

The Good and the Gross: Essays in Metaethics and Moral Psychology

by

Alexandra A. Plakias

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
In The University of Michigan
2011

Doctoral Committee:

Professor Elizabeth Anderson, Chair
Professor Allan Gibbard
Professor Peter Railton
Assistant Professor Chandra Sripada

© Alexandra A. Plakias 2011

*To my parents and my sister, for everything.
And to Corey Brettschneider, because I promised.*

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee— Liz Anderson, Allan Gibbard, Peter Railton, and Chandra Sripada— for their contributions to this project. Liz in particular has provided incisive comments and very helpful suggestions throughout the writing process, and has been an immense help in thinking about how to organize and present the ideas contained in this dissertation.

I first began thinking about moral disagreement and the ideas discussed in Chapter Two while working towards my M.A. at the University of California, Santa Cruz. There I was fortunate enough to write my M.A. thesis under the supervision of John Doris, without whom this dissertation would not have been possible. I will forever be grateful to John for showing me how empirical psychology can be brought to bear on philosophical debates, and for being a source of encouragement, support, and advice throughout this process; and for being a true mentor to me. Though they left after my second year in the program, Andy Egan and Peter Ludlow both helped shape how I think about relativism and disagreement; the ideas presented in Chapter Three grew out of an independent study I did with Andy Egan in the fall of 2007. I am also grateful to Valerie Tiberius for her wisdom and her support, and to Tim Sundell, whose work on how (not) to define disagreement has helped shape my own thinking on the matter. Thanks also to Ernie Lepore, Peter Ludlow, and Jason Stanley for providing me with the opportunity to examine these issues through the lens of philosophy of language and semantics.

My interest in disgust and its role in moral judgment developed during the course of a reading group I ran with psychologist Nina Strohminger from 2008 to 2010. What began as a weekly meeting to discuss recent publications in moral psychology soon grew into a friendship that has been a constant source of inspiration, comfort, and hilarity throughout the writing of this dissertation. I'm beholden to Nina for teaching me about disgust, humor, and how to deploy fart spray in a laboratory setting. Shen-yi Liao provided enormously helpful comments on an earlier draft of this chapter, as did Daniel Kelly, and I thank them both.

I have also been fortunate enough to have the chance to present earlier versions of these papers to audiences at ANU, Northwestern University, and Western Michigan University, and I'm grateful to those audiences for their feedback and for the opportunity to work through these arguments and ideas. Thanks are also due to the undergraduate professors who inspired me to pursue a life in philosophy: Katheryn Doran, Marianne Janack, and Rick Werner. And to my fellow graduate students both at Michigan and at other departments, especially Nate Charlow, Kevin Coffey, Eduardo Garcia-Ramirez, Megs Gendreau, Liz Goodnick, Alex Hughes, Lina Jansson, Eleni Manis, Alison Niedbalski, Bryan Parkhurst, David Plunkett, and Amanda Roth. Finally, I am grateful beyond words to my parents and my sister for their patience and their love, and for teaching me the meaning of unconditional support. Without them, none of this would have been possible.

Table of Contents

Dedication	ii
Acknowledgements	iii
Introduction.....	1
Chapter One: The Good and the Gross	10
Introduction	10
1. Disgust: a Description.....	12
2. Disgust as Distortion?.....	17
2.1 Moral Disgust as Auxiliary Disgust	18
2.2 The Plasticity of Disgust.....	20
2.3 Disgust as Distortion	22
2.4 Disgust as Magical Thinking.....	26
3. Doubting Disgust.....	31
3.1 Disgust as Constitutive of Moral Judgment	31
3.2 Disgust as Evidence of Moral Wrongness	32
3.3 Poisons, Parasites and... Persons?.....	33
Conclusion: Disgust, Affect, and Evaluative Skepticism	40
Chapter Two: How Moral Disagreement is a Problem for Realism	42
Introduction	42
1. A <i>Priori</i> Arguments from Disagreement.....	44
2. Explaining versus Explaining Away	48
2.1 Non-moral Facts.....	49
2.2 Irrationality or Insensitivity to Evidence	49
2.3 Partiality	50
2.4 Background Theory	51
3. Disagreements Defused?	53
3.1 Honor	53
3.2 Sex and Sexuality	55
3.3 Fairness.....	58
4. The <i>A Posteriori</i> Arguments.....	59
5. Objections and Replies.....	65
5.1 Overgeneralization/Companions in Guilt.....	65
5.2 The Argument is Self-Defeating.....	68
Conclusion.....	70
Chapter Three: Moral Relativism and Monolithic Metaethics	72

Introduction	72
1. The Objectivity Assumption	73
A Puzzle About Intuitions	76
1.2 Culture, Content or Both?.....	79
2. If the Objectivity Assumption is False, is any Metaethics Permitted?	80
2.1 How to be a Moral Relativist.....	82
2.2 How to be a Moral Relativist and a Moral Objectivist.....	88
3. Objections and Replies.....	91
3.1 Variantism, Redux	91
3.2 It's Normative, but What Makes it Moral?	93
Conclusion.....	96
Works Cited.....	97

Introduction

Our moral evaluations of an act are influenced by factors ranging from whether we've recently washed our hands to whether we've just heard a funny joke. A flurry of recent work in empirical moral psychology has focused on cataloguing these influences, but identifying them and understanding the circumstances under which they operate gives us only half the story. We intuitively recognize some causes of judgments or choices as arbitrary and irrelevant; if I decide to donate to a specific charity by flipping a coin, or by picking its name out of a hat, then arguably I've failed to make a judgment at all. Instead, I've let an external factor— one that's completely unrelated to the relevant features of the case— determine my action for me. On the other hand, if I choose to give to a charity that buys antiretroviral drugs to treat HIV and sends them to Africa rather than a charity that buys suits for laid-off stock brokers, I've made a decision based on my assessments of need, suffering, and perhaps desert. The territory between these two extremes is what this dissertation sets out to explore: how do we assess whether a given factor is distorting, rather than informing, a normative judgment? How much divergence— in both the content of judgments and the causes of those judgments— can a single metaethical theory accommodate?

The three papers in this volume stem can be read as an attempt to illustrate how empirical inquiry into the nature of moral judgment informs and constrains metaethical theorizing. In Chapter One, 'The Good and the Gross,' I argue that the ubiquity of affective and metacognitive influences on judgment means that a straightforward

skeptical argument from the influence of disgust on moral judgment is untenable, as it threatens to result in global evaluative skepticism. In Chapter Two, ‘How Moral Disagreement is a Problem for Realism,’ I show why the argument(s) from moral disagreement depend on facts about the actual extent and nature of moral disagreement, and argue that these facts create problems for moral realism. In Chapter Three I show that there is no straightforward answer to the question, ‘does moral discourse presuppose objectivity?’ Rather, the answer is, ‘it depends.’ This is because empirical research has revealed that judgments of moral objectivity vary depending on the content of the claim under consideration. But this does not mean that moral discourse is metaethically variable, inconsistent, or incoherent: a relativist analysis of moral claims can explain and accommodate this variation in objectivity.

The implications of the arguments presented in Chapter One extend far beyond disgust. I have chosen to focus on disgust here both because its involvement in moral judgment has recently become the focus of a flurry of empirical research, and because it presents a particularly challenging case. Disgust’s track record as an influence on policies and behaviors is mixed at best: it is often recruited to motivate exclusion and ostracism, and to paint certain ethnic or religious groups as unclean or polluting. This checkered past, taken in conjunction with disgust’s origins as a guard against contamination by parasites and poisons, makes disgust a particularly tempting target for a skeptical argument. In fact, however, this line of thought conflates first-order normative considerations with second-order questions about the appropriateness of certain emotions or affective states as inputs to (or constituents of) evaluative judgments. The fact that disgust can be used to motivate morally reprehensible behaviors or attitudes doesn’t show

that it is itself a mistake; in fact, I argue, disgust may function, in the socio-moral domain, to alert us to the threat of social contamination. But this means that determining when disgust is an appropriate response to some act or attitude is inextricably bound up with the normative status of the act or attitude under consideration, just as determining when some substance is potentially deleterious to our physical health requires an antecedent conception of physical health. The reason this is particularly vexing in the moral case is that it is precisely the intuitions whose legitimacy is under consideration which inform our judgments about the normative status of those same actions and attitudes whose status we need to determine before we can decide whether the affective response that informs our intuition is appropriate or whether it's mistaken or out of place. In the case of normative judgments, we are reliant on intuition to produce the very yardstick against which we assess the accuracy of intuition.

I've attempted to circumvent this problem by focusing on the characteristic ways that disgust presents its objects, and this approach suggests that disgust is not uniformly out of place in moral and other evaluative judgments. But adopting this approach means that talk of whether affect or emotion 'distorts' or 'biases' evaluative judgment is a gross oversimplification. Instead, we'll have to adopt a case-by-case approach to the question of how and why various emotions or types of affect do or do not have a legitimate role to play in our evaluative appraisals.

This complication is not, I argue, specific to affect or emotion. Much like 'affect' and 'emotion,' the terms 'framing effect,' 'heuristic,' and 'bias' pick out a wide variety of phenomena. The fact that these latter terms are often used pejoratively reflects confusion between the first-order judgments that result from their application and the

phenomena themselves. And if affective influences are ubiquitous, framing effects are positively inescapable. For example, the order in which a cafeteria presents foods has a significant impact on which foods people choose. But there is no frame-independent presentation here against which we can measure various alternatives: the foods must be presented in some manner or another. Framing effects occur when a choice is affected by inconsequential features of the formulation or presentation of the options. But which features are consequential is itself a normative question, and as such, may not be subject to investigation from a frame-neutral standpoint. What we can do is reach a decision about what outcomes are desirable and choose frames and heuristics that promote that outcome. But we must be aware that the intuitions we rely on to reach that decision will themselves be influenced by affect, heuristics, and framing effects.

This observation has implications that extend well beyond theories of moral judgment. Advocates of paternalism might draw on arguments like the one above to justify policies that ‘nudge’ (to borrow a term from the legal scholar Cass Sunstein) people towards one decision rather than another: if the food has to be presented in some order or another, the thought goes, and if that order is going to influence people’s choices, then this presents us with an opportunity to help people make *better* choices. Of course, cake is not killing, and which foods are healthiest is a far less controversial matter than which lives are most valuable, even among the supposed experts.

A similar problem arises in the discussion of defusing explanations in chapter two. Here, the relevant question is how to counterfactually characterize the parties involved in a moral disagreement so as to evaluate which factors are or aren’t causally implicated in the dispute. This is not a problem involving framing effects, but it is a

problem that stems from finding a neutral standpoint from which the two parties can conduct their debate. I'll argue in chapter two that moral disagreements are often caused by disagreements over theories of fairness, or personhood; these theories might be thought of as frames of a sort. While not inconsequential, they exert an influence over our evaluations and preferences, and there is often no way of formulating or assessing options or preferences independently of them. And the fact that people disagree over which features are consequential or properly granted consideration (is it okay to accord oneself or one's family a greater share of resources? What if they're life-saving resources? How serious a wrong is an insult to one's reputation?) complicates the matter of settling moral disagreements: not only are we unable to escape the influence of framing on our moral judgments, but we can't agree on the normative status or acceptability of the frames themselves. To a utilitarian like Peter Singer, giving money to a homeless person rather than a child three thousand miles away is no less arbitrary than flipping a coin; on the other hand, perhaps one has a greater duty to foster well-being and help members of one's own community.¹

These considerations might seem unduly skeptical or pessimistic: after all, we can surely all agree that it is better to donate money to buy lifesaving medicine than to buy a Brooks Brothers suit for a wealthy (and healthy) young man. I grant the significance of this point, and in chapter three I take up the question of whether a metaethical theory can countenance the existence of both objective and relative moral claims.

¹ Singer would likely respond by arguing that utilitarianism is the only real frame-independent moral theory, since it relies on reason rather than affect. I return to this point in chapter one, but notice two things: the assessment and measurement of utilities is itself highly susceptible to framing; and arguments for the claim that reason yields better results than emotion depend on intuitions that in turn are susceptible to framing effects.

Thus far I've been speaking rather broadly in terms of moral judgments and of normative and evaluative judgments. But the latter category includes judgments about the tastiness of food, the attractiveness of landscapes, and the funniness of jokes. The papers presented here raise, but do not settle, the question of what makes a moral judgment *moral*. That is, what is special or unique about moral judgment; what sets it apart from other species of evaluative judgments? I don't spend much time defending a specific answer to this question in what follows, but the arguments presented here do have implications for how we go about answering this question. If we don't want to abandon talk of moral truth or falsity, then we should be wary of defining morality in terms of objectivity and convergence. Therefore, the arguments from disagreement in Chapter Two rule out an account on which moral judgment is distinguished by its commitment to the idea that it has universal scope or authority. In Chapter Three I offer an analysis of moral judgment, but the form of this analysis does not necessarily distinguish moral judgment from other evaluative judgments (such as aesthetic judgment). I concede that this may strike some people as a weakness, but in fact I think it is an advantage. This is in part because I think that there is fundamental disagreement over, and variation in, the proper objects of moral judgment. So-called 'purity' norms—those governing food, sex, and hygiene—are one example of a domain whose status as moral is often contested, and the possibility of such debate is worth preserving. A second consideration stems from the psychological underpinnings of moral judgment. The idea of a 'moral grammar' has become increasingly popular in recent years, and has been elaborated and defended by John Mikhail and Gilbert Harman, among others. But this strikes me as implausible, simply because the idea of an innate cognitive architecture

dedicated specifically to the acquisition and deployment of moral norms would be extremely inefficient. Humans— and perhaps other animals, such as chimpanzees— must recognize and conform to a variety of norms if they are to succeed as social creatures, and we've evolved the capacity necessary to do so. But if the capacity to learn and to conform to norms is evolutionarily antecedent to the capacity for moral judgment, then it's mysterious why an entirely separate cognitive architecture or module would be constructed to subserve moral judgment. Perhaps, then, it's more accurate to say that we have a normative grammar.

I consider and reject other attempts to identify a signature pattern or psychological kind characteristic of moral judgment in chapter three, so I won't belabor the point here. But an alternative to the approaches described above is to think of the domain of morality as distinguished by its concerns. That is, what makes a judgment specifically *moral* (rather than some other kind of normative judgment) is the subject matter it is concerned with. If we adopt this approach, the next question is: what is the distinctively moral subject matter; or, what is the characteristic that makes a norm moral? At the end of chapter three, I suggest that this question presupposes— falsely— that our concept of moral can be analyzed in terms of necessary and sufficient conditions. It may be more fruitful to think of morality as a prototype or exemplar concept, in which some violations, acts, or decisions are paradigmatic instances of morality, and others are more peripheral. Some might be the subject of deep and irresolvable disagreement; disgusting norm violations involving incest, or cannibalism, for example.

By now it should be clear that the papers I present here raise a whole new suite of questions. And as we make progress in understanding moral disagreement, so will our

understanding of the significance of disagreements involving disgust-based judgments of moral wrongness; likewise, as we improve our understanding of how disgust influences moral judgment, we will be better able to understand its normative significance, which will in turn enable a more informed assessment of disagreements in which one or both parties' judgment is framed in part by disgust. As I argue in chapter one, the significance of various affective and cognitive frames (and here I am including emotion in the former category) will need to be assessed individually, so the papers that follow may be read as the first in a series of such examinations.

This is a particularly exciting time to be engaged in moral psychology; empirical research into disgust, moral judgment, and judgments about moral objectivity is progressing rapidly. This is both good news and bad news for philosophers such as myself who incorporate these findings into their arguments: it requires that we maintain a certain modesty, since empirical work in these areas is really in its infancy (see note 1 in chapter one below). At the same time it means that we are afforded the opportunity to do more than speculate *a priori* about the causes of moral disagreements, or about the nature of disgust's influence on moral judgment, or whether people think of morality as objective or not— we have the opportunity to engage with empirical research and improve our understanding of the phenomena, and we gain access to a plethora of new evidence. It's this opportunity I intend to capitalize upon in the papers that follow, and I hope that in doing so I can simultaneously provide a useful illustration of how discoveries in empirical psychology can be assimilated into philosophical theory.

