Deonna, J. A. (2014). Being moved. *Philosophical Studies, 169*(3), 447–466.

Cova, F., Deonna, J., & Sander, D. (2015). Introduction: Moral Emotions. *Topoi, 34*(2), 397–400.

D'Arms, J. & Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research, 61*(1), 65-90.

D'Arms, J. & Jacobson, D. (2003). "The significance of recalcitrant emotions (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement,* 52, 127-45.

Deigh, J. (1994). Cognitivism in the theory of emotions. *Ethics, 104*, 824-54.

Deonna, J. A., & Scherer, K. R. (2010). The Case of the Disappearing Intentional Object: Constraints on a Definition of Emotion. *Emotion Review, 2*(1), 44–52.

Deonna, J. A. & Teroni, F. (2012). *The emotions: a philosophical introduction.* New York: Routledge.

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.

Greenspan, P. S. (1998). Moral responses and moral theory: Socially-based externalist ethics. *The Journal of Ethics*, *2*(2), 103–122.

Griffiths, P. E. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*. New York: Oxford University Press, 852-870.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* New York: Pantheon Books.

Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education*, *42*(3), 281–297.

Lazarus, R. (1991). *Emotion and adaptation.* New York: Oxford University Press.

Lewis, H. B. (1971). *Shame and guilt in neurosis.* New York: International Universities Press.

Lyons, W. (1980). *Emotion.* Cambridge: Cambridge University Press.

Nichols, S. & Prinz, J. (2010). Moral emotions. In J. Doris and the Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook.* Oxford: Oxford University Press, 111-46.

Nussbaum, M. C. (2001). *Upheavals of thought: the intelligence of emotions.* Cambridge University Press.

Nussbaum, M. C. (2004). *Hiding from humanity: disgust, shame, and the law.* Princeton University Press.

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.

Prinz, J. (2009). The Moral Emotions. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion*. New York: Oxford University Press, 519-538.

Rawls, J. (1999). *A Theory of Justice.* Cambridge: Belknap Press of Harvard University Press.

Riek, B. M., Luna, L. M. R., & Schnabelrauch, C. A. (2014). Transgressors' guilt and shame: A longitudinal examination of forgiveness seeking. *Journal of Social and Personal Relationships*, *31*(6), 751–772.

Roberts, R. C. (1988). What an emotion is: a sketch. *Philosophical Review*, *97*(2), 183-209.

Roberts, R. C. (2003). *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

Roberts, R. C. (2013). *Emotions in the Moral Life*. Cambridge: Cambridge University Press.

Rorty, A. O. (Ed.). (1980). *Explaining Emotions*. Berkeley: University of California 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*(4), 574-586.

Scarantino, A. (2014). The motivational theory of emotions. In J. D'Arms & D. Jacobson (Eds.), *Moral Psychology and Human Agency.* Oxford: Oxford University Press, 156-185.

Solomon, R. (2007). *True to Our Feelings: What Our Emotions Are Really Telling Us*. Oxford & New York: Oxford University Press.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology*, *58*(1), 345–372.

Tappolet, C. (2016). *Emotions, values, and agency.* Oxford: Oxford University Press.

Teroni, F. (2007). Emotions and Formal Objects. *Dialectica*, *61*(3), 395–415.

Thomason, K. (2018). *Naked: The Dark Side of Shame and Moral Life*. New York: Oxford University Press.

Williams, B. (1973). Morality and the emotions. In *Problems of the Self: Philosophical Papers 1956-1972*. London & New York: Cambridge University Press, 207–229.

Williams, B. (1981). *Moral Luck.* Cambridge: Cambridge University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy.* Cambridge: Harvard University Press.

Williams, B. (1993). *Shame and Necessity.* Berkeley: University of California Press.

Chapter 3 – In Defense of Genuinely Moral Anger

1. Introduction

Anger is thought by many philosophers to be central to morality. Anger often occurs as a response to wrongdoing and seems to play an important role in the blaming and punishing of wrongdoers. As such, it's neither uncommon nor surprising for anger to be referred to as a moral emotion, though what precisely is meant by the term "moral anger" is not always clear; does generic, garden variety anger, which is likely familiar to us from computer malfunctions or heavy traffic, also show up in the moral domain, perhaps as a morally appropriate, fitting, or epistemically reliable response to (certain features of) wrongdoing?[44] Or is there a distinctive psychological state of *moral* anger that is differentiable from generic anger, and from other emotion types? The latter, ontological question is the target of this chapter. In what follows, I use the motivational theory of moral emotions I introduced in Chapter 2 to defend the claim that there is a distinctively moral subtype of anger. My overarching goal in this chapter is to offer a novel, empirically-supported account of moral anger that constitutes a positive answer to the ontological question about moral anger, thereby demonstrating that it is possible to vindicate the existence of a genuinely moral emotion while making sense of the idea that the moral emotions should be understood as a recognizable subset within the general class of the emotions.

The chapter proceeds as follows. In section 2, I motivate the claim that there is a distinctively moral kind of anger that is differentiable from generic anger, which typically takes non-moral objects and is characterized by action tendencies that aim to overcome goal-frustration. In section 3, I develop my motivational account of moral anger more fully. I argue that moral anger counts as distinctively moral primarily in virtue of its action tendencies, which are typically triggered by perceived injustice against oneself or others and aim to satisfy two moral goals or aims: a communicative goal, and a retributive goal. In developing my account, I draw from recent

[44] These distinctions between the various senses of "moral" as applied to anger should be familiar from my longer treatment of these issues in Chapter 1 with respect to "moral" disgust.

work by David Shoemaker (2015; 2018), who advances an illuminating account of moral anger, which he labels blaming anger. Shoemaker claims that the defining aim of blaming anger is communicative. He argues that any sanctioning or retaliatory action tendencies that are part of blaming anger are merely instrumental to what he considers its fundamental communicative aim. I argue that Shoemaker's account of moral anger is incomplete, because it does not adequately explain cases in which moral anger is partially satisfied despite its communicative aim not being met. My account of moral anger, which includes a retributive aim that can come apart from the communicative aim, is able to explain these cases, and thus extends Shoemaker's account in an important way.