CHAPTER ONE:
THE GOOD AND THE GROSS

Introduction

In *The Anatomy of Disgust*, William Miller writes that “Disgust has elicited little attention in any of the disciplines that claim an interest in the emotions: psychology, philosophy, anthropology” (1997: 5)². This is no longer true. Philosophers, in particular, have recently taken an interest in disgust, and with good reason: disgust is strongly implicated in moral and other evaluative judgments. However, the nature and status of this involvement remains unsettled. Does disgust have any legitimate role to play in moral discourse? How much (if any) importance should we accord it when assessing moral transgressions, or when engaging in moral deliberation? And should we be skeptical about moral judgments that involve or are justified by appeal to disgust? The answers to these questions, I argue, are not as straightforward as they may seem. I begin by briefly surveying the psychological literature on the origins and nature of disgust, focusing on its various elicitors and behavioral outputs. I then discuss some experimental and empirical evidence that motivates suspicion of disgust’s role in moral judgment,

² In a similar vein, Kekes (1992: 431) writes that disgust “has not been much discussed in philosophical literature, or, indeed anywhere else.” I initially found these claims implausible, but a search of the database Web of Science shows that prior to 1997, 206 papers were published on the topic of disgust and 72 on the topics of emotion and disgust; from 1998-2011, those numbers are 1639 and 778, respectively. Philosopher’s Index lists 16 papers on disgust prior to 1997, and 0 on emotion and disgust; from 1998-2011 those numbers are 14 and 53, respectively.

before arguing that finding an argument to substantiate that suspicion is more difficult than one might expect. It is tempting to appeal to the fact that disgust often appears to bias and distort moral (and other kinds of) judgments, and to lead to irrational responses, as a reason to distrust both disgust itself and the moral evaluations in which it is implicated. But this line of argument threatens to end in a widespread skepticism about evaluative judgments.

The skeptical arguments I'll be challenging rest on empirical evidence about the nature of disgust: its evolutionary origins; the nature and plasticity of the disgust response; disgust's influence on moral judgment; and the way in which disgust represents its elicitors. Therefore, I spend section one surveying the relevant empirical literature, before moving on in section two to discuss whether these empirical findings can support skepticism about disgust. I argue that neither the evolutionary origins nor the plasticity of the disgust response are sufficient to reach any normative conclusion about the significance of disgust. I concede that disgust's influence on moral judgment provides an initially plausible basis for skepticism, but go on to demonstrate that the argument overgeneralizes, generating widespread skepticism about evaluative judgments. To avoid this unpalatable conclusion, the disgust skeptic might instead adopt an indirect strategy of examining the different possible roles disgust might play in moral judgment and demonstrating that it is ill-equipped to play any of these roles. However, in section three, I show that this strategy also runs into trouble. Drawing on the observation that one of disgust's central roles is to protect against contamination, I argue that disgust as a response to moral wrongness is plausibly construed as tracking social contamination.

1. Disgust: a Description

Disgust is one of the basic emotions, and it is present in all cultures (Rozin et al, 2000). It appears that disgust, like language, requires cultural input; there is some evidence that feral children do not develop disgust (Miller, 1997: 12-13). Psychologists (see, e.g., Fallon & Rozin, 1983) hypothesize that disgust emerges from the more primitive distaste response. Very young infants and animals will display an aversive reaction to bitter substances, a reaction which most likely protects against the ingestion of poisonous or harmful substances. But disgust goes beyond distaste insofar as it need not be a response to oral incorporation of an unpleasant taste; it can be elicited through multiple sensory modalities. Disgust also encompasses a global rejection of its object. Distaste does not. A flower, for example, might elicit a distaste response if it were put in the mouth and found to taste bitter, but this would not entail a rejection of the flower itself, just a rejection of the flower *qua* food.

Disgust represents its elicitor as “something revolting, primarily in relation to the sense of taste, as actually perceived or vividly imagined; and secondarily to anything which causes a similar feeling, through the sense of smell, touch, and even of eyesight” (Darwin, cited in Tybur et al, 2009: 103). The disgust response is characterized by a facial expression consisting of a kind of gape in which the labii levator muscles contract to raise the upper lips into a grimace. Other aspects of the response include a withdrawal from the elicitor, physiological feelings of nausea, including increased salivation, and feelings of revulsion, aversion, offensiveness, and the impression that the object is contaminating. Rozin, Haidt & McCauley (2000: 758) note that “contamination sensitivity is a basic feature of disgust.” Another feature distinguishing disgust from the

distaste response is that it endows the offending object with “contamination potency”: if something is disgusting, then so is whatever comes into contact with that thing. If one’s hand touches something disgusting, such as feces or vomit, it is contaminated. If a dead body is perceived as disgusting, then a sweater that has been worn by a dead body will also be perceived as disgusting.³

The recognition of the possibility of contamination by parasites or poison presented our ancestors with a “potentially crippling” problem: contaminants are typically so small as to be invisible, so our ancestors were forced to confront the fact that “everything we might eat or touch is potentially contaminated” and we have no way of knowing, in any given case, whether the object is contaminated (Rozin et al, 2000: 640; see also Fallon & Rozin, 1983). Disgust offers one solution to this problem by proscribing certain types of bodily fluids and products, decaying bodies, and foods. While humans consume only a small subset of available foods, there is a significant difference between foods deemed distasteful and those deemed disgusting—we may force ourselves to eat a few bites of a badly overcooked stew out of politeness, but a food which truly disgusts us will be hard to choke down and may even cause us to gag, thereby expelling it from our mouths, or, in extreme cases, to vomit, thereby expelling it from our body.⁴

³ This is one of the features that leads Nussbaum to complain that disgust involves “magical thinking”; see also Frazer’s discussion of the laws of contagion and similarity in chapter 3 of *The Golden Bough*. I discuss this point at length in section 2.4.

⁴ This is evident in recent discussions of the viability of insects as a food source—insects are actually ideal food in many ways, but their adoption as a food is up against the disgust many people feel when presented with the idea.

Disgust may have arisen to protect against literal physical contamination, but it also protects against perceived social and spiritual contamination. Rozin et al (2000) describe four types or domains of disgust: core disgust, which protects against contamination, and is directed at certain foods, body products or animals; animal reminder disgust, which is directed at things that remind us of our animal nature and therefore our mortality, such as sex, death, and “body envelope violations” (e.g. a gaping wound); interpersonal disgust, which protects the soul and the social order and is directed against contact with undesirable others; and moral disgust, which is elicited by moral offenses.⁵

Empirical studies of disgust suggest three ways it influences moral judgment: by amplifying/strengthening moral judgments; as a consequence of appraisals of moral wrongness; and by leading to appraisals of moral wrongness even when the actions in question are paradigmatically *nonmoral* (for a more detailed discussion, see Pizarro et al in press).

Several studies have shown that individual sensitivity to disgust correlates with both political attitudes and specific moral judgments. Inbar and Pizarro (2009a) found that higher disgust sensitivity correlates with political conservatism. Highly disgust-sensitive subjects also judged gay marriage and abortion to be *more* morally wrong than

⁵ These categories are descriptive and do not attempt to identify what really is or isn't disgusting, or which things or behaviors merit disgust. Tybur et al (2009) have called into question the utility of Rozin's four categories, proposing instead the categories of “mating, microbes, and morality.” In section 2.4 below I discuss doubts about the claim that reminders of animal nature would be experienced as aversive or as linked to mortality. Animals are worshipped in some societies, such as the indigenous tribes of the Pacific Northwest (for discussion see Levi-Strauss 1966), and Miller disputes the idea that disgust is concerned with reminders of death, arguing instead that it is often triggered by “the capacity for life” (1997: 40). Here I am mainly concerned to point out ‘core’ disgust and its typical elicitors, since these will be relevant in thinking about disgust's adaptive function. I discuss animal disgust and its relationship to mortality in greater detail below.

their less sensitive counterparts. In a subsequent study (2009b), the authors report finding that disgust sensitivity predicts negative attitudes towards homosexuality.⁶ Again, highly disgust-sensitive subjects evaluated homosexual behavior as *more* morally wrong than their less easily disgusted counterparts. Schnall et al (2008) found that subjects exposed to a disgusting smell judged a variety of moral violations as more wrong than did subjects not exposed to the smell. They also had subjects fill out their surveys at either an extremely messy or a clean desk; again, moral judgments were more severe in the disgusting condition. This suggests that the presence of disgust intensifies the moral judgment— it's not that subjects in the nondisgusting conditions didn't judge the transgressions morally wrong, they just judged them less harshly than their disgusted counterparts.

Chapman et al (2009) found that, when subjects playing an ultimatum game were presented with unfair offers, they responded with the facial expression characteristic of disgust. Haidt (2001) has argued that moral judgments are affectively laden intuitions, which do not result from reasoning but are caused solely by (and therefore are distinct from, though a direct result of) our emotional response to the act in question.⁷ However,

⁶ It should be noted, however, that this study used implicit measures of attitude-- that is, subjects were not explicitly queried about their attitudes towards homosexuality; instead, these were inferred from subjects' performance on an implicit association test (IAT). The use of the IAT might be problematic insofar as the subject himself is often unaware of the attitude or bias revealed by the IAT-- for example, virtually everyone, regardless of race, reveals a bias favoring whites over African-Americans. Would we therefore want to impute this attitude to everyone? Part of the difficulty here is that there's often ambiguity over the type of attitude we're interested in (implicit, explicit, consciously accessible or not). This is a project worth pursuing in its own right, though I don't discuss it here.

⁷ As a rule, it is difficult to empirically disentangle the claim that disgust is an affective response to the perception that a moral violation has occurred from the claim that disgust just *is* the moral judgment, since the question turns heavily on one's theory of what a moral judgment is and under what conditions we are willing to attribute a judgment to someone (for example, certain sorts of sentimentalists might be happy to concede that moral judgment consists in an affective response; others might insist that since a moral judgment is a belief, affect is insufficient). I would like to remain as neutral as possible on the question of

some studies have gone beyond documenting that disgust occurs as a consequence of moral judgment and attempted to show that disgust suffices to bring about appraisals of moral wrongness, even in the absence of any violation that could plausibly be described as morally wrong.

In a striking demonstration of the power of disgust, Wheatley and Haidt hypnotized subjects to feel disgust at certain innocuous, affectively neutral words, such as 'often' and 'take'. Subjects were then presented with a series of vignettes and asked to rate each for wrongness. When the vignette contained the word 'often' or 'take', subjects judged it more wrong than when it did not. Surprisingly, however, this effect persisted even when subjects were presented with vignettes containing no moral content whatsoever: when given the case of Dan, a student council representative who "often" tries to bring interesting speakers to campus or "takes" topics of interest to the other students, subjects who had been hypnotized judged Dan's actions *morally wrong*. When asked, subjects sought to justify this judgment, saying things like, "it just seems like he's up to something" and suggesting that Dan might be "a popularity-seeking snob" (2005: 783). This suggests that the feeling of disgust is sufficient to bring about an appraisal of moral wrongness even in the absence of a moral violation.⁸

which account of moral judgment is correct; my concern is primarily with whether disgust can be plausibly accorded or denied a role in any such theory, and part of what I suggest below is that the difficulties one encounters in pursuing this question are not unique to disgust or even to sentimentalist theories of moral judgment—they extend to rationalist theories as well. Hence I try to stick to the less loaded term 'appraisal' when possible.

⁸ An objection that immediately presents itself is that disgust triggered by nonmoral objects, such as rotten meat, doesn't result in a moral judgment. I discuss this point and its significance below.

2. Disgust as Distortion?

The descriptive claim is well-established: as the studies above demonstrate, disgust *is* often implicated- as an input to, output of, and/or concomitant with- in moral judgment. But ought we trust it as a guide to moral wrongness? Does the presence or absence of disgust signify anything about the presence or absence of a moral violation? If not, then moral judgments based on or involving disgust will, in the absence of independent corroboration or justification, be called into question.

My goal in the next section is not to show how to generate a skeptical argument against affect-driven moral judgments—though those looking for a skeptical argument for global skepticism about evaluative judgment may find some suggestions below. Instead, I want to show how not to generate a skeptical argument against disgust. (I assume throughout what follows that global evaluative skepticism is undesirable and something to be avoided if possible.) Part of my aim is to show how claims about the biasing or distorting effects of an emotion must appeal to specific features of the emotion in question, and to show why those specific features make it especially likely to distort or bias evaluative judgment in problematic ways. The ubiquity and potency of affective, cognitive, and metacognitive influences on moral judgment means that the skeptic can't just gesture at the tendency of these influences to distort judgments and leave it at that, on pain of global evaluative skepticism. Disgust skeptics, I argue below, owe us- but have thus far failed to give- a sufficiently detailed and empirically supported basis for skepticism about disgust-based moral judgments. I'll consider four arguments. The first three appeal to features of disgust's etiology and its role in moral judgment, and as such, fall prey to overgeneralization worries. The final argument I consider relies on features it

claims are specific to disgust. It therefore has more promise than the first three; unfortunately, it fails because the empirical support it requires is lacking.

2.1 Moral Disgust as Auxiliary Disgust

One consideration that has been invoked to discredit disgust is the fact that it originally evolved to track contamination by poisoning or parasites, and was later “co-opted” to track moral and social violations. Its role in moral judgment is therefore an auxiliary role, or a byproduct of its original adaptive function. Byproduct hypotheses typically proceed by showing that some trait originally evolved to serve one function and later was recruited to serve another; therefore, what appears to be a poor design or performance is actually a result of the trait serving a different purpose from the one it originally evolved to fulfill. In the case of disgust, the hypothesis is that it evolved to track the threat of physical contamination—which was, in ancestral environments, a real and problematic threat, especially in light of the difficulty of detecting contaminants—and later was recruited to enforce moral and social norms by presenting violations and violators as “tainted, contaminating, immoral and somehow less or lower” (Kelly, in press: 157). Because it did not evolve to track them, disgust is somehow mistakenly targeted at social and moral violations: “In revealing [moral judgments linked to disgust] as byproducts... this view also suggests that these aspects are also baseless” and that therefore “this view shows that in these cases, feelings of offensiveness... even when vivid, are misplaced” (ibid.)

This argument fails because it commits a kind of naturalistic fallacy, by confusing the evolutionary purpose or origin of a trait or response with its “proper” place or current

purpose. Consider parental love, which evolved to compel parents to make potentially costly investments in their offspring and kin. This emotion is often co-opted, however, insofar as parents feel love for, and invest in, genetically unrelated children; in some cases, parents exert tremendous energy and expense to acquire children who have no genetic relation to them. From an evolutionary perspective, this is irrational. Feeling parental love for a child from halfway across the world would seem to be, in Kelly's term, a "mismatch."

Likewise, Tooby and Cosmides (2005) have argued that humans have cognitive capacities that evolved for the purpose of reasoning about social exchange and coordination. One of these capacities is the ability to tell when rules or contracts have been violated—a capacity that requires conditional reasoning. If that's right, then the ability to do formal (and informal) logic may be a byproduct of this capacity, yet this does not make us skeptics about logic.

Byproduct hypotheses can be useful in demonstrating why some behavior or capacity regularly goes awry. Yet the fact that something is a byproduct does not show that it does, in fact, go awry. In order to know this, we would need to have some independent or antecedent method for checking correctness. But in the case of moral judgment, this kind of method is absent—the question of how to verify moral claims (or whether they can be verified at all) is vexed, and does not look to be settled any time soon. Without such a method, the fact that disgust is a byproduct tells us nothing about the probable accuracy of disgust. Thus the argument based on the fact that moral disgust is some sort of auxiliary disgust or side-effect of natural selection fails to lead to a normatively significant conclusion.