2. Why think there are two kinds of anger?

In this section, I attempt to motivate the claim that there are two distinct kinds of anger: generic anger, and moral anger. I do this in two ways. First, I provide various snippets of views found in the philosophical literature that are predicated on the (mostly unargued-for) assumption that there is a special moral subtype of anger that is distinct from generic anger. Second, I examine empirical evidence regarding anger's psychological profile. It will be seen that the evidence supports the idea that there are two kinds of anger that are distinguishable from each other on the basis of their respective aims, action tendencies, appraisals, and characteristic elicitors.

It is quite common for philosophers to distinguish between the everyday variety of anger we're all familiar with (what I call generic anger) and *moral* anger. In so doing, it seems that many such theorists are implicitly assuming, though not explicitly defending, the existence of a new emotion kind that is a distinctively moral subtype of anger. For example, Victor Kumar writes, "[m]oral emotions are distinctive in the sense that they differ from various non-moral counterparts, e.g. in the way resentment differs from anger" (Kumar 2016: 789). This differentiation between generic anger and resentment is particularly prevalent in the contemporary moral responsibility literature, thanks primarily to P. F. Strawson's seminal paper, "Freedom and Resentment" (1962). In that paper, Strawson introduced a special class of emotions[45] called the "reactive attitudes." The

[45] Gary Watson (1996), for example, describes the reactive attitudes as reactive *emotions* and considers both resentment and indignation among the class of the reactive emotions. Antony Aumann & Zac Cogley describe reactive attitudes as "emotional responses to interpersonal situations" and also take anger, understood as something narrower than generic attitude, to count as a reactive emotion (n.d.: 1).

"reactive attitudes" approach, which has been taken up by R. Jay Wallace (1994) and, more recently, by David Shoemaker (2015; 2018), seeks to account for the crucial role blame plays in our moral responsibility practices by understanding blame as an essentially emotional response.

The two most widely-discussed negative reactive attitudes are resentment and moral indignation. These are typically understood as kinds of moral anger (see e.g. McKenna 2013; Pereboom 2013; Pereboom 2014; Shoemaker 2018; Scanlon 2013). Resentment is typically understood as the moral anger someone feels when they have been personally wronged by the perpetrator. Derk Pereboom expresses this view when he describes resentment as "anger with an agent due to a wrong he has done to oneself" (2014: 179). By contrast, indignation is the moral anger elicited in response to a wrong done to someone else (Shoemaker 2015; 2018; McKenna 2012). In both cases, the anger is directed at the perpetrator. David Shoemaker observes that Pereboom, along with most responsibility theorists, takes resentment (and indignation) to incorporate a constitutive judgment that what the offender did wronged the resenter (or a third party) and/or that the offender deserves blame or is responsible for having perpetrated the wrong (Shoemaker 2015). This is a *cognitivist* analysis, on which resentment and indignation are defined by a constitutive judgment about the moral responsibility, wrongdoing, and/or blameworthiness of the perpetrator. According to this account of resentment and indignation, we can think of these two emotional states as "cognitively sharpened" forms of anger (D'Arms & Jacobson 2003: 143); that is, they belong in the same emotion family as generic anger, but are plausibly differentiated from it in virtue of their (partially) constitutive judgments. Resentment is constituted by a judgment like, "You wronged me [and you're blameworthy]!" and indignation's defining judgment is, roughly, "You wronged him/her/them!" (Shoemaker 2018: 69-70).

As I argued at length in Chapter 2, the cognitivist method for type-identifying emotions faces a number of problems, especially when used in an effort to define the specifically moral emotions. I argued that moral emotions can centrally *involve* moral content (including narrowly moral notions such as wrongness, blameworthiness, moral responsibility, unfairness, etc.) by way of their appraisal component, without necessarily being *constituted* by moral judgments containing moral concepts or principles. If this is right, then tokens of genuinely moral anger need not be constituted by moral judgments. This means that, on the view of moral anger I endorse, not all tokens of the moral anger emotion-type are tokens of resentment or indignation, understood as cognitively sharpened forms of anger.

To lend support to the claim that appraisals of wrongness need not contain a constitutive normative judgment, David Shoemaker offers a compelling analogy between instances of blaming anger and aesthetic responses:

> Suppose you are moved to tears on first hearing a live symphony play Samuel Barber's *Adagio for Strings*. This response arises in you unexpectedly as you hear the mournful chord progressions and perceive the interplay of violins and viola. It's not as if your response is constituted by a judgement *that this is beautiful*: in fact, making the judgment *that this is beautiful* would take you *out* of that aesthetic experience. Sure, your emotional response may count as the source of a later independent judgment that the piece was beautiful, but it seems a mistake to say that the judgment of beauty is constitutive of that aesthetic response at that time (Shoemaker 2018: 71).

Shoemaker claims that, just as this aesthetic experience is not constituted by a judgment (of beauty), being in a state of blaming anger does not seem to require a judgment about wronging or responsibility. This insight explains the sudden, irruptive nature of the anger one might feel when being shoved in a crowd or realizing that a snide remark was intended as a personal insult (71).

To spell the point out further, Shoemaker gives the following "more dramatic" example:

> Suppose in seeing your daughter's black eye you immediately realize that she was beaten up by her boyfriend. Your (presumed) angry response will likely come upon you immediately, arising independently of any judgments about whether he's responsible or blameworthy for wronging her (Shoemaker 2018: 70).