2.2 The Plasticity of Disgust

Christopher Knapp (2003) argues that the plasticity of the disgust response rules out any possibility of using disgust to “anchor” an evaluative discourse. Knapp’s argument rests on the fact that the disgust acquisition mechanism is “designed to be triggered by different things in different people,” as evidenced by the evolutionary origins and ontogenic development of the disgust response. If disgust is to support any normatively loaded discourse, Knapp contends, “it must be the case that the properties that make something fit-for-disgust are subject-invariant” (2003: 274). And they are not. Therefore, Knapp argues, “the structure of the disgust response is too plastic to anchor an evaluative property.” (2003: 277) If that’s right, there’s no sense in debating whether something actually is disgusting, and no interesting “evaluative conception of disgustingness”: debates over whether something is disgusting are “pointless, or uninteresting, or both” (276). Normative disputes are not like this, however, so there can be nothing normative about disgust ascriptions. To report finding something disgusting is to report some fact about one’s own (likely idiosyncratic) reaction to it, but this is not sufficient to ground an evaluative judgment, much less a moral one.

There are two problems with this line of argument. The first is the move from design to plasticity, and the second is with the claim that plasticity precludes normativity. To see why the first move is dubious, consider Knapp’s analogy with language. On the Chomskian picture, the innately prepared mechanism via which we acquire language is designed to be triggered by the local language. Which specific language we end up with is dependent on which language we are exposed to during a constrained/limited developmental window. But this does not necessarily mean that the mechanism is

plastic, insofar as it can only acquire the grammar(s) it's exposed to during that period. (Though it can, of course, be augmented; one can learn a second, third, or fourth language, but this is a much more effortful process.) This is because language aims at coordination. The mechanism can only acquire whichever language it is exposed to during a specific window, and this limited acquisition period facilitates coordination. The mechanism is plastic in that it requires cultural or environmental input to operate effectively.⁹ The same can be said for the disgust acquisition mechanism: it may be plastic insofar as it requires environmental or cultural input in order to deliver a set of outputs, but this can also be put as: it is designed (at least, as Knapp describes it) so that the outputs it delivers are fixed by something very specific: the prevailing cultural norms.

Even if disgust were completely plastic, would Knapp's denial of normativity follow? I think not. At the very least, it would not follow straightforwardly, for two reasons. First, the biological or innate mechanism by which some norm, emotion or faculty is acquired is not the sole source of constraints on that norm, emotion, or faculty. It may be that the faculty through which moral norms themselves are acquired is infinitely plastic, but that considerations about the nature of human sociality and cooperation constrain—perhaps even uniquely—which norms are available in the environment for acquisition. Cultural transmission, unrelated cognitive limitations, and other environmental factors can all narrow the set of available inputs. This brings us to the second flaw in Knapp's argument, which is its failure to explain why some response has to be “subject invariant” to ground an evaluative property. As Gert (2007: 347) observes, “Knapp's markers of evaluative properties turn out to be nothing more than

⁹ For a detailed description of the French doctor Jean-Marc Itard's efforts to educate a feral adolescent in the early 19th century, see Lane 1976.

markers of objective properties.” This suggests that Knapp equates normativity with objectivity, but this is tendentious to say the least—absent additional argument, that some facts/judgments/claims may be both relative and normatively significant. Indignation, for example, is aroused in different people by different things, and yet it plainly evaluates its object. So plasticity alone is insufficient grounds for denying disgust a role in evaluative or normative discourse.

2.3 Disgust as Distortion

One might be tempted to argue that based on the experiments discussed above, it is simply self-evident that disgust distorts our moral judgment and is therefore not to be trusted: if the mere occurrence of disgust can lead to the judgment that bringing interesting speakers to address a campus group is morally wrong, then how can we trust judgments of wrongness that stem from or co-occur with a feeling of disgust? The answer, some have suggested, is that we can't. Disgust is, at bottom, untrustworthy. And in the pantheon of emotions, disgust is thought to be especially and uniquely untrustworthy: Nussbaum (2004: 13) argues that “shame and disgust are different from anger and fear, in that they are especially likely to be normatively distorted” while Kekes (1992: 438) points out that disgust is “fickle” and therefore may be unreliable. Using disgust as any sort of guide to immorality is “problematic and irrational... across the board” (Nussbaum, 2004: 102).

The problem with this argument isn't that disgust doesn't cause distortions in judgment or cognitive biases—it's that so many other emotions do. For example, Isen (1984; see also Levine and Pizarro, 2006) found that inducing positive emotions made

subjects significantly more likely (72% vs. 40% in the control condition) to give an erroneous, but intuitively plausible, answer to a physics problem. More generally, positive affect correlated with subjects' reliance on heuristics to solve problems, and led to more erroneous responses to a wide variety of questions. Schwarz (2002) surveys a variety of ways in which emotions influence cognition; subjects experiencing negative emotions are more likely to engage in bottom up, data-driven processing.

More specifically, anger and sadness have implications for cognition: anger makes people more likely to attribute their misfortune to an individual rather than nonagentic factors, while sadness has the opposite effect (see, e.g., Keltner, Ellsworth, & Edwards, 1993). Anger also increases reliance on stereotypes (Bodenhausen et al, 1994; Park and Banaji, 2000). Sadness, on the other hand, leads people “to rely less on stereotypes and to perform better on deductive reasoning tasks” (Levine and Pizarro, 2006: 8).

Emotions affect moral judgment even when they are not directed at the judgment as such: Valdesolo and DeSteno (2006) found that subjects who had viewed a humorous video were more likely to favor a utilitarian solution to a moral dilemma. Strohminger et al (in press) demonstrate that while mirth made subjects more likely to favor utilitarian solutions over deontological ones, another positive emotion, elevation,¹⁰ made subjects less likely to do so. A third emotion, empathy, may explain the rather bizarre finding that as the number of people harmed by an action increases, the punishment that subjects—and real-life jurors—assign to the perpetrator *decreases*. So the punishment assigned for

¹⁰ Strohminger et al (in press) explain that elevation comes from “witnessing acts of moral beauty,” and is associated with “reverence”; experimentally, the emotion is induced by giving participants stories from *Chicken Soup for the Soul*, or, in another experiment, by showing footage of crowds on the night of Obama’s election.

defrauding three people is actually *more* severe than the punishment assigned for defrauding thirty—both in the lab and in the courtroom (Nordgren & McDonnell, 2010). One explanation is that it is easier to empathize with three victims than with thirty, and it is this increased empathy that causes the increased punishment. In this case, empathy leads us to make irrational decisions—surely harming thirty people is worse than causing that same harm to three people. But to rule out moral judgments based on or involving empathy would be to rule out a lot.

Concluding that because of disgust’s role in cognitive biases it has no legitimate role to play in moral judgment would mean concluding that other emotions, both positive and negative, likewise have no legitimate role to play. And this would be trouble for noncognitivists and sentimentalists, among others.

One might think, ‘so much the worse for sentimentalism’. Josh Greene (2001, 2008) and Peter Singer (2005) have both suggested that we should throw out or discount affect-driven moral judgments as unreliable and adopt a utilitarian consequentialist view, since the latter has its origins in, and is supported by, reason rather than affect. But even if one were inclined to grant this dubious dichotomy, why think reason is any more reliable than affect as a guide to morality? In particular, why think reason any more innocent of distortion or bias than emotion? In fact reason is just as susceptible to this line of argument as emotion: it is not only affect, but also cognitive and metacognitive states that influence our moral judgments in ways that might be called ‘distorting’. If emotion is guilty of distorting judgments, then reason is not innocent in this respect. As Hauskeller (2006: 578) notes, “The problem of distinguishing between mere prejudice and legitimate concerns is not less of a problem for the rationalist than it is for the

sentimentalist.” So the argument threatens skepticism regardless of whether one’s theory locates moral judgment in affect, reason, or both.

Consider the relationship between spatial location and evaluative judgment. Casasanto (2009) found that when presented with a pair of job applicants, products, or alien creatures, subjects tended to identify the ‘good’ member as the topmost member, or the rightmost (when the pairs were oriented horizontally). It’s hard to see how spatial location is morally relevant. Yet it is also ubiquitous. In real life as well as in many experimental designs, objects are definitely positioned relative either to ourselves or to each other (and note that even where the prompt is a single sentence smack in the middle of the page, the subject is typically asked to respond by circling a point on a Likert scale—which is arrayed left to right). Indeed, Laham, Alter and Goodwin (2009) found that something as innocuous as typeface influenced subjects’ moral evaluations (see also Schwarz, 2004; Song & Schwarz, 2010). Anger, sadness, empathy, knowledge of the number of victims, spatial location, and typeset— all these factors lead to cognitive biases and instances of what appear to be distortions in reasoning and/or judgment. Yet they are ubiquitous enough that skepticism about judgments involving these factors would amount to skepticism about most- if not all- moral judgments.

The moral of the story is this: we ought to be extremely wary of inferring from the fact that seemingly irrelevant framing effects can affect a judgment to the conclusion that judgments of that type are unwarranted. Avoiding wholesale skepticism requires adopting a more refined treatment of framing effects, biases and ‘distortions’. The fact that a moral judgment is influenced by a seemingly irrelevant emotion, piece of information, or other feature, licenses us to conclude neither that that feature itself is

always distorting, nor that the resultant judgment is untrustworthy. While a principle that dictates when and how affect results in an untrustworthy judgment or appraisal is desirable, the goal of this paper is not to arrive at such a principle. Indeed, the considerations discussed above reveal the difficulties any attempt to do so would face. Most likely, we will need to draw our distinction more finely, perhaps approaching the issue on a case-by-case (emotion-by-emotion) basis. The next argument I'll discuss illustrates how the skeptic can avail himself of this strategy. It appeals to features particular to disgust in an attempt to establish skepticism about disgust without implicating other emotions or cognitive processes.

2.4 Disgust as Magical Thinking

Witches don't sink in water. How do I know? Because there are no witches. People used to think there were; they attributed magical powers to certain individuals (usually women) and accused these individuals of being witches. But then they figured out that there really was no such thing as magic- hence Frazer's and subsequently Nussbaum's pejorative use of the phrase 'magical thinking'- so nothing had the property of having magical powers. Given the centrality of magic to the concept of a witch, they concluded that there were no witches. An analogous move in the case of disgust would identify a property or feature that is central to our concept <disgusting> and show that nothing has that feature. Therefore, disgust, in attributing disgustingness to its object, is systematically misleading, erroneous even. In short, we have arrived at an error theory of disgust.

What is unique to disgust, recall, is its tendency to elicit a sense of offensiveness, contamination, and contamination potency. Moral disgust is therefore based on magical thinking—it represents a kind of category mistake. Nussbaum introduces her disgust skepticism by raising precisely this concern:

Disgust, I shall argue, is very different from anger, in that its thought-content is typically unreasonable, embodying magical ideas of contamination, and impossible aspirations to purity, immortality, and nonanimality that are just not in line with human life as we know it.¹¹

Nussbaum argues that it is not its attribution of contamination potency per se, but associated beliefs about animality- in particular disgust’s role as a reminder of our animal nature- that is problematic: “Because disgust embodies a shrinking from contamination that is associated with the human desire to be nonanimal, it is frequently hooked up with various forms of shady social practice, in which the discomfort people feel over the fact of having an animal body is projected outwards onto vulnerable people and groups. These reactions are irrational, in the normative sense, both because they embody *an aspiration to be a kind of creature that one is not*” (2004: 74, emphasis mine) (Notice also that if this line of reasoning is correct, it is no mere accident of history that disgust has motivated morally problematic policies.) Nussbaum’s argument seems to be:

1. Disgust embodies contamination fears associated with a desire to distance ourselves from our animal nature

¹¹ Nussbaum thinks that emotions have cognitive content and so are subject to evaluation as reasonable and unreasonable. See also D’Arms and Jacobson’s (2000) discussion of the ‘fittingness’ of emotions, and in particular disgust. While D’Arms and Jacobson disagree with Nussbaum’s claim that emotions have cognitive content, and deny that they embody judgments, they nonetheless agree that judging something disgusting involves presenting that thing as possessing certain properties or attributes; disgust and other emotions can represent objects correctly or incorrectly. One does not, therefore, have to adopt a cognitive account of emotions in order to view their occurrence as being more or less warranted, and therefore amenable to criticism.

2. Our animal nature is a source of anxiety because it reminds us of “a type of vulnerability that we share with other animals, the propensity to decay and become waste products ourselves.” (Nussbaum, 2004: 92)
3. Disgust polices the boundary between human and animal nature
4. Thus disgust presents its object as less- or other-than human; it dehumanizes its object
5. This in turn permits us to treat those individuals or groups towards whom we feel disgust as less-than-human.

The above argument, if successful, establishes that the link between the content of the disgust response and the dehumanizing treatment associated with moral disgust is not just historical, but conceptual.¹² I will take up the question of contamination in section 3; here I focus on the claim that disgust is a response to anxiety about our animality and mortality. Haidt et al (1997: 110) have argued that “most of the objects that meet... [Americans’] definition of disgust are indeed animals (including humans)... or are disgusting by virtue of their contact with them,” and Rozin et al (1993) argue that at its core, the emotion is really “animal-origin disgust,” and that other types of disgust are elaborated out of this core concern. Our revulsion to animals and reminders of our animal nature is, according to Rozin, a way of avoiding reminders that we, too, are animals, and as such, are mortal. Whereas Haidt et al, and Rozin, include humans as part of the category of ‘animals,’ Nussbaum associates the term with a kind of dehumanization: to view someone as animal is to view them as less than a person.

Unfortunately, like Haidt and Rozin before her, Nussbaum provides little support for the claim that disgust is about avoiding reminders of death and decay per se as opposed to avoiding them because death and decay tend to present contamination dangers

¹² Compare: the connection between romantic love and greeting cards, though empirically robust, is not conceptual; rather, it’s mediated by a series of historical events that, while partly enabled by some of the features of romantic love (it motivates people to make costly declarations and displays of their feelings) don’t reveal anything deep and meaningful unique to romantic love itself.

of their own. Nor does she fully explain why these concerns should be linked to animals. As she acknowledges, certain traits such as “strength and speed, and the animals who exemplify them” (2004: 92) are admired. Many cultures have celebrated, rather than resisted, the continuity between human and animal, either by deifying animals or by constructing cosmologies in which humans are descended from or created by animals (see Levi-Strauss, 1966). As Miller (1997: 46) points out, if disgust were aimed at distancing reminders of our animal nature, we’d expect that animals most like us would be most disgusting, and this is simply not so. And “animal secretions,” which both Rozin and Nussbaum suggest are the properties in virtue of which animality and death and decay become linked, have been used by many cultures in magic and religious rituals, often as a means of purification (in rituals involving sympathetic magic, it is sometimes assumed that only the contaminant itself can remove contamination; this suggests that contamination is a primary rather than an incidental concern of such performances).¹³ Miller has suggested that disgust is triggered by a superabundance of life, rather than reminders of death (1997: 40-42), but both claims, and the evidence used to support them, can be explained in terms of contamination concerns without adverting to claims that they trigger anxiety by reminding us either of death or of life.

Nussbaum might respond that contamination alone cannot account for disgust’s tendency to lead to the dehumanization of certain individuals or groups. But this tendency is not sufficient to establish the causal conclusion Nussbaum needs. She argues that “Disgust... revolves around a wish to be a type of being one that one is not, namely

¹³ The earliest blood transfusions were from animals to humans, forcing people at the time to grapple directly with the question of whether humans and animals were essentially different; see Tucker (2011). Tucker notes that at the time, animal blood was actually often thought to be more pure than human blood.

nonanimal and immortal,” and that concerns about contamination are instrumental in our attempts to realize this wish: “thoughts about contamination serve the ambition of making ourselves nonhuman,” (2004: 102) but why not think instead that an aversion to bodily fluids and decaying flesh (both human and animal) and to snakes and insects (which Nussbaum argues signify decay, and Miller argues signify an overabundance of life) both serve the ambition of protecting ourselves from contamination? If that’s right, then dehumanization and ostracism aren’t implicated in disgust’s cognitive or conceptual content/nature/structure; rather, they are associated with disgust insofar as they are effective, psychologically and physically, at removing or warding off the threat of social contamination. To show that moral disgust is intrinsically dehumanizing, the skeptic needs to demonstrate disgust’s tendency to dehumanize or lead to the ostracism of its elicitors as necessarily a result of the way it presents its elicitor- in this case, as less than human- rather than as an optional reaction to some third, mediating presentation (in this case, contamination potency). But isolation is, in the absence of a cure, an effective way of preventing the spread of contamination. Thus there is no strong argument for thinking of disgust as necessarily dehumanizing rather than as capitalizing on/resulting from our capacity to dehumanize others (see Zimbardo, 2007; Doris and Murphy, 2007) as a strategy for avoiding potential contaminants. Therefore this route to an error theory fails because we have no reason to think that disgust is committed to magical thinking or some kind of superhuman aspirations on our part.

3. Doubting Disgust

The disgust skeptic at this point can adopt a less direct strategy. Rather than advancing a skeptical argument, they can challenge the disgust advocate to more specifically state how disgust is involved in moral judgment and then argue against the proposed account. If no plausible role for disgust can be found, the skeptic's case is bolstered; certainly, the burden of proof is shifted back onto the disgust advocate to come up with a more promising proposal. In this section I discuss three ways the disgust advocate might spell out the nature of disgust's involvement in normative assessment in order to vindicate its role in moral judgment. The first two accounts fail, though for different and illuminating reasons. The third, which takes up the issue of contamination potency, may succeed in providing an account of how and why disgust is a justified response to moral wrongness, though it leaves open legitimate concerns about whether disgust is a particularly efficacious way of coping with it.

3.1 Disgust as Constitutive of Moral Judgment

Most accounts of moral judgment accord at least some role to affect. Disgust is a specific type of affect. It also, as Nussbaum and D'Arms and Jacobson point out, presents its objects a certain way, so it's possible to give a cognitivist analysis of disgust discourse.¹⁴ Perhaps, then, disgust *is* the moral judgment? Above, I mentioned that Haidt suggests that moral judgments are no more than our affect-laden intuitive responses to moral violations. If this is right, maybe to be disgusted by something just *is* to judge that is morally wrong. Again, above I examined some evidence suggesting that disgust

¹⁴ Though not necessary; see note 10 above.

generates appraisals of moral wrongness; perhaps the disgust advocate can offer an account on which disgust doesn't have a role in moral judgment; it just is the moral judgment. To get this proposal off the ground, its proponent will have to explain why we are not constantly judging objects on the street, our own bodily functions, babies' diapers, and so on immoral. Cockroaches, though repugnant, are not immoral. Likewise for some foods—types of offal, for example. So the claim that the disgust reaction is constitutive of moral valuation requires distinguishing these types of disgust from the kind of disgust we feel for the child rapist, or the businessman who embezzles from a charity. An alternative approach might be to appeal to the fact that the cases involve persons and their actions toward one another- that moral disgust is distinguished by its concern with people's attitudes and behaviors- but this will not do the trick. We feel core disgust towards people, and fail to feel moral disgust towards them, even where this is elicited by a person or an action. We may feel disgust when looking at a burn victim, or I might feel disgust watching someone vomit, but these are not moral judgments. Indeed, some conduct that disgusts us is morally neutral or even lauded, perhaps even more so because it is disgusting: sanitation workers and nurses perform jobs that are disgusting but socially necessary.

3.2 Disgust as Evidence of Moral Wrongness

Another approach would be to take disgust as evidence of moral wrongness. This may be what Leon Kass has in mind when he writes that though disgust is “not argument,” nonetheless “repugnance is the emotional expression of deep wisdom, beyond reason's power fully to articulate it” (1997: 18). In a similar vein, Kahan (1998) has

argued for allowing disgust to play a role in criminal sentencing, with murders that cause disgust receiving higher sentences, on the grounds that disgust helps identify legally significant features such as excessive cruelty; Nussbaum writes that Kahan takes disgust to be “a useful... criterion, giving us information that is relevant to the legal regulation of certain types of acts” (2004: 85). If a murder that contains elements that provoke disgust deserves a stiffer sentence than one that does not, presumably this is because it is *more wrong*, as evidenced by its disgustingness.

The problem arises when we ask in virtue of what, exactly, disgust evinces wrongness. Consider guilt again: as an emotion, guilt carries with it a motivation to make reparation, which is prima facie evidence that one believes one has done some kind of damage which needs to be repaired. Guilt therefore tracks a kind of damage that’s linked to moral violation, and therefore is prima facie evidence of moral wrongness. Disgust, on the other hand, motivates withdrawal; if it is evidence of anything, it is of a contamination threat. So one way of spelling out the evidential link is: disgust tracks contamination and therefore immorality. But contamination is far from obviously immoral. Indeed, the kind of contamination much disgust protects us from is patently amoral. Isn’t this a problem? And doesn’t treating people as contaminants seem like a kind of category mistake?

3.3 Poisons, Parasites and... Persons?

The claim that disgust is evidence of wrongness is unconvincing unless it can identify some property that disgust is tracking—and show that some things do have that property. Earlier I considered- and rejected- Nussbaum’s suggestion that disgust’s

central role was tracking violations of the human-animal divide; now I want to consider another proposal: disgust tracks contamination. This is not accidental, but one of disgust's primary functions. The idea of contamination potency is part of what distinguishes disgust from mere aversive distaste, and it is one reason why disgust is so effective at motivating avoidance of "poisons and parasites" (Kelly, in press). As we saw above, Kelly argues that moral disgust's presentation of its elicitors as polluting or contaminating is systematically misleading. This should be distinguished from the claim that disgust itself is sometimes misleading; based on its evolutionary role in protecting us from "pathogens and poisons" (Kelly, in press), we should expect disgust to be 'cautious' and overgeneralize. But this moral disgust overgeneralized to the point of being co-opted into a whole new domain, one which it was never designed to handle; therefore, moral disgust committed a category mistake. Conduct or behavior, the purview of morality, is not the kind of thing that can be polluting or contaminating. But this, as I'll now show, is false.¹⁵

That the language of contamination is used to describe social phenomena is, as noted above, undeniable: drug use is 'an epidemic'; marketing campaigns, videos, and songs are 'viral'. But is this language metaphorical? Do behaviors and beliefs spread through contact and exposure? In answering these questions I will side-step the issue of how best to define disease (though for a good overview see Murphy, 2010), since my interest is less in the etiology of the phenomena under consideration and more on their transmission.

¹⁵ Kelly could redefine 'contamination' so that it is a property only of physical or biological entities, and thereby evade the criticisms I raise here, but doing so risks rendering his discussion of moral disgust vacuous.

We also need to distinguish the related questions of whether thinking of behaviors as contaminating leads to the adoption of morally problematic policies, or whether the language of contamination is the most effective approach to changing patterns of behaviors or attitudes. The conflation of these questions is an instance of what D'Arms and Jacobson dub 'the moralistic fallacy': "the question of whether F is fitting in the sense relevant to whether its object X is ϕ ... is indeed logically distinct from the moral, prudential, and all-in practical questions... considerations which bear on these further questions are irrelevant to property ascription." (2000: 71) Identifying these type of considerations is not the goal here; I won't be aiming to answer the question of whether disgust is the *all things considered* right response to wrongness. Rather, I will argue that contrary to what some disgust skeptics have claimed, there are behaviors and attitudes to which disgust is a fitting response. Some behaviors and attitudes really do have the properties disgust ascribes to them, namely, contamination potency.

Above I suggested that if disgust dehumanizes or leads to the ostracism of its object, it may be because isolation or ostracism are in fact often the most effective way to guard against contamination. Therefore, I argued, disgust may co-opt a preexisting propensity on our part to dehumanize other groups or people. But the language of contamination could itself be essentially linked to this dehumanizing or ostracizing attitude, in which case it, too, would be essentially problematic. So if a tendency to present its elicitor as less than human and to motivate ostracism of that object is essential to moral disgust, then the disgust skeptic might argue, via a strategy similar to the error-theoretic argument discussed above, that there are no behaviors or attitudes that merit

moral disgust. Before asking the question of whether actions or attitudes can be contaminating, then, it's worth pausing for a moment to consider this possibility.

Certainly the language of disease and contamination has been invoked to justify policies ranging from questionable to reprehensible. The idea of crack as an 'epidemic' was used to motivate the passage of stringent and punitive laws— from 1986 until recently, crack cocaine was the only drug for which a first-time possession conviction carried a federally mandated minimum sentence. In upholding New York State's exceptionally punitive drug sentencing laws (the 'Rockefeller Laws') the appellate court (in *People v Thompson*, 83NY 2nd 3/30/94) wrote that "drug-related crimes may be much more prevalent, that is, have a higher and rising incidence, than other crimes comparably punished or equally grave crimes not as severely punished, requiring greater isolation and deterrence," appealing to the idea that drug use and dealing is a greater social threat because of its propensity to spread from person-to-person than a similarly "grave" crime (under the Rockefeller laws, a first-time possession conviction could lead to a stiffer sentence than a rape or homicide conviction). This might seem like sufficient evidence that an attitude of contamination is the wrong way to approach moral or social problems such as drug abuse.

But leaving aside the tricky question of the most effective policy solutions to drug problems, let me make a few points on whether the language of contamination is essentially or necessarily linked to more punitive or ostracizing attitudes. Thibodeau and Boroditsky (2011) found that when subjects were given a description of a town with a crime problem, presenting the problem in terms of a disease or 'virus' actually made subjects more likely to suggest adopting policies that involved social reform and

rehabilitation; they were also more likely to describe the primary role of a police officer as preventing crime, educating children, and serving as role models (as opposed to capturing and punishing criminals). These results, though of course far from conclusive, suggest that presenting behaviors as contagions may lead to a less punitive view of those behaviors and shift the focus from punishment to prevention. If the problem with the proposal that disgust tracks contamination isn't that it necessarily results in the adoption of problematic attitudes or policies, the next question to ask is whether there is any factual support for the claim. Is moral disgust, in presenting moral violations as contaminating, committing some sort of category mistake? Should we read the language of disease and contamination here as merely metaphorical, or is there evidence that behaviors and attitudes, and the people who possess them, are contaminants in much the same way as poisons and pathogens?

The Framingham Heart Study began in 1948 with 5,209 men and women between the ages of 30 and 60; its aim was to identify the causes and correlates of cardiovascular disease. In order to do so, researchers obtained detailed physical and behavioral data on each of their subjects; more importantly, they followed up on their subjects every two years, and subsequently enrolled the children of these subjects in the study: in 1971 5,000 offspring of the original cohort joined the study, and in 2002 a third generation was added. All in all, the researchers obtained detailed medical and behavioral data on almost 15,000 subjects from the town of Framingham, Massachusetts. On intake forms, subjects were asked to "name a close friend who can find you in case we can't." (Couzin 2009)

Researchers subsequently realized that the inclusion of this question meant that the data could be mined for information on how behaviors, attitudes, and other conditions

such as obesity spread through social networks. They found that subjects' happiness, and their chances of becoming obese, quitting smoking, or getting divorced, all increased as their friends developed these conditions or behaviors (or, in the case of happiness, increased as their friends' happiness increased). They also found that the behaviors spread not just between friends but between friends of friends. So, for example, an individual is 50% more likely to drink heavily if a friend does so as well, but the effect remains strong even at two degrees of separation—if a friend of a friend drinks heavily, the individual is 36% more likely to do so. The same effect holds for all the factors examined so far—from obesity to divorce to happiness. (Physical proximity doesn't explain the effect; the behavior of neighbors, for example, had no significant effect.) This data shows how behaviors spread through a population, and demonstrate that one need not even come into direct contact with a person to be affected by their behaviors or attitudes—third parties (mutual friends) can act as 'carriers'.

This last finding is the truly striking one. That people are influenced by the behaviors of those around them isn't surprising; after all, it's what makes imitative learning possible. But that behaviors spread not just through direct contact but through contact with those who have come into contact with it is a powerful evidence that they do have contamination potency—any object- or in this case, person- that has come into contact with or been exposed to the behavior has the power to contaminate others. This result is not specific to alcohol—the findings have been replicated with respect to obesity, smoking, divorce, happiness, and suicide.

Nor are the results specific or unique to this study. The idea that behaviors or attitudes can be a kind of social contagion has informed some of the theoretical work on

racism or prejudice (e.g., Allport 1954) and recently, Grim et al (2005) show, using computational modeling, that contact with other groups both reduces prejudice and thwarts the spread of prejudice in the first place. The contagion effect of suicide has been noted and remarked upon for centuries- it's sometimes referred to as the 'Werther effect', after the wave of suicides following in the wake of Goethe's *Young Werther*, in which the protagonist kills himself¹⁶- and is potent enough that most modern media outlets purposely play down reports of or simply refrain from reporting suicide and refrain from giving details of the method used, so as not to encourage imitators. Positive attitudes and emotions also spread; as Hill et al (2010: 3828) note, "happiness could be thought of as a form of social infection." Cooperative behavior, as measured by performance on public-goods games, also demonstrates a contagion effect; subjects playing with generous individuals were more likely to behave generously in future interactions, even when these were played with different individuals than the prior round. Most striking, perhaps, is that subjects' behavior was influenced, not just by whom they played with, but by whom the people they played with had played with. That is, the effects of cooperation could be seen up to three degrees of separation.

Behaviors and attitudes spread through a group in much the same way a disease does. And while an important disanalogy might seem to be the fact that germs are invisible, while our friendships are not, the fact that these behaviors and attitudes affect parties who don't come directly into contact with them means that these, too, have the potential to be invisible contagions.

¹⁶ The effect was first named by Phillips (1974: 340), who quotes Goethe: "my friends... [thought to] imitate a novel like this in real life and, in any case, shoot themselves; and what occurred at first among a few took place later among the general public."

The disgust advocate's strongest argument, then, is that moral disgust tracks social contamination. But as mentioned above, not all contagion is *bad* contagion. Contagion itself is neither good nor bad; it is the attitudes and behaviors in question that make something contaminating in a deleterious sense. In the case of disease, these deleterious consequences are rather easily observed, and tend not to be controversial—most people would agree that pain, vomiting, weakness, and death are all bad things. But in the case of moral contamination matters are less settled: the extent and depth of controversy over many first-order moral questions means that what one man or group considers a contaminant, another might consider laudable and worth encouraging. The question of whether some behavior or attitude has the kind of contamination potency disgust ascribes to it, then, can be disambiguated into the question of whether the behavior has potentially deleterious effects and the question of whether the behavior is contagious. Likewise, the question of whether moral disgust is an appropriate response to some act can be disambiguated into the question of whether that act is really immoral, and whether the act is potentially contaminating.

Conclusion: Disgust, Affect, and Evaluative Skepticism

I have argued that claims to the effect that disgust 'distorts' moral judgment or is somehow an illegitimate component of moral judgment are problematic insofar as they are too coarse-grained to distinguish disgust from other affective and cognitive features of moral judgment; they risk throwing the baby out with the bathwater. The considerations adduced above highlight both the need for and the difficulty of specifying exactly when and how supposed 'distorting' factors, biases, or framing effects should

make us suspicious of the resultant judgment. The issue is especially pressing for evaluative judgments (as opposed to, say, judgments about probability) as we have no agreed-upon, independent procedure or standard against which to measure our judgments. But it is not specific or unique to sentimentalist theories of evaluative judgment, as revealed by the data suggesting the existence of (potentially distorting) metacognitive influences on such judgments. Furthermore, many judgments cannot be presented independently of some kind of frame—this paper, for example, had to be written in *some* font or another. The case of disgust serves as a useful reminder that if we are to avoid widespread skepticism about evaluative judgments, we must take a more fine-grained approach to specifying when and under what circumstances an emotion or other cognitive process is deserving of suspicion. Such accounts must carefully examine and engage with the specific content or presentation associated with the emotions and cognitive processes involved in evaluative judgment. A global assessment of the reliability of affect, emotion, or cognition is desirable from a theoretical standpoint, but problematic from an empirical one.