I agree with Shoemaker that angry flashes like these count as genuine instances of blaming anger; indeed, Shoemaker holds that they are "just as much blaming responses as those angry responses [like resentment or indignation] that incorporate some judgment about having been wronged" (70). As such, we should – in line with my recommendations in the previous chapter – reject a cognitivist account of moral anger and instead opt for a motivational view that can explain the moral content associated with moral anger by way of its characteristic appraisal.

At it turns out, not everyone who discusses "moral anger" has a cognitivist theory of emotion in mind. More simply, some theorists seek to differentiate moral anger from generic anger according to the kinds of objects each type of anger typically takes. For example, Myisha Cherry claims that "moral anger arises out of a moral wrong whereas [non-moral] anger can arise without the presence of any "moral" injustice at all" (Cherry m.s.: 1). Similarly, C. Daniel Batson says "moral anger is anger provoked by the perception that a moral standard (principle, ideal) has been or will be violated" (Batson 2016: 156). For both Cherry and Batson, moral anger is elicited by

moral objects, i.e. cases of (perceived) injustice. Moral anger is thus distinguished from non-moral (generic) anger, which is elicited by non-moral objects.

As Antony Aumann & Zac Cogley observe (n.d.) it is less common for psychologists to use the term "moral anger," preferring to stick with just "anger."[46] Without the assumed distinction between moral anger and a broader, non-moral kind of anger, we find ourselves with an expansive emotion type that is elicited by all kinds of non-moral objects as well as moral objects such as perceived injustice. For example, Haidt (2003) claims that anger's typical elicitors are goal blockage and frustration, as well as slights or insults[47] that are taken to be unjustified, which includes perceived moral norm violations. In a similar vein, Tangney et al. write in their review article about moral emotions that "people may experience anger for a very broad range of situations – e.g., when insulted, frustrated, inconvenienced, or injured" (2007: 356). Tangney et al. (2007) take most appraisal theorists (e.g. Lazarus 1991; Roseman 1994; Smith & Ellsworth 1985) to hold that anger is elicited when people appraise an event (i.e. the object) as personally-relevant, goal-frustrating, and in some way caused, often intentionally so, by a responsible other. This approach suggests that anger and its characteristic appraisal has all the anger-elicitors, non-moral and moral alike, covered. So too does anger's characteristic action tendencies, which include aggressive, retaliatory behaviors as well as actions aimed at overcoming goal-frustration (Prinz & Nichols 2010). If this broad understanding of anger is right, there seems to be no need to posit the existence of a distinctively moral subtype of anger, as I plan to do.

I think, though, that we have good reason to resist this approach, and insist on drawing a distinction between generic anger and non-moral anger, which is a move also made by David Shoemaker (2018). To assume that there's just one, broad kind of anger that is elicited by so many different objects is to assume that these wildly varied anger tokens can be sufficiently unified by anger's characteristic appraisal, action tendencies, and goal(s). And I don't think that's plausible: an emotion that has such disparate eliciting conditions seems like it would be an odd sort of emotion. In what follows, I argue against the idea that there's only one, broad kind of anger and instead make the case that there are *two* kinds of anger: generic anger, and moral anger. It should

[46] Aumann & Cogley also note, to echo a point I made above, that "philosophers almost always call the relevant emotion 'resentment' or 'indignation'" (n.d.: 1, n5). The emotion they have in mind is the anger involved in blame that can be overcome by forgiveness.

[47] Haidt is by no means alone in taking anger to be elicited by slights, offenses, and insults. See e.g. Lazarus (1991), D'Arms & Jacobson (2014) and Shoemaker (2015).

be noted that throughout this chapter, I proceed on the assumption that the theory of emotions I developed in chapter 2 is broadly correct, and that we ought to type-identify emotions *primarily* according to their characteristic action tendencies and goals. In line with the weaker, syndrome view, we may also attend to the nature of emotions' elicitors, and the characteristic thoughts associated with emotions' appraisals.

As we've already seen, psychological research tells us that anger is elicited by a wide range of elicitors, a salient subset of which involve perceived injustice.[48] To look more closely at the range of objects that elicit anger, let's look to one of the first, and largest-scale, studies on anger. Hall (1898) collected detailed reports from over 2000 participants, including children and adults, about things that made them angry. There were numerous references to anger elicited by injustice, with participants naming elicitors such as "injustice," "injustice to others," being "accused of doing what I did not do," and "self-gratification at another's expense, cruelty, being deceived." There were also many references to anger elicitors that did not obviously have anything to do with perceived injustice, such as "girls talking out loud and distracting me in study hours," "an over tidy relative always slicking up my thing," being "given a seat in church behind a large pillar," "being kept waiting, being hurried, having my skirt trodden on, density in others," "if I am hurrying in the street and others saunter, so that I cannot get by," and "slovenly work, want of system, method and organization" (Hall 1898: 538-539, discussed in Prinz & Nichols 2010: 129).

Impressed by the rich variety of anger elicitors reported by the participants in Hall's study, and wanting to capture the fact that not all episodes of anger are elicited by perceived injustice, Prinz & Nichols (2010) suggest all of these anger elicitors, from being slowed down on the street by saunterers to being a victim of cruelty or deception, can be unified and understood as violations of the individual's autonomy, with instances of perceived injustice being the paradigm of such violations.[49] They write,

> Being annoying or disruptive, thwarting goals, violating personal possessions of space, being insulting or offensive – all these things have a negative impact on a victim, and thus fail to respect individual rights or autonomy (Prinz & Nichols 2010: 129-130).

[48] See Prinz & Nichols (2010: 128-130) for a useful discussion of the empirical strategy employed in many of these studies, and how it likely leads to an overrepresentation of injustice-induced anger episodes.
[49] Cf. Rozin et al. (1999), who also claim that anger is characteristically elicited by violations of autonomy norms, but do not seem to construe violations of autonomy as broadly as Prinz & Nichols.

Further, they claim,

> That anger is frequently triggered by [objects] other than perceived injustice should not be surprising if, as seems likely, the anger system is evolutionarily ancient… In older phyla, the homologues of anger may be more typically elicited by physical attacks… or the taking of resources (129).