CHAPTER TWO:
HOW MORAL DISAGREEMENT IS A PROBLEM FOR REALISM

“I think that if one were to get all the people in the world to gather together the things they found unacceptable, and then to take from this pile the things they found acceptable, not a single custom would remain, but in the end they would all have been distributed among the peoples of the world.” -Anonymous, Dissoi Logoi

“...If we had been ignorant, say, of the custom amongst the Egyptians of marrying sisters, we should have asserted wrongly that it was universally agreed that men ought not to marry sisters, -even so, in regard to those practices where we notice no discrepancy, it is not proper for us to affirm that there is no disagreement about them, since, as I said, disagreement about them may possibly exist among some of the nations which are unknown to us.” -Sextus Empiricus, Outlines of Pyrrhonism

Introduction

Philosophers often speak of *the* argument from disagreement, as if it were a single thing, when in fact there are many arguments from disagreement that take disagreement to be evidence for moral antirealism or evidence against moral realism. In this paper I discuss four arguments from disagreement and assess their strength. Some philosophers have suggested that actual moral disagreement is irrelevant to the arguments from disagreement; the existence of *a priori* versions of the argument that appeal to possible disagreement, the thought goes, renders appeals to the empirical record otiose. Part one is devoted to showing this claim is false: I discuss two versions of the *a priori* argument from disagreement and show that they can- and should- garner support from an empirical inquiry into instances of moral disagreement. I go on to show that, though the empirical

evidence supports the *a priori* arguments from disagreement, *a posteriori* arguments proceed more smoothly and are able to sidestep debates over the semantics of moral discourse. Using the aforementioned empirical evidence and the realist's responses to the *a priori* arguments, I argue that the *a posteriori* argument can be reformulated as a kind of dilemma for the realist: they are left open to either an inference to the best explanation argument from disagreement or a kind of skeptical argument from disagreement. I conclude by considering some objections.

Before discussing the arguments, a word about terminology is in order. For the purposes of this paper, I refer to the proponent of the argument from disagreement as a moral antirealist. What I mean by this is just that the proponent of the argument denies moral realism; antirealism picks out a variety of views, but for present purposes, it is enough to distinguish realist from nonrealist views. The term 'realist' has also been applied to a wide variety of views, so it will be useful to say something about how I use it in what follows. I follow Michael Smith in taking 'realism' to refer to the view that "moral questions have correct answers... made correct by objective moral facts... and that by moralizing we can discover what these objective moral facts... are." (1991: 399) Smith goes on to offer an analysis of objectivity: "the term... signifies the possibility of a convergence in views" (1991: 400). This is an instance of what Horgan and Timmons (2008: 270) call the "rationalist" conception of objectivity, which is the conception I will be using throughout the paper. On the rationalist conception, a domain is objective "if there is a method of thinking or reasoning whose use would yield... convergence in belief about the subject matter in question." With this terminology in place, I'll begin by explicating the *a priori* arguments from disagreement.

1. *A Priori* Arguments from Disagreement

Since realism is committed to claims about moral facts and about moral knowledge, the antirealist has two targets at which to aim the argument from disagreement. The metaphysical argument takes aim at moral facts; the epistemic argument takes aim at moral knowledge. I discuss each in turn before arguing that, despite being *a priori*, the arguments are best evaluated via an empirical investigation into actual moral disagreement.

A priori metaphysical:

1. If two individuals disagree over a claim about an objective matter of fact, then one of them must be mistaken.
2. It is possible for two individuals to disagree over moral claims and for neither of them to be mistaken.
3. Therefore, moral claims do not concern objective matters of fact.

A priori epistemic:

1. If two individuals disagree over the truth of some claim, and they are epistemic peers with respect to that claim— i.e. neither has any reason for thinking his opponent more likely to be mistaken than he is— then they ought to suspend judgment about that claim; their belief in the truth or falsity of the claim is unjustified.
2. The disagreement need not be actual; merely possible disagreement suffices to undermine justification in the way described above.
3. It is possible (for all or for some subset of moral beliefs) for two epistemic peers to disagree over the truth of a moral claim.
4. Therefore, we should suspend judgment about (all or some subset of) our moral beliefs; they are unjustified.

The metaphysical version of the argument targets the realist's claim that there are objective moral facts or properties; the epistemic version targets the idea that our moral beliefs or judgments can be justified. The metaphysical argument takes the analysis of objectivity above as its starting point: if it's possible for two people to have conflicting

beliefs about the truth of some moral claim, and for both to be correct/neither to be mistaken, then there are no objective moral facts. It is possible, and so there are no objective moral facts. Obviously, a lot hangs on the second premise here. The first premise follows straightforwardly from the analysis of objective facts above. In experimental work, denial of the possibility of faultless moral disagreement is often taken as diagnostic of moral realism (see, e.g., Goodwin and Darley 2007, Sarkissian et al in press). Realists and antirealists alike should be inclined to accept the first premise of the metaphysical argument.

The epistemic argument is more controversial; many epistemologists disagree with premise one, maintaining that even in the face of disagreement with an equally rational, informed peer, one is licensed to maintain one's level of confidence in the belief at issue. Premise one, though, claims that if one holds a belief that is denied "by another person of whom it is true that: you have no more reason to think that he or she is in error than you are" (McGrath 2008: 91) then that belief does not amount to knowledge; there is no reason why the belief must be actually denied by another person. The mere possibility of such disagreement suffices to undermine claims to knowledge (cf. Kelly, 2008; Wedgwood, 2010). Given that disagreement is possible for a great many— if not all— of our moral claims, we lack knowledge of, and should suspend judgment about, a significant number of moral claims/answers to moral questions. But for reasons I discuss below, in the case of the epistemic argument, it is difficult to evaluate premise one without a discussion of premise two.

What about the second premise? This is the real sticking point. That's in part because the first premise is so ecumenical: if realists and antirealists both agree on what

objectivity requires, then their disagreement is over whether those requirements are met. Premise two is, essentially, the antirealist's thesis; it amounts to a denial of moral realism. Therefore asking a realist to concede it is unreasonable, and pointless: getting to antirealism via an argument that assumes the denial of realism accomplishes nothing. David Enoch (2009) has suggested that this argument cannot go through without begging the question against moral realism, and therefore can be summarily dismissed. After all, he suggests, why can't the realist simply respond by denying that such cases are possible and maintaining that the very fact that two parties disagree is evidence that one of them must be making a mistake? If two parties disagree, for example, about the truth of some unproven mathematical conjecture, we feel confident that one of them is making a mistake, even if we have don't know which one of them is mistaken, or what the mistake is.¹⁷ Why not make an analogous response in the case of moral disagreement? Well, it's certainly one way to avoid begging the question against realism. But it seems to do so by begging the question against antirealism.

This is one reason why the arguments about how best to respond to disagreement between epistemic peers in other discourses are not straightforwardly generalized to include moral disagreements. In disagreements over arithmetic, it's known that one party is making a mistake, and the issue is which party that is and whether one is entitled to assume that it must be the other party, or whether one ought to concede that the mistake might be one's own. But even the staunchest realist ought to admit that moral realism is less securely established than its mathematical counterpart. (The question of whether and

¹⁷ I return to this point below when I discuss objections to the argument from disagreement, since the analogy is sometimes taken to show that arguments from disagreement would, if successful, also undermine realism in science and mathematics.

to what extent realism enjoys a presumption in its favor in cases like these is one I take up in Chapter 3.) In disagreements over moral claims, it's certainly possible that one party must be mistaken, in which case it may be permissible to retain one's confidence that the mistake belongs to one's opponent or it may be that one's confidence should be shaken. But it's also possible—that is, it should not be ruled out—that both parties are correct, in which case a third possibility is introduced: one is entitled to retain one's original view, but not to conclude that one's opponent is mistaken.¹⁸ This latter possibility is what the second premise aims to establish. To dismiss such a possibility at the outset does mitigate the force of the argument against moral realism—but only by assuming the truth of moral realism.¹⁹ I suggest that what Enoch's objection actually demonstrates is the difficulty of establishing premise two a priori—the realist denies such cases are possible, because his theoretical commitments rule it out. The antirealist maintains it is, because he has no such commitments. Therefore, I suggest that premise two is best established *a posteriori*. That is, the best evidence that faultless moral disagreement is *possible* is that it actually exists. In order to avoid begging the question, I'll refer to cases of disagreement that aren't explained by any ignorance, irrationality, or epistemic defect on the part of one or both of the parties as *fundamental disagreements*.

¹⁸This suggests an intriguing possibility regarding the puzzling phenomenon of belief or attitude polarization (see Kelly 2008). Suppose that in such cases, when exposed to good evidence for an opposing viewpoint, subjects actually conclude that there is no uniquely correct answer to the question under discussion. Then increasing one's credence in one's original view is less irrational than it appears, since countervailing evidence is actually evidence that one's view is not and need not be challenged by apparently conflicting positions or evidence.

¹⁹ There is an interesting issue here about what, in metaethical debates, constitutes a neutral starting point or standpoint from which to debate, but I don't tackle that issue here.

The suggestion I am making is that the best evidence for the possibility of faultless disagreement is an actual instance of fundamental disagreement.

The natural strategy for the proponent of either of these arguments is to find such cases; the natural strategy for the realist confronted with such cases is to deny that they are instances of fundamental moral disagreement, either by showing that one or both of the parties actually is at fault, or by showing that the disagreement is really about some other, non-moral issue. If the disagreement is really about, say, whether capital punishment has a deterrent effect, then it turns on some matter of fact that, though perhaps obscure, is nonetheless discoverable in principle. Therefore the disagreement is not in fact fundamental—it is caused by ignorance on the part of one or both of the disputants. Possible explanations of disagreement abound, and realists have identified several. What they have not done is examine whether these actually explain specific instances of moral disagreement. In what follows I discuss the realist's proposed explanations—call these *defusing* explanations (a term introduced in Doris & Plakias 2008)—and see whether they seem like promising explanations of instances of actual moral disagreement.

2. Explaining versus Explaining Away

Moral realists have suggested that the moral disagreement we observe is no threat to realism, because it can be attributed to one or more of the following defusing explanations. I'll explain each in turn; in the first two cases, I concede the explanations are indeed defusing, but I will go on to question whether they actually *explain* specific instances of disagreement. In the latter two cases, I question the assumption that the

proposed explanation would be *defusing*; that is, I deny that the explanation represents an epistemic shortcoming.

2.1 Non-moral Facts

Nearly every realist writing on disagreement has pointed to disagreement over relevant non-moral facts as an explanation: Boyd thinks, and Brink writes that “many genuine moral disagreements depend on disagreements over non-moral facts” (1989: 199), and Boyd goes further, arguing that “agreement on non-moral issues would eliminate *almost all* disagreements about the sorts of moral issues which arise in ordinary moral practice” (1988: 213; emphasis in the original).

2.2 Irrationality or Insensitivity to Evidence

Another possible explanation of moral disagreement is that it’s caused by faulty reasoning, or by a failure to take into account or properly weigh the relevant evidence. Some writers have claimed that the failure of imagination is a particularly pernicious problem in moral discourse—many disagreements are thereby supposed to be caused by a failure “to sympathize and imagine what it is like to occupy a different position in the relevant interaction,” (Enoch 2009: 25) or to imagine other people’s feelings. Others (e.g. Shafer Landau 1994: 331) have suggested that emotions may “stand as a barrier to convergence” by interfering with our exercise of rationality. But, as Brandt points out, emotion itself is not sufficient evidence of irrationality: for that, we need a “reason to believe that [the agent’s] desireless opinion is more correct than the one we regard as the creation of his desire” (1944: 489). We need grounds beyond the dissenting judgment

itself for thinking that the disputant is irrational. Furthermore, this explanation is less plausible as an explanation of inter-*cultural* disagreement: it's one thing to claim a person is being irrational, but to accuse an entire culture of being irrational or unduly influenced by emotion is another matter entirely, a much stronger claim.

2.3 Partiality

Yet other realists have appealed to the fact that we so often have a personal stake in the outcome of the dispute. Enoch thinks this kind of explanation is “especially important” because “it is extremely powerful” and it can “explain the difference in the scope of disagreement in morality and in other discourses” (2009: 26), since the parties to a moral disagreement are more likely to have some interest in the outcome than the parties to a mathematical disagreement. Boyd points out that morality is subject to a high degree of “social distortion,” and people’s judgments are often influenced by “class interests,” (1988: 212) two facts that he thinks explain much of the moral disagreement we observe. Proponents of this explanation do not address the question of how exactly these factors *distort*; presumably, they lead people to cling to views even though, in the absence of a personal stake, they’d concede the argument. And they do so by rendering the person blind to some piece of evidence, some intuition, some error in reasoning that they would be able to appreciate were they not personally invested in the outcome.

I’m not as taken with this explanation as some of its proponents, not just because I’m skeptical of its explanatory power (though I am) but because demonstrating that self-interest is implicated in a belief’s genesis doesn’t itself undermine the belief. To do that, we’d want a more detailed explanation of how self-interest inclines us to error. In many

cases, self-interest will make us *more* careful reasoners—if we have a stake in the outcome, we want to make sure to get it right. If I am choosing between medical treatments for myself or a loved one, for example, I will presumably be motivated to learn as much as possible about my options. So the idea that partiality on its own explains a moral disagreement is neither here nor there in terms of assigning fault to one of the disputants. If this is to be an explanation, we need evidence that partiality makes people irrational in a manner that is likely to affect their moral judgment and reasoning.

2.4 Background Theory

The disputants may come to the table not just with different factual beliefs, but with different background theories, and this is especially likely to be true in cases of cross-cultural disagreement. If the disagreement can be explained in terms of these theories, then perhaps they aren't really moral disagreements after all. Daniels (1979: 274) calls the state in which one's background theories, principles, and intuitions about cases all align 'wide reflective equilibrium,' and suggests that "One traditional worry, that moral judgments are not objective because there is insufficient agreement about them, may be laid to rest by seeking wide reflective equilibrium," in part because "It may allow us to reduce moral disagreements (about principles or judgments) to more resolvable disagreements in the relevant background... theories."

Daniels is right that background theory is likely to be implicated in some/many moral disagreements, but overly optimistic in thinking that tracing moral disagreement to this level will make it more tractable, or lead to a quicker resolution. That's partly because disagreements over things like the correct theory of the person, or "the role of

morality in society” are still normative, and we are likely to find that our disagreements go ‘all the way down,’ as it were—for example, there may be no factual discovery that settles whether to think of a person primarily in terms of their relationships and responsibilities or their individual attributes and behavior. This distinction corresponds roughly to what Nisbett (2003, see especially Ch. 3) describes as the difference between the Western, ‘individualist’ conception of the person and the Eastern ‘interdependent’ conception. This is plausibly construed as a debate over background theory, and it likely exerts some influence over people’s judgments of obligation and permission, but it is a normative debate nonetheless. Charles Larmore (1987: 57-58) puts the dilemma well:

If we imagine that under ideal conditions others continue to hold their own view of the world, and that their view is significantly different from our own (imagine them to be Bororo, or Tutenkhamen and Li Po), we cannot expect that they could come to agree with us about the justification of some substantial claims of ours. And if... we imagine the supposedly ideal conditions as detached from our general view of the world as well as from theirs, we have no good notion of what would take place, if anything, and it is certainly unclear what sense there would be to saying that it is with the Bororo that we would be conversing.

So while this is a plausible explanation of some disagreement, it’s not a *defusing* explanation; it fails to show that the disagreement is the result of an epistemic failure or defect on the part of one of the disputants. Locating a moral disagreement in a disagreement about background theory doesn’t show that it’s not fundamental.

The fact that moral disagreements *might* be due to these defusing explanations doesn’t show that they are. The debate over the existence of fundamental moral disagreement has too often been conducted in terms of hypotheticals. Instead, I will now survey some cases of actual moral disagreement, and see how well they are accounted for by defusing explanations. Spoiler alert: not well.

3. Disagreements Defused?

3.1 Honor

Nisbett and Cohen (1996; see also Sommers, 2009 for a discussion of honor cultures) coined the term ‘Honor Culture’ to describe cultures with the following features:

- There is little cooperation among strangers, rather, cooperation usually occurs between members of tight-knit groups or among kin;
- Protection of resources is critical for survival, and resources are relatively scarce;
- Attempts at theft or raids on property are common;
- The society is relatively lawless.

The above characteristics are commonly found in herding societies, where one’s livestock is one’s livelihood, but where property (i.e. one’s herd) is relatively vulnerable to theft. Honor cultures are prominent in Mediterranean Europe, the Middle East, and the American South, but also to gangs and tribal societies. According to Nisbett and Cohen (1996), in such cultures one’s reputation plays an important role in securing one’s resources and deterring attempts at theft. Thus, in honor cultures, having a reputation as someone who violently responds to transgressions or attempted transgressions will be valuable, since it will deter others from attempting to take one’s property. This leads to the prediction that members of honor cultures will be more likely to respond severely and violently to offenses and to insults that might threaten one’s reputation. And in several studies, this is just what Nisbett and Cohen found.

Nisbett and Cohen predicted that, since the Southern United States was originally settled by cattle herders, honor culture norms would still prevail in the American South,

despite the fact that most Southerners no longer herd livestock. Thus they hypothesized that Southerners would be more likely to respond violently to insults and would also be more tolerant of others' violent responses to insult. To test this hypothesis, they performed both laboratory and field studies. In the first case, subjects hailing from both the Northern and Southern U.S. were brought into a laboratory (supposedly to perform an unrelated task) at which point their levels of cortisol (a hormone associated with stress) were tested. Subjects were subsequently instructed to go to a different room. While walking down the hallway, an experimental confederate bumped into the subjects and muttered 'asshole'. Cortisol levels were then re-tested. Nisbett and Cohen discovered that Southern subjects showed greater levels of cortisol increase than their Northern counterparts. In other words, Southerners had a stronger physiological response to insult than Northerners.

In the field study, Nisbett and Cohen composed a letter purporting to be from a man who had recently been released from prison after serving a sentence for manslaughter and was seeking employment. In the letter, the man explained that he had been in a bar when another man insulted his fiancée, to which he responded by suggesting that they step outside, at which point things escalated. The letter was sent to potential employers in both the Northern and Southern U.S., and the differences in response were striking. Southern employers tended to be far more sympathetic, with one responding: "anyone could have been in the situation you are in". No Northern responses were comparably sympathetic. The case illustrates Sommers' (2009: 37) observation that "honor cultures and non-honor cultures have radically different perspectives regarding responsibility and punishment." But does this represent a fundamental disagreement? In

this case, it doesn't look as though there are any differences in subjects' understanding of the relevant non-moral facts. Nor are the subjects obviously irrational in any way. And in the field study at least, the potential employers had ample time to think about the case before responding to the letter. So there's no reason to think that their judgments were unduly rushed or caused by a lack of deliberation.

3.2 Sex and Sexuality

The question of with whom one is permitted to have sexual relations, and under what circumstances, is highly culturally variable. Contemporary Western societies treat one's sexual orientation as relatively fixed throughout one's life. Homosexuality has, until recently, been viewed with moral disapprobation, and sex with children and family members is strictly proscribed. In contrast, the Ancient Greeks "regarded homosexual desire by a man or youth for a boy, or by a man for a youth, as natural." Sexuality was fluid, and identifying oneself as exclusively homo- or heterosexual was uncommon enough that, Dover reports, "An Athenian who said, 'I am in love' would not have taken it amiss if asked, 'With a boy or a woman?'" (1994: 213) The Greeks "would certainly not have regarded [an adolescent boy's] homosexual activity in adolescence as incompatible with the enjoyment of women or with his eventual prospect of a harmonious marriage."

While the practice of sexual relations between boys and adult men may seem completely foreign to a contemporary American, the norms governing the practice are not: "Public attitudes to the lover and his boy in fourth-century Athens seem to have been remarkably similar to modern attitudes to pre-marital sexual relationships." (Dover 1994:

215) The Greeks did not have markedly different factual beliefs about children or youths, and they were cognizant of, and had norms in place to protect against, the potential for physical harm caused by sexual relations between men and boys. Consent was emphasized— boys were actively courted, but were free to assent to or reject the advances of an older suitor, and there is evidence that sex with boys deemed too young to grant consent was condemned. Anal penetration was rare, and the preferred form of sex was intercrural (where the man’s penis is rubbed between the boy’s thighs), which may indicate concern for the well-being of the youth, and therefore signify that the Greeks were aware of the physiological facts surrounding sexual intercourse and its potential physical effects on a young boy’s body (for a discussion of the norms governing the Greeks’ practice of pederasty, see Nussbaum 1994). It is likewise implausible to write off the Athens of Plato and Aristotle as an irrational culture, given that much of what we know about philosophical inquiry, we learned from the Greeks.

The proscription of incest is one of the strongest candidates for a universal moral norm, in part because it is thought to have a biological basis: Evolutionary psychologists argue that humans have an innate aversion to sex with family members.²⁰ This ‘incest-avoidance mechanism’ (first proposed by Westermarck 1891) is thought to protect against the potentially deleterious effects of inbreeding/lack of genetic diversity. Different cultures differ on where, exactly, the boundaries of family versus potential mates are to be drawn: in some cultures, cousin marriage is not just permitted but encouraged, while in the contemporary United States, it is forbidden. But evolutionary

²⁰ More recently, Shepher (1972) has refined the theory, arguing that the mechanism operates via imprinting, resulting in sexual avoidance of individuals who were cosocialized during a certain developmental window.

psychologists and anthropologists have tended to assume that all cultures proscribe sexual relations between siblings and between parents and their children. However, it is now established that the Egyptians practiced what is known as “Royal Incest,” wherein siblings of royal families married one another. More recently, some scholars have uncovered evidence suggesting that the practice extended beyond royal families; the classicist Walter Scheidel (1996: 322), “about one-third of all young men with marriageable sisters married them instead of women from outside the nuclear family.” Zoroastrian religious texts make frequent and approving reference to “next-of-kin marriage”; Scheidel quotes from one such text that appears to extol the virtues of parent-child procreation: “blessed is he who has a child of his child...” (1996: 326).²¹ Goody (1956: 292) reports that the Tallensi of northern Ghana lack a word for incest, and that “offences [i.e. sexual relations] between brother and sister are merely ‘disreputable’,” but that the Ashanti, who live in central Ghana, treat sibling incest as an offense punishable by death.

While it is plausible to claim that the permissibility of royal incest derives from beliefs about the special status of royals, who were thought to be more than human, and therefore not subject to the same standards as others, the prevalence of the practice among non-royals is not explained by these factual beliefs. Nor, again, does irrationality seem like a plausible explanation: why should we suppose the Ashanti are more rational than their neighbors to the north?

²¹ Sextus Empiricus reports, in the *Outlines of Pyrrhonism*, that “the Persians... marry their mothers; and the Egyptians take their sisters in marriage.” (1933: III.203) See also Frandsen (2009).

3.3 Fairness

In an ambitious and wide-ranging field study, Henrich et al (2005) performed ultimatum games in 15 small-scale societies. Their results are especially interesting, because while earlier results of studies performed using university students as subjects revealed little cross-cultural variation in which offers were considered acceptable, Henrich et al's experiments, which were performed in small-scale societies, revealed substantial variation between groups. Whereas among university students playing an ultimatum game, "UG offers are consistently 50%"; among subjects in small-scale societies, offers ranged from well below 50% to up to 72%. Likewise, rejection rates were also highly variable. This may be an instance of a disagreement about 'background theory,' and while there may be agreement on the importance of fairness as such, there is little substantive agreement across these different groups on what constitutes a fair offer. It's difficult to see what factual disagreement could be driving the differing offers in these games, which are after all fairly simple to explain and understand; there is likewise no reason for thinking that some of the groups studied were irrational in any relevant way. So I agree with Machery et al's (2005: 831) conclusion that "subjects distribute windfall gains differently because they hold different views about fairness, specifically about how to fairly distribute such windfall gains." And these differences may emerge fairly early in development: Rochat et al (2009) observed cross-cultural differences in performance on tasks involving fairness and distribution among children as young as five years old, with children in modern urban communities behaving more self-interestedly than children in small rural communities.

4. The *A Posteriori* Arguments

The realist's defusing explanations have failed to satisfactorily account for the above cases (or have turned out not to be defusing to begin with), so we are now in a position to return to the *a priori* arguments above, and to premise two. And we can now point to the existence of fundamental moral disagreement as evidence for the possibility of faultless moral disagreement. In other words, we've arrived at the basis for an abductive argument for premise two of both the epistemic and the metaphysical *a priori* arguments: the existence of fundamental moral disagreement is best explained by the hypothesis that these are faultless moral disagreements, and that therefore there are no moral facts (via the metaphysical argument) and/or we do not have moral knowledge (via the epistemic argument).

But the realist has another response up his sleeve. The appeal to defusing explanations was intended to show that the putatively faultless disagreements that the *a priori* arguments appeal to are not really *moral* disagreements (because they turn on an issue of non-moral fact, or because they're really just failures of rationality). Another strategy the realist can take in response is to argue that they're not really moral *disagreements*. That is, the realist can refuse to grant the possibility of faultless disagreement by denying that faultless moral disagreement is conceptually coherent and insisting that if such a case is found, it's not really a *disagreement*. Disagreement, on this account, requires that one party be asserting a proposition and one party denying that same proposition (or asserting its negation). And in such cases, mutual correctness is just not possible. Hence on any plausible definition of disagreement, faultless disagreement is an incoherent notion.

The question of how best to define disagreement is vexed, and one I won't settle here. The antirealist is within her rights to resist this objection, and maintain that the realist's response begs the question against the argument from disagreement, by ruling out the possibility of faultless disagreement as a matter of definition. And she could point to the fact that many of the moral disagreements in question bear all the hallmarks of uncontroversially 'genuine' disagreements (such as, for example, disagreements about the correct way to divide a dinner check). But many antirealists themselves are sympathetic to the aforementioned definition of (or constraint on a definition of) disagreement. I do not want to enter into the debate over the correct account of what it is to be a disagreement here. Nor do I need to, since the argument can be reformulated so as not to require or even mention 'disagreement'. The two arguments I discuss below—the *a posteriori* arguments from disagreement—do not turn on the possibility of faultless disagreement, but rather on the nature of actual moral disagreement. And they do not turn on our definition of disagreement—in fact, while I will continue to use the word 'disagreement' for continuity's sake, the arguments I give below would lose none of their force if 'disagreement' were replaced with the more neutral term 'diversity'. The move from *a priori* to *a posteriori* arguments from disagreement allows the antirealist to sidestep the question of how to define disagreement. I now discuss two *a posteriori* arguments from disagreement, one epistemic and one metaphysical. I'll argue that responding to these arguments creates a dilemma, in that the realist's response to the metaphysical argument leaves him vulnerable to the epistemic argument.

Specifically, I'm suggesting that the antirealist draw on the empirical evidence described above to bypass the *a priori* argument entirely and make an inference to the best explanation argument:

1. The anti-realist can explain moral disagreement without appealing to objective moral facts or properties, by showing that our different moral views reflect different "ways of life"²² rather than (flawed, inaccurate, or distorted) perceptions of objective facts or properties.
2. The realist explanation must not only appeal to moral facts or properties but also explain how we fail to recognize or be guided by them; he must explain moral disagreement as the result of error.²³
3. The anti-realist explains disagreement better than the realist, therefore;
4. Most likely, there are no objective moral properties or facts.

Since this is an inference to the best explanation argument, its success hinges on whether its proponent's preferred explanation really does account for disagreement better than its realist competitor. We've already seen that the realist's proposed defusing explanations appear unable to account for the disagreements discussed above. Mackie thinks the best explanation of disagreement is that our values and moral judgments are really conventional, inculcated by society. So it's not surprising that we see different judgments in different cultures.

The realist might protest at this point that his proposed explanation has been given short shrift. Just because we have failed to identify the mistake, irrationality, or other defect responsible for the disagreement, he might argue, doesn't mean it's not there.

²² I borrow this phrase from Mackie (1977: 37-38)

²³ This is a bit too strong. The realist can appeal to indeterminacy and vagueness to explain *some* cases of moral disagreement—for example, it may be that some of the moral disagreements over the permissibility of abortion, or euthanasia, stem from vagueness or indeterminacy about personhood, or the fact that these are borderline cases of life. This explanation must be used sparingly, though, since a morality that allowed lots of vagueness and indeterminacy would be unsatisfying and would fail to deliver on its promise of epistemic access to moral facts, and since many of the disagreements discussed in this chapter are not plausibly ascribed to these factors (for example, the question of whether or not it is permissible to sleep with family members is not, one suspects, due to an indeterminate concept of siblinghood).

After all, this is difficult stuff: many realists have proposed a sort of fifth defusing explanation, which boils down to: morality is hard. With this consideration in mind, the realist may want to deny that he has failed to explain the disagreements; he can maintain that the defusing explanations are still the best explanation of the disagreement, even if we are not yet in a position to show where, exactly, the defusing explanations apply. But the problem with this strategy is that it makes morality *too* hard, and therefore exposes the realist to yet another version of the argument from disagreement—a skeptical challenge:

1. The process by which we form moral beliefs is unreliable (either because it is riddled with error, as the realist alleges, or because its aim and/or design is not to track truth or any real property of things, as the anti-realist alleges)
2. Beliefs formed as a result of unreliable processes are unjustified.
3. Our moral beliefs are unjustified.

This skeptical argument follows from the responses from the inference to the best explanation discussed above, though it aims not at the moral facts themselves, but at our ability to form justified beliefs or make justified claims about them. The IBE argument introduces two competing explanations of moral disagreement, and the skeptical (or epistemic) argument contends that on either explanation, our confidence in the processes by which we form moral beliefs should be undermined. If the antirealist's explanation of moral disagreement is correct, then our moral judgments are the product of convention and of our "participation in... ways of life." (Mackie 1977: 37) And this explanation renders moral facts explanatorily gratuitous, since it explains why we have the moral beliefs that we do without any mention of moral facts or truth. In doing so, it also introduces a story about moral belief formation that does not paint the process as one aimed at tracking truth. Rather, the process is one of socialization into, and conformity

to, ways of life and social convention. The argument here is not so different from skeptical arguments based on evolution (see, e.g., Kahane 2011, Joyce 2007, and Street 2006, among others), except that it takes culture, rather than evolution, to be the relevant explanans.