Prinz & Nichols think that the anger system in humans should reflect its evolutionary history and make sense of Darwin's proposal that anger's adaptive function is to motivate retaliation (Darwin 1872, discussed in Prinz & Nichols 2010: 124). So, on this expansive view of anger, anger's characteristic appraisal is about violations of rights or autonomy, and its action-tendencies aim at the goal of retaliation in light of such violations. If we include the insights from appraisal theorists considered earlier, a move which is consistent with Prinz & Nichol's explicit attempt to give a unified account of anger, we might add that anger includes action tendencies that aim at overcoming goal-frustration in addition to its retaliatory action tendencies.

To say that these features are characteristic of *all* instances of anger, which is in essence what proponents of the expansive view of anger are committed to, strikes me as something of a stretch. To return to an example from Hall's study about typical anger elicitors, being kept waiting can certainly be angersome, but many aggravating situations that involve waiting, such as being in the doctor's office, are not plausibly construed as violations of autonomy. Most of us know that medical appointments often run behind schedule, and as frustrating as it is to have one's plans for the rest of the day messed up, we don't tend to appraise these annoying delays as interfering with our autonomy, nor do we harbor a genuine retaliatory desire to get back at the medical office's staff or take them to have intentionally kept us waiting. Consider, too, other common elicitors of anger, such as bad weather (hello, Michigan winters) that makes one's commute much more difficult, or a computer application that freezes at an especially inconvenient time. Though I might like to think I have a right to a fast, easy bike ride to campus, no matter the time of year, or to a glitch-free computer experience, it seems less plausible to think of poor weather or malfunctioning software as having violated my autonomy than to conceive of them just as instances of goal frustration. Admittedly, it is conceivable that I may be motivated to retaliate against my computer, insofar as I am appraising it as an agent that intentionally caused the software to crash, thereby screwing up my work day – it's not crazy to think that I might direct a few choice swear words at the laptop, or feel an urge to break it – but even if I grant this much in the computer case, I'm doubtful the same analysis can be applied to the bad weather case.

To lend further support to this last point, David Shoemaker (2018) rightly notes that while action tendencies that aim at overcoming goal frustration and action tendencies that aim at retaliation sometimes overlap, as suggested above with the computer case, there are pure cases in which only one kind of action tendency is operative. Shoemaker imagines a case where a rockslide blocks his way to work. The anger he feels when he sees the rockslide "will not result in any sort of motivation to retaliate against the rockslide!" Instead, he "will be motivated to look for a way around or through the rockslide, so that [he] can still get to work" (Shoemaker 2018: 73).

If these thoughts are right, then I think we should reject Prinz & Nichols' suggestion that there is one kind of anger, unified by an appraisal that one's rights or autonomy have been violated, that takes such a wide variety of objects in virtue of its disparate eliciting conditions. Instead, I think it's more straightforward, and empirically plausible, to think that cases of generic anger are typically elicited by *goal frustration*, whereas instances of (seemingly, still to be vindicated) moral anger are elicited by perceived injustice. Additionally, and relatedly, I suggest we should think of generic anger as having the sole aim of overcoming the goal-blocking obstacle, and consider action-tendencies that aim at retaliation distinctive of moral anger.

I am now in a position to properly spell out what I take generic anger to be. On my picture, generic anger is a syndrome of appraisal, feeling, and motivation that aims at overcoming goal frustration and obstruction. Generic anger appraises its objects as instances of goal-frustration, where the goal is personally relevant to the subject, and its action tendencies aim to fulfill its characteristic goal by removing or getting around the obstruction. This fits with Frijda's (1986; 2007) view of anger, according to which the action tendencies associated with generic anger aim to regain control or freedom of action, usually by removing whatever obstacle is restricting one's freedom or control. Someone in the grip of generic anger will experience feelings of heat (e.g. higher skin temperature and heart rate), and an urgent desire to be rid of the obstacle. Angry feelings, thoughts, and action tendencies are irruptive and will seek precedence over and control one's attention and behavior. The motivational theory predicts that in cases where the obstacle can't be removed – it's unfortunately quite difficult to get rid of a blizzard or a MacBook's pesky spinning rainbow wheel – the experience of anger will persist, because generic anger's goal has not been satisfied.

To see the motivational analysis of generic anger in action, imagine that I have recently planted some tomato plants in my backyard. I've become very invested in seeing them thrive, and

am very much looking forward to reaping the fruits (so to speak) of my gardening labor. Unfortunately, a squirrel lives nearby who has taken to eating all the tomatoes. One day, I glance out at my backyard and see the squirrel nestled among my tomato plants, chomping away. My attention is immediately diverted from what I was doing as it comes upon me that the squirrel is yet again thwarting my tomato goals. I run outside, shouting expletives at the squirrel, stomping loudly, and waving my fists. These tactics successfully scare the squirrel away and, reassured that my tomato plants are (temporarily) safe, my anger subsides, at least until the squirrel's next tomato mission.

On the motivational picture, a constitutive judgment about goal-frustration is not required for someone to experience generic anger. In the example given above, I never formed a judgment about the squirrel frustrating my goals – but I nevertheless appraised the situation in a way that captured these key elements. The motivational theory thus accommodates a datum cognitivist theories of emotion struggle with: that pre-verbal infants (and some animals) are just as capable of experiencing genuine bouts of generic anger (and other emotions) as adults, despite not yet possessing the relevant concepts or being able to form thoughts with reportable, relevant propositional content, as long as they are capable of making goal-frustration appraisals (see e.g. Deigh 1994 and Griffiths 1997 for a general formulation of this objection to cognitivism). It seems quite clear that babies experience real anger, and that their experiences of anger involve the perception that they have in some way been prevented from getting what they want, be it "food, drink, attention from mommy" (Shoemaker 2018: 72).