This is partly because an evolutionary explanation would seem to favor a cross-cultural uniformity of moral views. Guy Kahane (2011: 121) points out that the phenomenon of moral diversity precludes explanation in terms of “the diversity of evaluative beliefs over time and across and within cultures— a diversity not fully explained by differences in non-evaluative belief— makes the suggestion that all evaluative beliefs can be given a straightforward evolutionary explanation extremely implausible.” The psychologist Solomon Asch argues that this implication may be so implausible as to actually provide evidence against relativism. The relativist explanation of moral diversity in terms of culture presupposes, according to Asch, an implausibly plastic view of human psychology and motivation, one on which man is “a dynamically empty organism,” and on which “human nature is like water, which takes on whatever shape is imparted to it.” (Asch, quoted in Brandt 1944: 479) But cultural explanations of morality needn’t deny that our moral psychology is shaped in part by evolution, nor need the existence of diversity preclude an evolutionary story about the nature and development of the cognitive architecture underlying moral judgment. (Asch suggests that we view cultural differences as “the necessary consequence of permanent human tendencies coming to expression under particular conditions,” suggesting that he agrees with this assessment while disagreeing that it is compatible with relativism.) So the two arguments are not incompatible; we can give an explanation of moral norms and beliefs

in terms of culture, evolution, or both. The important thing is that whichever feature we invoke can explain our moral judgments without invoking moral facts. Richard Joyce succinctly explains the debunking strategy: “I contend that on no epistemological theory worth its salt should the justificatory status of a belief remain unaffected by the discovery of an empirically supported theory that provides a complete explanation of why we have that belief while nowhere presupposing its truth.” (Joyce 2006: 219)

On the other hand, if the realist’s explanation of disagreement is correct, then our moral reasoning and the resultant judgments are subject to distortion by irrationality, self-interest, and ignorance. Furthermore, if the realist is right, these influences operate so subtly that we’re often unaware of their influence, hence the persistence and intractability of moral disagreement—we stick to our guns because we fail to recognize that defusing explanations apply to us. This combination of pervasiveness and subtlety is a recipe for skepticism: our belief forming mechanism, in the case of morality, is often thrown off-track, and we are often unaware of this fact. This means both that the process by which we arrive at moral beliefs/judgments is unreliable, and that for any particular belief, there is a non-negligible chance that the belief is the result of error, ignorance, and/or irrationality. Therefore, if defusing explanations give the best account of moral disagreement, so much the worse for our confidence in our ability to form moral judgments.

5. Objections and Replies

5.1 Overgeneralization/*Companions in Guilt*

Any argument that derives an antirealist conclusion from the existence of disagreement faces a companions in guilt response: the objection that most if not all discourses have areas of persistent, apparently faultless disagreement, and yet we are not thereby inclined to abandon realism about them.²⁴ Earlier I mentioned mathematics, and there are many mathematical conjectures the truth of which remain contentious and which may never be proven true or false. Brink points to scientific disagreements that once seemed interminable: “no one concluded from the apparently interminable disagreement among astronomers a short while ago about the existence of black holes that there was no fact of the matter concerning the existence of black holes.” (1989: 198) The debate over whether God exists, for example, is an instance of disagreement in which both parties are rational, privy to all the relevant facts, and seem to be epistemically faultless. And it isn’t just that the disagreement is unresolved— it’s that it’s unresolvable. It is highly unlikely that evidence could be procured which would satisfy both sides that the issue is settled. Indeed, one of the reasons the debate remains so thorny is that there seems to be a lack of consensus on what form such evidence would have to take. Nonetheless, we remain confident that there *is* a fact of the matter about which side is correct in the debate over God’s existence. If disagreement doesn’t undermine realism about this question, why think it undermines realism about moral discourse?

²⁴ Loeb (2003) points out that this strategy can also be coopted by the antirealist, who can point to the similarities between moral disagreements and disagreements in areas where realism is not tempting, such as taste.

In responding to this objection, it's important first to distinguish between different levels of disagreement. The proponent of the argument from moral disagreement is basing his claims on disagreement about the correct answers to first-order moral questions, such as whether euthanasia is morally permissible or whether incest is morally wrong. He is *not* basing his claim on the existence of higher-order disagreement over whether there is a correct resolution to these first-order debates (though this is another sort of objection that can be leveled against the argument from disagreement, one which I discuss below). So the analogy with the debate over God's existence is misguided: in neither case does the anti-realist want to deny that the ontological question has a uniquely correct answer. The more apt analogy would be with debates over what attributes God has, and these do exist and have been taken as evidence that God does not exist. Likewise, fundamental disagreements over what properties different actions have (i.e. rightness, wrongness, permissibility) are, or so I've argued, good evidence that those properties don't exist.²⁵

But the realist can point to other debates as well, in areas such as mathematics, medicine, history, and science that aren't about ontology, and reiterate the objection. And it's here that the antecedent plausibility of realism becomes relevant. We approach debates in math, medicine, and science with a presumption in favor of realism;²⁶ to that

²⁵ For a discussion of religious disagreement and its significance, see Feldman (2007), who gives a thorough explication of the issues, but draws slightly different conclusions from the ones I've drawn here.

²⁶ The question of whether realism and anti-realism are best understood the same way across all these areas is potentially tendentious; in referring to mathematical realism, I certainly don't mean to commit to a Platonist ontology or anything like that. By 'realism' I mean the view that there are objective facts about these discourses; what undergirds those facts- the metaphysical nature of their 'truthmakers'- is an issue on which I remain agnostic, though it is an interesting and worthy question to what extent and why realism can or should be consistently defined across different areas. I discuss this briefly in Chapter 3.

extent, we suppose even apparently intractable debates in these areas will (or at least can) be resolved. To the extent that they cannot, we should suspend realism about those questions. (See Plakias and Doris 2008) This is why the realist needs to motivate the idea that we approach the issue with a presumption in favor of realism in order for the objection to succeed. Many of the considerations that motivate realist treatments of the other discourses under consideration are absent in the case of morality: moral facts don't play the same explanatory role as scientific facts (Harman 1988), nor has there been the kind of progress in morality that has characterized math, science, and medicine (see Williams 1985, especially chapter 8; also Leiter 2009). Realists have sometimes appealed to commonsense moral thought and practice to motivate the claim, arguing that we do in fact appeal to objective moral facts in our explanations, judgments, and justifications; this, combined with a preference for conservative/preservationist over revisionary theories, would create a presumption in favor of realism, and rehabilitate the analogy at issue. The trouble is— as I argue at length in Chapter 3— the claim just isn't true. While we do talk about morality an awful lot, recent empirical work suggests that what we talk about when we talk about morality isn't always (or even usually) objective (see Goodwin and Darley 2008, Nichols 2004, and Sarkissian et al in press). So this appeal to folk morality also falls flat as a way of motivating a presumption in favor of moral realism, without which the companions-in-guilt strategy fails as a response to the arguments from disagreement.

5.2 *The Argument is Self-Defeating*

A second, related objection to the argument from disagreement is that it's self-defeating. This is like the overgeneralization objection, except here the complaint is more specific: the argument overgeneralizes to include itself in its scope, and therefore undermines its own claim. If the existence of fundamental disagreement is a reason for skepticism or antirealism, then given the degree of disagreement over the argument from disagreement, we should be skeptics or antirealists about it. In fact, the problem extends beyond the argument itself, and may encompass most if not all of philosophy. Philosophical disagreements are especially troubling, because they have persisted for so long among reflective, informed, rational disputants. Leiter (2009: 8) refers to “what seems to me the single most important and embarrassing fact about the history of moral theorizing by philosophers over the last two millennia: namely, that no rational consensus has been secured on *any* substantive, foundational proposition about morality,” but the observation extends beyond moral theory: a lack of consensus characterizes many if not most areas of philosophy.²⁷ Enoch (2009: 47) notes, “disagreement is widespread not just *in* morality, but also *about* morality, in metaethics, and indeed in philosophy in general.” For this reason, he thinks it will be difficult if not impossible to generate an argument from disagreement to the denial of realism that doesn't defeat itself, either by defeating realism about metaethical theorizing— a kind of meta-metaethical realism— or by defeating realism about philosophical debates altogether.

One response might be to embrace a kind of Pyrrhonian skepticism, either about moral theorizing or about philosophical theorizing more generally (for a discussion of

²⁷ See also Shafer-Landau (2003: 219-220).

Pyrrhonism about metaethics, see Sinnott-Armstrong 2009). However, there are other responses available that require us to do less bullet-biting. First, there is the question of how closely philosophical disagreements really do resemble fundamental moral disagreements. A striking dissimilarity is that philosophers typically have a lot invested in the views they defend, and in their continued participation in philosophical debates: it is, after all, their job. So defusing explanations may be more plausibly applied to disagreements between philosophers than to those between laypeople. This is, of course, an issue that merits consideration on a case-by-case basis, which is beyond the scope of this paper. But experimental philosophers have begun investigating whether there are correlations between certain personality traits and philosophical views; Nichols (2007) has argued that the rise of compatibilism can be explained in terms of the fact that compatibilism is motivationally attractive: to put it bluntly, the persistent popularity of compatibilism is an instance of wishful thinking. If this line of argument is correct, then *fundamental* philosophical disagreement may be less pervasive than it appears to be.

At this point, the concern might shift from realism about philosophy to skepticism about philosophical arguments and reasoning, but this is a somewhat different claim than the claim that the argument from disagreement leads to antirealism about philosophy itself. Furthermore, the question will have to be decided on a case-by-case basis, by looking at specific philosophical disagreements: it may well be that defusing explanations are more plausibly applied to (for example) debates in free will than to debates in ontology. This is an empirical question and should be dealt with as such. But notice that at this point, the argument no longer threatens philosophy as a whole, but rather different subfields or even particular debates within philosophy, which makes this a slightly more

palatable bullet to bite. And skepticism about metaethics, even if it does extend to the argument from disagreement, might be a satisfactory enough conclusion for the irrealist/proponent of the argument from disagreement; if moral disagreement forces the suspension of judgment regarding competing metaethical theories, this might not be a victory for antirealism, but it is certainly a strike against moral realism. Pyrrhonian skepticism is, in this case at least, not a Pyrrhic victory.

Conclusion

The observation that arguments from moral disagreement benefit from empirical research is not new; as the epigraphs to this paper reveal, the Greeks were acutely aware of the variation in moral codes and the possibility that claims about the universality of moral beliefs or practices are always vulnerable to new discoveries. Nor has the empirical record been supplanted by newer, *a priori* forms of the argument from disagreement— as I've argued above, the existence of moral disagreement continues to make trouble for the realist's assertions that moral facts exist and that we can claim to know them. Realists such as Smith and Jackson predict that continued inquiry and argument will yield increased convergence, but this is an empirical prediction, and it's worth noting that just as the significance of disagreement depends on its causes, so does the significance of convergence. Furthermore, discoveries about the nature of the cognitive processes underlying moral judgment may lead us to revise our conception of what constitutes a defusing explanation. Therefore, a metaethical theory that does not stand or fall with convergence is preferable to one that does. Finally, we should expect and be prepared for the fact that there are likely to be areas of convergence and areas of fundamental

disagreement; a theory that can accommodate both types of moral fact will therefore enjoy an advantage over a theory (such as Smith's) that rules out this possibility *a priori* by building convergence into its definition of moral facts.

CHAPTER THREE

MORAL RELATIVISM AND MONOLITHIC METAETHICS

Introduction

Contrary to what error theorists and many moral realists have claimed, moral discourse is not committed to full objectivity.²⁸ In fact, thinking of objectivity in terms of degrees can make sense of and reconcile apparently inconsistent features of moral discourse and judgment. In this paper I describe and discuss evidence showing that morality can be vindicated, and can be said to have a subject matter, even if it lacks full objectivity. I then discuss an objection to the effect that any morality that contains both objective and non-objective facts is inconsistent, incoherent or cannot be captured by a single metaethical analysis, a claim I dismiss by showing how moral relativism can account for the existence of both types of facts. In other words, relativism makes sense of why some moral judgments seem to be more objective than others. In the previous chapter I described evidence that this is in fact the case; in this chapter I focus on the

²⁸ Michael Smith holds a view like this, on which moral facts are those which ideal agents would all converge on, and on which morality is committed to objectivity. If there are no objective moral facts, then, for Smith, there are no moral facts. See, e.g., *The Moral Problem*, esp pages 9 & 13. Despite holding a very different metaethical view, error theorists like Mackie and Richard Joyce also endorse the conditional: see Joyce's *The Myth of Morality*, especially chapter 2, and Mackie's *Ethics: Inventing Right and Wrong*, especially pages 23 and 40.

claim that a metaethical theory that includes both objective and relative moral facts is theoretically viable and practically necessary.

1. The Objectivity Assumption

Horgan and Timmons (2008: 270) suggest that definitions of moral objectivity fall into one of two categories: ontological or rationalist. On the ontological conception of objectivity, moral properties and facts are “out there to be experienced” in a way analogous to “ordinary physical-object thought and discourse.” This means that the moral properties are *not* “certain idiosyncratic facts about the psychology of the agent making the moral judgment.” On rationalist conceptions of objectivity, “a realm of thought or discourse is objective... if there is a method of thinking or reasoning whose use would yield... convergence in belief about the subject matter in question.” These two conceptions of objectivity aren’t mutually exclusive; rationalist objectivity is sometimes seen as evidence for ontological objectivity.²⁹ But the rationalist conception is more modest insofar as it does not take a position on what kind of facts the convergence of belief is about. In this paper, I follow the rationalist conception of objectivity, and take objectivity to be a matter of convergence in belief among rational, informed agents. One argument in favor of the ontological conception is that it best captures our intuitive, pre-theoretical understanding of objectivity; an objection to the rationalist conception might be that it makes moral objectivity look significantly different from objectivity in other domains. But both points overlook the considerations that make

²⁹ For example, Williams (1985: 136) takes science to be a paradigmatically objective discourse, and points to the fact that in science, not only is there convergence, but “the best explanation of the convergence involves the idea that the answer represents how things are.”

moral objectivity distinctive. While it is natural- and, in the case of physical object talk, apt- to think of objectivity as a matter of facts being “out there” in the world, existing independently of us and any of our beliefs about them, this conception of objectivity is complicated when applied to morality by the fact that morality is, essentially, about persons and their relations to one another. So while we can say that physical objects existed and would continue to exist without any people to perceive them, the same cannot be said for morality. Thus, the fact that a conception of moral objectivity- specifically, the rationalist conception- is discontinuous with conceptions of objectivity in other domains is not a strike against it, but a point in its favor.

Moral relativism is often supposed to be a significantly revisionary view: “Moral absolutist philosophers often portray relativism as an exotic skeptical doctrine delivered by some special philosophical theory, and they see (and portray) themselves as defenders of common sense against the bizarre...” (Dreier, 2006: 241). For example, Michael Smith writes that common sense morality is committed to the idea that “moral questions have correct answers; [and] that the correct answers are made correct by objective moral facts” (Smith 1994: 9). This supposition is not unique to realists; relativists themselves have tended to take a defensive posture, conceding that “relative morality may be less than common sense can hope for” and that relativism is “revisionary of common sense... at odds with common sense moral thinking” (Dreier, 2006: 241). Darwall writes: “Ethical thought and feeling have ‘objective purport.’ From the inside, they apparently aspire to truth or correctness and presuppose that there is something of which they can be true or false... something objective and independent of the perceiver (e.g., some objective fact or an objective property of some substance)” (1998: 25, 239).

Call this idea—the idea that common-sense morality is committed to, or aspires to, objectivity-- *the objectivity assumption*. The objectivity assumption is typically paired with a second, less explicit, assumption: that metaethical theory is, to some degree or another, constrained by common sense moral theory. This constraint might be so strong that we are forced to reject any theory that counts as revisionary of common sense, or it might simply mean that that common sense morality is one factor among many we must weigh when constructing a theory. In what follows, I will assume a moderate reading of it, which says that all things being equal, we should prefer a theory that vindicates and accords with common sense moral intuition/theorizing/discourse over one that is revisionary. Call this moderate position *metaethical conservatism*.

The rise of experimental philosophy has led to suspicion about unsubstantiated claims about common sense belief or intuition. In this case, that suspicion turns out to be warranted. The objectivity assumption turns out to be dubious, at best.

But this same evidence that undermines the objectivity assumption raises a new puzzle. As I'll demonstrate shortly, it turns out that moral judgments are not all equal with respect to objectivity. The reason for doubting the objectivity assumption is that, in the experiments I will describe below, subjects- users of common sense morality- have evinced a willingness to treat a moral violation as both morally wrong and relatively true. Furthermore, the degree of objectivity or relativity they assign to moral violations varies depending on the nature of that particular violation. This is puzzling: are people being inconsistent or incoherent? Are they judging as moral realists at one moment and moral relativists the next? I argue that a particular kind of moral relativism makes sense of this puzzle, explaining how and why moral judgments can vary with respect to objectivity.