With this picture of generic anger in mind, I now want to briefly sketch my view of moral anger, which I will develop more fully in the next section. As suggested above, I take moral anger to be characterized by an appraisal of perceived injustice. Thus, moral anger's paradigm elicitors include moral violations perpetrated against oneself or others. These violations are, unsurprisingly, usually appraised as wrong, but may be more specifically appraised as wrong in virtue of being unfair (Mikula 1986; Scherer 1997), demeaning or offensive (Lazarus 1991), and/or blameworthy (Ortony & Turner, 1990). Since I take moral anger to belong to the same emotion family as generic anger – i.e. it is a genuine form of anger – we should expect moral anger to resemble generic anger when it comes to some of its characteristic features, such as its facial expression, which typically includes a furrowed brow, raised eyelids, and thin lips and/or clenched teeth (Ekman & Friesen 1971), a propensity to shout, and feelings of heat and aggression. Importantly, however, the action

tendencies and goals that define moral anger, on my view, go beyond mere obstacle-removal and are best thought of as distinctively moral in character. As I hope to establish in the next section, moral anger's action tendencies and goals are the most helpful element of the emotional syndrome that help set moral anger apart from generic anger and justify my positing its existence as a distinctive emotion type.

3. What are the distinctive aims of moral anger?

In this section, I develop my account of moral anger in more detail. I contend that moral anger is primarily defined by two distinctively moral aims: a sanctioning goal, and a communicative goal. First, though, I will first lay out some findings regarding moral anger's action tendencies. As several emotion theorists have pointed out (e.g. Frijda 1986; 2007; Deonna & Scherer 2010; Frijda & Scherer 2009), an effective strategy for identifying an emotion's distinctive aim(s) is to examine the various actions the emotion motivates, and consider what goal (goals) unifies (unify) the various, possibly disparate actions.

First, moral anger elicited by perceived moral violations in economic games, such as low ball offers in ultimatum games and free-riding in public goods games, motivates both victims and third-party bystanders to take action to punish the wrongdoer, despite incurring material costs to do so (see e.g. Fehr & Fischbacher 2004; Fehr & Gachter 2002; Fischbacher et al. 2001; Pillutla & Murnighan 1996; Tangney et al. 2007; Haidt 2003; Prinz & Nichols 2010). Participants who choose to punish perpetrators describe their punitive actions as expressions of anger (Pillutla & Murnighan 1996; Fehr & Gachter 2002). Participants also choose to punish even if they know they will have no further interactions with the wrongdoer, which, coupled with the fact that they must give up some of their own money in order to punish, suggests that the motive for punishing can't be explained by the expectation that punishing will lead to future economic benefit in the game (Fehr & Gachter 2002). Punishing behavior in economic games thus cannot be understood as motivated by self-interest.[50] Rather, is typically interpreted as a form of retaliation aimed at

[50] Robert Frank (1988) argues that anger, which he considers a moral emotion that brings with it a "sense of justice" (54), is best understood as an evolved commitment device. Commitment devices enable people to pursue a more rewarding long-term strategy of cooperation by acting in a costly (to one's self-interest) way in the short term. This explains why people choose to punish wrongdoing in economic games, despite incurring material costs in doing so.

remedying perceived injustice by getting even (Tangney et al. 2007; Haidt 2003; Prinz & Nichols 2010; D'Arms & Jacobson 2014; Gibbard 1990).

The idea that retaliation is a key action tendency of anger is famously found in Aristotle. In the *Rhetoric,* Aristotle writes,

> Anger may be defined as an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification towards what concerns oneself or towards what concerns one's friends. If this is a proper definition of anger, it must always be felt towards some particular individual (*Rhetoric* 1378a-b, 70-71).

Though Aristotle does not explicitly call the emotion he has in mind "moral," insofar as he is concerned with explaining what role the emotions play for a properly virtuous person, I think it's reasonable to interpret him as having a conception of anger in mind that is narrower than a broad view of anger that's also intended to capture goal-blockage.

There are, however, competing interpretations of the data regarding the punishing behavior seen in economic games that go beyond an aim of retribution. For example, Brendan Dill & Stephen Darwall (2014) argue that the moral emotions of guilt and blame (what I'm calling moral anger) serve the central moral motives of condemnation and conscience. The core motivation of moral anger, which is elicited by an appraisal of wrongdoing and directed at the wrongdoer, is condemnation, which aims to get wrongdoers to hold themselves to account for their wrongdoing. Guilt's core moral motive is conscience, which is elicited by one's own wrongdoing, which functions to hold oneself to account. Dill & Darwall argue that their account is better supported by the empirical data than a rival, purely retributive hypothesis, which holds that the moral emotions, particularly moral anger functions only to make wrongdoers suffer.

I agree with Dill & Darwall that moral anger doesn't aim *solely* at retribution. In addition to moral anger's retributive aim, I hold that moral anger is *also* characterized by a communicative aim, and that this element of my view can be made consistent with Dill & Darwall's accountability aim(s). I suggest that some (but, as I will try to make clear below, not all) of moral anger's characteristic action tendencies, which may include obvious communicative and expressive acts like yelling or physically lashing out, should be understood as communicative attempts to hold the wrongdoer to account. If communication is successful and there is uptake of the condemnatory message, guilt will likely be elicited in the wrongdoer (Aumann & Cogley n.d.; cf. Macnamara 2015; Gibbard 1990). A wrongdoer in the grip of guilt will, as I argued in the preceding chapter,

recognize that they are responsible for having wronged the victim and engage in reparative actions, aimed at repairing the relationship that's been damaged by the wrong. In so doing, they are holding themselves to account. Once the victim sees this, the communicative aim of moral anger, and moral anger itself, will presumably be satisfied, which can in turn allow the wrongdoer's guilt to subside.

3.1 The communicative goal of moral anger

David Shoemaker (2015; 2018) has recently advanced a compelling account of moral anger (what he calls "blaming anger," or, equivalently, "angry blame"). Shoemaker, who, like me, makes use of a motivational theory of emotion (following Frijda 1986; 2007 and Scarantino 2014) argues that the fundamental, encompassing aim of angry blame is communicative (Shoemaker 2018). Unlike me, Shoemaker argues that any sanctioning or retaliatory action tendencies that are part of angry blame, such as the desire to get back at, lash out at, or punish a wrongdoer, are merely instrumental to angry blame's fundamental communicative aim. In what follows, I try to show that Shoemaker's account of moral anger is incomplete because the retributive goal of moral anger can come apart from its communicative aim. I argue that it's possible for moral anger to be partially satisfied even when its communicative aim is not fulfilled, something that my account, but not Shoemaker's, predicts and explains.

First, however, I will examine Shoemaker's arguments in favor of the communicative aim, which are illuminating and have convinced me that moral anger is partially defined by a goal that aims at communication. Shoemaker's arguments also help make clear why another rival view of moral anger – the protest account of blame – is deficient. Shoemaker notes that satisfaction of moral anger's communicative aim is partially achieved just by "getting something off your chest" (2015: 104). Being able to give voice to one's anger as opposed to keeping it bottled up can yield great, if not total, relief. Shoemaker takes this as a reason for why we should prefer an account of moral anger that includes the communicative aim over the protest account of blame, a rival account of angry blame that does not include the communicative function. On the protest account, the central purpose of expressing one's moral anger is to stand up for oneself (see e.g. Smith 2013, discussed in Shoemaker 2015), perhaps to restore one's self-respect or one's standing relative to others after suffering a wrong or a slight. Though the expression of anger may well serve to reestablish one's self-respect, especially in the eyes of other third-party bystanders, it seems as if

something important is missing when the anger is not successfully communicated to the perpetrator of the slight. Shoemaker brings this point out nicely by asking us to consider two pairwise cases. In the first case, the communication of one's anger to the wrongdoer is achieved, whereas in the second, it is not. Shoemaker argues that if the wrongdoer, who is the primary target of the expressed moral anger, fails to hear and understand this expression as a protest of the wrong he committed, moral anger's aim is frustrated (2015: 105). Anger which is expressed but not taken up by the wrongdoer thus falls short, even when it's heard by bystanders. As Shoemaker writes, moral anger "aspires to be more than just public noise" (105). It essentially involves a communicative function which is only fully satisfied when the expression of anger is successfully communicated to, and heard by, the wrongdoer.

Here, it's worth noting that Shoemaker (2015; 2018) no longer believes (cf. his earlier work) that anger's communicative, expressive tendencies aim to sanction the wrongdoer. He denies that retaliation is built into the moral anger felt by victims of wrongdoing. Such victims "may, without frustration, merely be motivated to shun, or shoot a dirty look at, the slighter" (Shoemaker 2015: 107). To Shoemaker, the fundamental point of these and other actions motivated by moral anger is to make the slighter fully aware of what he has done. Moral anger is thus "a demand to get him to appreciate, to acknowledge, the emotional havoc (and worse) that he has wreaked" (107). To Shoemaker, although an angry demand leveled at a wrongdoer to acknowledge what he has done might bring pain to the offender, this pain is just a contingent side-effect of the communicative demand. The sanctions brought about by angry actions are not, then, part of moral anger's characteristic aim. Shoemaker says that though it's true that the retributive motivation of moral anger is often associated with its communicative aim, "that's essentially because retribution is perhaps the most effective and dramatic *form* the communication of anger can take" (2018: 75).

To support the claim that the fundamental aim of moral anger is communicative, Shoemaker cites some empirical findings I myself drew on in Chapter 1.[51] To recap: Gollwitzer and colleagues (2009; 2011) conducted two studies which sought to clarify what goal is motivating 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

[51] Dill & Darwall (2014) also take these findings to support their claims about the core moral motive of condemnation aiming at accountability, and the conclusion I drew from the results in Chapter 1 are entirely consistent with Dill & Darwall's interpretation.

the opportunity to punish their exploitative partner by signing them up for an unpleasant follow-up task. One 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 if, and only if, it was followed by a message from the perpetrator that showed that he was holding himself accountable for his earlier wrongdoing. Notably, participants who got no such message from the wrongdoer were dissatisfied; interestingly, they were just as dissatisfied as the participants who elected not to punish the wrongdoer at all.

In Chapter 1, I concluded from these results that what people are seeking when they punish people who wrong them is not mere retribution – i.e. people are not content just to have an outlet for their anger. I suggested that punishing behavior is best explained as being driven by a desire to hold the wrongdoer to account. Importantly, I said that "the functional goal of [moral] anger, which is achieved by its distinctive action tendencies, is to redress the perceived injustice that has taken place" (see page 21). According to my current, more detailed account of moral anger, moral anger will be properly fulfilled when both its communicative and retributive aims are satisfied, and for the participants in the study who both meted out punishment *and* received a satisfying message from the wrongdoer, this indeed seems to be the case. The result that nicely confirms Shoemaker's claim and thus spells trouble for my account is the finding that participants who punished yet got no message from the wrongdoer were just as dissatisfied as participants who neither punished nor (successfully) communicated with the wrongdoer. If I am right in claiming that moral anger has two distinct aims, we should expect those who punished to have at least partially satisfied moral anger's retributive aim, and be slightly more satisfied than those who neither punished nor had any communication with the offender.

At this point, I could try to tell a story about how the boring follow up task that was meant to serve as the punishment in the study wasn't sufficiently unpleasant to count as punishment, or explain the results away in some other way. I don't think that would be very fruitful, though, so instead I want to raise the possibility that in general, the infliction of suffering (via punishment, etc.) aims in part at the goal of retaliation. As Aumann & Cogley (n.d.; see also Cogley 2013) observe, the elicitation of guilt *is itself a sanction*, in virtue of its being unpleasant to feel. When shooting a dirty look at or pointedly ignoring your slighter, the point is not just to make clear to

them that you're angry. Such actions often seem designed to hurt in order to get back at the wrongdoer for having hurt you (or a third party) in the first place. Punishing behavior certainly helps fulfill moral anger's communicative aim in many cases, but I want to suggest that it *also* aims at a distinctive, retributive goal to get back at the wrongdoer. In the next section, I discuss some other results from psychological studies to see if my suggestion has at least some empirical support.

3.2 The retributive goal of moral anger

Haidt and colleagues (2010) ran a study in which they showed clips from Hollywood films that portrayed injustice, and then asked participants to rate a variety of alternate endings. (All participants reported being very angry in response to the injustices they saw.) I'll discuss three of the ending options that are of particular interest for present purposes. The first kind of ending is one in which the victim made the perpetrator suffer in a way that parallels the original wrongdoing; for example, one of the movies involved a slave owner who had cut the limb off a slave. By way of punishment, the slave owner had one of *his* limbs severed by the slave. It's important to note that in these endings, the wrongdoer understood that he was being punished by the victim for the wrong he committed. The second type of ending also involved the perpetrator suffering in a way that paralleled the original wrong, but without the perpetrator knowing that it was the victim who inflicted the suffering. The final type of ending did not involve any suffering at all for the perpetrator; rather, the victim was depicted as coming to terms with the injustice (e.g. by becoming more active in their church community or joining a support group) and learning to forgive the perpetrator for the injustice.

Haidt et al. found that participants were most satisfied by endings that depicted the perpetrator being punished by the victim *and* knowing that their suffering was being inflicted as payback for the original injustice. Participants were moderately satisfied by endings in which the perpetrator suffered, yet didn't know they were being punished by the victim. Interestingly, participants were not at all satisfied by the endings in which the victim dealt with the injustice on their own and learned to forgive the perpetrator. Taken together, these findings suggest that moral anger elicited by perceived injustice is fully satisfied when it achieves its retributive aim – the infliction of suffering or punishment – *and* its communicative aim – the perpetrator gets the

message that they are being punished for the wrong they committed. Infliction of punishment without successful communication was less satisfying, which is consistent with what my account of moral anger predicts: even though the communicative aim was not reached, the retributive goal was satisfied, so we should expect only partial satisfaction of the moral anger episode.[52]

In addition to the movie ending experiment, Haidt et al. (2010) also asked participants to describe satisfying revenge scenarios from their own lives. They write,

> Most of the satisfying revenge scenarios involved some form of immediate retaliation, generally verbal. But a few described efforts to retaliate behind the scenes. A male who felt insulted by a professor wrote a scathing letter about him, which he thinks was instrumental in preventing the professor from obtaining tenure (Haidt et al. 2010: 10).

I take this example to demonstrate that, in some cases, moral anger can be at least partially satisfied by knowing that the wrongdoer has suffered, even if the wrongdoer doesn't know that their suffering has been inflicted on them by the victim of the original injustice. The satisfaction of this man's moral anger can't just be explained in terms of the release that comes from expressing his anger and getting it out onto the page. Rather, it makes more sense to interpret the act of writing the letter as a means to the retributive goal of making the professor suffer: here, by derailing his tenure case. In this case, it seems likely that the man's moral anger was satisfied only when he heard that the professor was, in fact, denied tenure, thereby receiving confirmation that his retaliatory efforts had succeeded.

Finally, let me briefly consider some results from a study conducted by Doob & Wood (1976, discussed in Haidt et al. 2010). The study found that victims of injustice who observed their "tormentor" being punished were less likely to administer additional pain to the wrongdoer, even though they themselves were not the ones actively making the wrongdoer suffer and they had no way of communicating with the wrongdoer. I think that we can interpret the victims' decreasing urge to continue to inflict pain on the wrongdoer as evidence that their moral anger has been partially satisfied by seeing the wrongdoer punished. There is no communication in this scenario, but there is retribution. The partial satisfaction of moral anger in the absence of communication seen here supports my claim that moral anger has a distinctive retributive goal that can come apart from its communicative aim.

[52] Unfortunately, the study did not include a class of endings that depicted communication without retaliation.

With these considerations in mind, I think that we should accept that my account of moral anger builds on Shoemaker's view in an important way. I have tried to establish that Shoemaker's account of moral anger, which considers moral anger's retaliatory action tendencies as merely instrumental to its fundamental communicative aim, cannot adequately explain cases in which moral anger is partially satisfied despite its communicative aim not being met, and is thus incomplete. By contrast, my account of moral anger, which includes a retributive aim that can come apart from the communicative aim, is better able to predict and explain these cases.

4. Conclusion

In this chapter, I defended a novel, motivational theory of genuinely moral anger. Drawing from the motivational theory of moral emotions I developed in Chapter 2 and building on David Shoemaker's insightful account of angry blame, I argued that there is a distinctively moral kind of anger that is differentiable from generic anger. I claimed that moral anger counts as distinctively moral in virtue of its action tendencies, which are typically triggered by perceived injustice against oneself or others, and aim to satisfy *two* moral aims: a communicative goal, and a retributive goal.

I argued that my motivational theory is preferable to a cognitivist account because it can make sense of moral anger's moral content without committing to the claim that instances of moral anger are necessarily (partially) constituted by moral judgments. If my account of moral anger is right, then not all tokens of distinctively moral anger are tokens of resentment and indignation, where these emotions are understood as partially constituted by a judgment about wrongness, blame, or responsibility. This insight has important implications for ongoing debates in the moral responsibility literature, in which a cognitivist analysis of resentment and indignation is mostly assumed.

A potential objection that must be acknowledged is whether I have done enough to establish the conclusion that there are two distinct kinds of anger. What I have tried to do in this chapter is show that the empirical evidence from the psychological literature and Shoemaker's (2018) proposed bifurcated view of anger dovetail to support my claim that there is a distinction in psychological kinds of anger worth making. If the distinction I have attempted to draw between moral anger and non-moral anger does not hold, then I would need to find another way to support

my claim that there is a distinctively moral kind of anger without reverting to a cognitivist approach and claiming that moral anger is partially constituted by a narrowly moral judgment of some kind.

One reason in favor of my claim that there are two kinds of anger, moral anger and generic anger, is that it presents a promising solution to a widespread problem in both the psychological and philosophical literatures. The problem is that the accounts of anger offered by moral psychologists about how anger appraises its object and what the goal of anger is do not fit together neatly enough to solve the following puzzle: how to make sense of the motivation to retaliate or punish that seems central to some instances of anger, while *also* explaining why there seem to be paradigm cases of anger that are not about perceived injustices or slights – the sort of things to elicit retaliatory or punishing action tendencies – so much as they are about goal frustration, which do not seem to involve the same kinds of action tendencies. It is difficult to see what goal frustration and perceived injustice have in common, other than the fact that we seem to get angry both when we take ourselves to have been wronged and when our goals have been frustrated. My proposal – i.e. that we should posit the existence of a distinctively moral kind of anger that captures the instances of anger elicited by perceived slights and offenses that seems central to our moral psychology and practices that is, importantly, distinct from generic anger, which explains the instances of anger elicited by goal frustration – is an empirically-informed attempt to address this problem without adopting a broad account of anger that must unite such a disparate set of eliciting conditions, appraising thoughts, and action tendencies.

The overarching goal of this chapter, which brings together key aims of the dissertation project as a whole, was to offer a novel, empirically-supported account of moral anger that constitutes a positive answer to the ontological question about moral anger. It is my hope that I have done enough to show that it is possible to vindicate the existence of a genuinely moral emotion – specifically, moral anger – while staying true to the important idea that the moral emotions should be understood as a recognizable subset within the general class of the emotions.

References

Aristotle. (1954). *The Rhetoric and the Poetics of Aristotle,* trans. W. Rhys Roberts. New York: The Modern Library.

Aumann, A. & Cogley, Z. (n.d.). Forgiveness and the multiple functions of anger. Unpublished manuscript.

Cherry, M. (n.d.). Moral anger, motivation, and productivity. Unpublished manuscript.

Cogley, Z. (2013). The Three-Fold Significance of the Blaming Emotions. In D. Shoemaker (Eds.), *Oxford Studies in Agency and Responsibility.* Oxford: Oxford University Press, 205–224.

D'Arms, J. & Jacobson, D. (2003). "The significance of recalcitrant emotions (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement,* 52, 127-45.

Dill. B. & Darwall, S. (2014). Moral psychology as accountability. In J. D'Arms & D. Jacobson (Eds.), *Moral Psychology and Human Agency.* Oxford: Oxford University Press, 40-83.

Doob, A. N., & Wood, L. E. (1972). Catharsis and aggression: Effects of annoyance and retaliation of aggressive behavior. *Journal of Personality and Social Psychology, 22*: 156-162.

Ekman, P., & V. Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–9.

Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution & Human Behavior, 25,* 63-87.

Fehr, E. & Gachter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137-40.

Fischbacher, U., Gachter, S. & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397-404.

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions.* New York: Norton.

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology, 45*(4), 840-44.

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology, 41*(3), 364-374.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*. New York: Oxford University Press, 852-870.

Haidt, J. Sabini, J., Gromet, D., & Darley, J. (2010). What exactly makes revenge sweet? Unpublished manuscript.

Hall, G. S. (1898). A study of anger. *American Journal of Psychology, 10,* 516-591.

Kumar, V. (2016). The Empirical Identity of Moral Judgment. *The Philosophical Quarterly, 66*(265), 783–804.

Lazarus, R. (1991). *Emotion and adaptation.* New York: Oxford University Press.

Macnamara, C. (2015). Blame, Communication, and Morally Responsible Agency. In R. Clarke, M. McKenna, & A. Smith (Eds.), *The Nature of Moral Responsibility: New Essays*. New York: Oxford University Press, 211-236.

McKenna, M. (2013). Directed blame and conversation. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms.* New York: Oxford University Press, 119-140.

Mikula, G. (1986). The experience of injustice: toward a better understanding of its phenomenology. In H.W. Bierhoff, R. L. Cohen & J. Greenberg (Eds.), *Justice in Social Relations.* New York: Plenum Press, 103-124.

Ortony, A. & Turner, J. (1990). What's basis about basic emotions? *Psychological Review, 97:* 315-331.

Pereboom, D. (2013). Free will skepticism, blame, and obligation. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms.* New York: Oxford University Press, 189-206.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*(3), 208-224.

Prinz, J. (2009). The Moral Emotions. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion*. New York: Oxford University Press, 519-538.

Prinz, J. & Nichols, S. (2010). Moral emotions. In J. Doris and the Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook.* Oxford: Oxford University Press, 111-46.

Roseman, I. J., Wiest, C., and Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67*: 206-221.

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*(4): 574-586.

Scanlon, T. M. (2013). Interpreting blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms.* New York: Oxford University Press, 84-99.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology, 73:* 902-922.

Shoemaker, D. (2017). You oughta know: defending angry blame. In M. Cherry & O. Flanagan (Eds.), *Moral Psychology of Anger.* London: Rowman & Littlefield, 67-88.

Smith, A. (2013). Moral Blame and Moral Protest. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms*. New York: Oxford University Press, 27-48.

Smith, C. A. & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*: 813-838.

Strawson, P.F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1-25.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology*, *58*(1), 345–372.

Wallace, R. J. (1994). *Responsibility and the moral sentiments.* Cambridge: Harvard University Press.

Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, *24*: 227-248.