1.1 A Puzzle About Intuitions

In his 2004 paper, 'After Objectivity,' Shaun Nichols set out to determine whether undergraduates are moral objectivists or subjectivists³⁰. He asked undergraduates to imagine a case in which two individuals are arguing over the permissibility of hitting people for fun. Both insist that among all the people they know, such a behavior is considered okay. Nichols found that a significant percentage of respondents were inclined to agree with the statement "there is no fact of the matter about unqualified claims like 'it is ok to hit people just because you feel like it.'" Different cultures believe different things, and it is not absolutely true or false that it is ok to hit people just because you feel like it." Moreover, these respondents did not differ significantly from their moral objectivist counterparts when it came to their ability to distinguish moral violations (for example, hitting) from conventional violations (for example, talking out of turn, or chewing gum in a classroom where there is a rule against doing so). This study casts doubt on the objectivity assumption— moral judgments are being made even in the absence of a judgment of objectivity. It does not, however, address the question of whether these judgments of objectivity vary depending on the question at hand.

There are also problems with the design of the experiment itself. First, the sample size is very small—about forty undergraduates total (interestingly the sample was composed almost entirely of females—and the extent to which these intuitions vary by gender is an underexplored area of experimental philosophy at this point). And some people have raised concerns about the use of a subject pool composed entirely of

³⁰ This is Nichols' terminology, so I've followed him in using it to describe his experiment. However, he intends it to contrast with objectivity, so for present purposes, it works much the same way as the term 'relativity'.

undergraduates- are college students representative of the population as a whole? Second, the forced-choice design of the experiment may raise concerns—subjects were forced to either endorse the wrongness of hitting people or to deny that such wrongness was absolute. But it seems that subjects might want to say both that hitting people is wrong, *and* deny that this wrongness is absolute. And subjects may have read the wording of the third and final choice as ambiguous between denying that there is any fact of the matter at all about such claims, and maintaining that there are facts but denying that the facts are absolute. In fact, perhaps even more subjects would have chosen the non-objectivist option had it been presented in a more perspicuous way.

In a related series of experiments, the psychologists Geoffrey Goodwin and John Darley (2007) set out to see whether subjects would treat ethical statements as objective and how this might differ from the way they would treat statements of scientific fact, social convention and taste. The experimenters gave subjects a range of statements in all three areas and asked them to rate their agreement with each statement, as well as to indicate whether they thought it was true, false, or an opinion or attitude. In the second phase, subjects were told that another person strongly disagreed with them, and asked whether they thought one of the two parties must be mistaken, or whether it is possible that neither party is mistaken. Their results showed that overall, people tended to treat ethical statements as more objective than statements of convention or taste, but less objective than statements of fact. The more interesting result, however, was the degree of variation *within* the category of ethical statements: subjects tended to strongly agree with statements concerning the permissibility/wrongness of bank robbing or opening gunfire on a crowd, and were likely to judge these statements ‘true,’ but, while levels of

agreement with statements concerning the permissibility/wrongness of abortion, euthanasia and stem cell research were similar, subjects proved extremely reluctant to describe these as ‘true.’ For example, 68% of subjects said it was a true statement that it’s wrong to open fire on a crowd, but only 2% of subjects said it was a true statement that abortion is permissible. This indicates that whether subjects regarded a statement as expressing an “opinion or attitude”³¹ or an objective fact is heavily influenced not just by the domain (aesthetic, ethical, scientific) but also by the specific content of the statement.

Note that the two-phase methodology described above addresses the concern I raised about Nichols’ experiment—subjects were given the option to separately rate wrongness, whether they thought the statement described a fact or an opinion or attitude, and objectivity, thereby eliminating the ambiguity of Nichols’ results. In addition, while the initial paper uses an undergraduate subject pool, subsequent studies (Goodwin and Darley, ms) replicate the result among both faculty and nonacademic adult subjects, ameliorating concerns that the result is just an artifact of relying on an undergraduate population.

In another recent study, Hagop Sarkissian et al (in press) describe Horace, an agent who “finds his youngest child extremely unattractive and therefore kills him.” Subjects were then asked to consider two individuals judging Horace’s action, one of whom judges it morally wrong while the other judges it morally right. One of the judges is described as a classmate of the subject; the second judge is described either as a typical

³¹ For a discussion of why this experiment fails to distinguish between these two, and whether any experiment can distinguish attitude from opinion or belief, see the discussion in section 3.1 below.

American college student, a member of an Amazonian tribe with very different values, or as an extraterrestrial being with a very different psychology from human beings. Subjects are then asked to rate their agreement with the statement, “since [the two judges] have different judgments about this case, one of them must be wrong.” What the authors found was that subjects’ relativism, as measured by their disagreement with this statement, increased as the distance/differences between the two judges increased. One might argue that this is due to some kind of framing effect that subjects would repudiate were they made aware of it, but subjects who were given all three cases at once continued to respond as most relativist in the extra-terrestrial case, significantly less relativist in the Amazonian case, and as objectivists in the same-culture case. This indicates that subjects are aware that they are treating the three cases differently and that they do not feel pressure to revise their judgments so as to respond the same way in each of the three cases. It would seem that “People. . . reject the idea that a single absolute standard can be applied to all moral judgments of a given agent and to operate instead with a system that applies different standards to different judges” (Sarkissian et al, in press).

1.2 Culture, Content or Both?

The foregoing experiments suggest that the degree to which subjects are willing to say that a dispute has a single right answer is a factor of the similarity or dissimilarity between the disputants. Subjects treat disputes between individuals within the same culture as having an objectively correct resolution, but objectivity decreases as the distance between disputants increases. However, Sarkissian et al only tested extremely violent transgressions, involving harm to non-culpable bystanders. Goodwin and

Darley's results demonstrate that varying the type and nature of the transgression also affects judgments of objectivity. This suggests that cultural distance is not the only dimension along which judgments of objectivity vary. Rather, content also plays a role: in Goodwin and Darley's study, the transgressions most likely to be judged objectively wrong involved violent harm to innocent bystanders. It's also worth noting that euthanasia and abortion, the two actions judged to be least objective, are extremely controversial. This is suggestive: perhaps the folk are also operating with a rationalist conception of objectivity.

2. If the Objectivity Assumption is False, is any Metaethics Permitted?

The experiments discussed above raise doubts about the objectivity assumption and introduce a further observation: that some moral claims are treated as more objective than others. One objection to according these experimental results any significance is that they are plainly inconsistent or incoherent: that subjects treat some cases as objective and others as relative only shows that they lack a coherent concept of morality, and therefore their concepts should be disregarded. Furthermore, if it is true that subjects lack a unified, coherent concept of morality, but are instead deploying two distinct or different concepts, then any account we give of moral judgment that treats it as homogeneous will be significantly revisionary.³² So not only should we not try to account for these

³² This objection is similar to a line of argument put forth in MacIntyre, who writes that "the integral substance of morality has to a large degree been fragmented" (1981: 4). See also his discussion of the emergence of emotivism as a competitor to objective moral theory, pp 15-20. It is also echoed in Larmore's discussion of the heterogeneity of morality (1987: ch 6). However, the objection discussed here bears, not on the content of moral theories, but on their metaethical features.

judgments, we should view them as licensing accounts that depart significantly from common sense morality—because there is no coherent common sense morality.

Or perhaps we should view these judgments as a reason to give up on the project of vindicating moral discourse and instead adopt an error theory concerning morality. Don Loeb argues that variations in judgments about objectivity like the ones discussed above license a move to an error theory: “inconsistent elements—in particular, commitments both to and against objectivity—may be part of any accurate understanding of the central moral terms.” These conflicting commitments have metaethical consequences, since “if the moral vocabulary is best understood as semantically incoherent, the metaphysical implication is that with respect to that vocabulary there is nothing to be a realist *about*” (Loeb 2008: 358). Call this view metaethical *incoherentism*.

If, however, we take seriously the idea that metaethics should, insofar as possible, capture the tacit commitments of ordinary moral discourse, then we should attempt to account for prima facie inconsistent intuitions. This has led some to argue in favor of a more ecumenical view than incoherentism. *Metaethical variantism* claims that "(a) some uses of moral language definitely fit expressivism, (b) other uses definitely fit realism, and (c) neither kind of use is primary or aberrant" (Sinnott-Armstrong 2009: 239.) Note the starkness of this dichotomy: there's plenty of interesting theoretical territory between realism and expressivism, but Sinnott-Armstrong's definition treats relativist, response-dependence, constructivist, and subjectivist accounts as fly-over country. Yet, as I'll show in the next section, fly-over country is where we find the best of both worlds: an

account that reconciles and makes sense of our seemingly inconsistent metaethical commitments.

2.1 How to be a Moral Relativist

Apparently conflicting intuitions can be reconciled using a version of moral relativism sometimes called ‘self-locating’ relativism, first proposed and developed by Andy Egan (ms). This type of moral relativism draws on the idea that truth can be assessed, not just at worlds, but at *centered worlds*. Whereas analyzing the truth of a proposition relative to a world only requires us to know which possible world we are looking at, or rather which possible world is accurate, analyzing its truth relative to a *centered world* requires us to look at the triple (w, t, i), where ‘w’ picks out a world of evaluation, ‘t’ picks out a time, and ‘i’ picks out an individual.³³ This version of relativism is called *self-locating* because it doesn’t just tell us what world we’re in, but where we are in that world. So, for example, ‘Chicago is north of New Orleans’ is true in all the possible worlds where Chicago is north of New Orleans, but ‘New Orleans is nearby’ is true only when uttered by a speaker occupying a certain location within those worlds.³⁴ The moral agent is, when uttering moral claims, attempting not just to describe the world, but to locate himself within it by self-ascribing some property. For Egan, the property we are ascribing when we make moral claims is “being someone whose ideally rational self would desire that they ϕ .” This is an amended version of Smith’s (1994)

³³ This idea goes back to Lewis (1979). As far as I know, however, Egan is the first to apply it to ethical theory.

³⁴ The example is borrowed from Egan (2006) who emphasizes that the point about location is not to be taken too literally—the properties in question when discussing centered worlds need not be geographic properties.

view, which analyzes moral rightness in terms of the desires that ideally rational agents would have. But whereas Smith requires convergence among ideally rational agents regarding those desires in order for morality to have a subject matter, Egan drops this requirement, so it's guaranteed that morality has a subject matter whether or not convergence obtains.

One might worry that on this account, what we are doing with moral language is, essentially, reporting on our desires. And this doesn't seem to be what moral discourse is concerned with. Brandt, himself a relativist, is careful to note that "what anyone is definitely committing himself not to do, when he uses moral language, is to give expression to what he personally wants" (1950: 316). The problem is illustrated by comparing a desire-based self-locating relativism with Michael Smith's view, which identifies moral rightness with the desires that *all* ideal agents would converge upon. Since it can't avail itself of the convergence requirement, though, this account leaves it unclear why anyone else should care about our moral claims, and how we can make moral demands on someone. Smith has a story to tell here: if moral rightness is what the desires of all ideally rational agents would desire, and I tell you that ϕ is morally right, I am telling you something about *your* desires in ideal conditions. But Egan can't appeal to this story. Another way of putting the problem is that, while Egan has a way of accounting for why and how my judgment that it would be wrong to ϕ gives *me* a reason not to ϕ , he has trouble accounting for why my judgment that it would be wrong to ϕ has any bearing whatsoever on whether *you* ought to ϕ . Stevenson makes a similar point: "Note that the ethical sentence centres the hearer's attention not on his interests, but on the objects of interest, and thereby facilitates suggestion" (1997: 78). Stevenson draws

our attention to the fact that not only are moral sentences not purely egocentric, they attempt to influence the beliefs and behavior of others. But why would a claim about my desires have any power to recommend an action to, or demand certain treatment from, others?

Egan might respond to this objection by appealing to an account of assertions as an attempt to add propositions to the ‘conversational scoreboard’ or common ground of one’s interlocutors. This explains why we bother to make moral assertions at all: we are trying to get our interlocutor to self-attribute the same property we are self-attributing. Another way to put the point is that there is a presupposition of relevant similarity in play. If I am trying to get you to self-attribute the property I’m ascribing to myself, it’s because I think we’re similar in both having the property. This response can explain why we make moral assertions; it’s because we want to get the other person to self-attribute our desires. And if we assume that both parties are aware of the presupposition of relevant similarity in play, we can explain how the act of assertion succeeds. However, this response is fairly limited insofar as it still doesn’t answer the question of why we care to get others to self-attribute in this way, and how we are able to make moral claims across conversations. It also seems unable to account for why we are motivated to punish wrongdoers, even if they have not wronged *us* but rather some third party.

This objection can be solved by taking Egan up on his suggestion that “Anybody can build a self-locating theory at home: start with your favorite naturalistic theory... and just take the universal quantifier off the front” (ms: 6). One way of amending the analysis to restore the social, reason-giving nature of morality is to substitute the attitudes of approval and disapproval in place of desire. At first glance this might not seem like

any improvement at all: can't we just run the same objection against 'approval'? The key to understanding why this is an improvement lies in a particular gloss on the attitudes of approval and disapproval, one discussed by Philippa Foot, who writes that "it is no more possible for a single individual to approve or disapprove than it is for him to vote" because "approval and disapproval can, logically speaking, exist only against a background of agreement about the part that other people's views shall be given in decision making" (2003: 190, 199). This point is echoed by Darwall's discussion of the second-person standpoint; Darwall argues that in making moral claims upon others, we offer reasons whose validity depends on presupposed authority and accountability relations between persons. "[B]eing subject to moral obligations includes accountability to those with the normative standing to demand compliance... morality as equal accountability understands the moral point of view to be fundamentally *intersubjective*" (Darwall 2006: 102-103). Approval and disapproval capture this intersubjectivity, since "the attitudes of approval and disapproval would not be what they are without the existence of tacit agreement on the question of who listens to whom and about what" (Foot 2003: 198). Approval and disapproval have an affective component, but also make reference to what we expect from others and what they can expect from us. Sripada (2005: 779) suggests that "the most commonly used [and most effective at stabilizing/maintaining compliance with moral norms] kind of punishment in human societies is reputation-based punishment," either via public denunciation or shaming, or by gossip and ostracism. This kind of punishment can only be effective, however, when there is coordination on "a certain assumption about the determination of conduct in a certain area," to borrow a phrase from Foot (2003: 203).

Substituting approval for desire, then, avoids the complaint that self-locating relativism is overly egocentric, while retaining the link that desire provided to motivation (via the affective component of approval). Rather than analyzing the claim ‘ ϕ -ing is morally right’ as ‘ ϕ -ing would be desired by any rational agent’ (as Smith does) or as ‘my (rational, informed) self desires ϕ -ing’ (as Egan does), I am proposing we analyze it as ‘my (rational, informed) self approves of ϕ -ing’. So when I judge that keeping promises is morally right, I am saying that I approve of keeping promises. Rather than building the imperative force of my claim into the form of the analysis itself, this account builds it into the psychological state of approval. This might seem to be just a special instance of, or variation on, desire; as Foot notes, “approving and disapproving are thought [incorrectly] to be rather like wanting and not wanting... the ill-defined term of art ‘pro-attitude’ is used to slide between wanting and approving” (2003: 192). But Foot points out that approval is essentially interpersonal in a way that wanting is not. The attitudes of approval and disapproval presuppose “some kind of influence... those able to approve or disapprove are as a general rule taken account of or listened to” (2003: 198). While approval *is* a pro-attitude, it is a special kind: it assumes “the existence of tacit agreement on who listens to whom and about what... where we have approval or disapproval we necessarily have such agreements, though it may not be necessary that we have just the agreements that we do.” (2001: 198) One way to put the difference between the analysis I’m proposing here and Egan’s analysis is that whereas Egan analyzes moral claims in terms of desire, and then appeals to the presupposition of relevant similarity— as a general feature of all assertion— to explain why we assert moral claims to others, the analysis I’m offering builds that into the moral claim itself, by building it into the attitude

of approval. The very possibility of the attitude, as Foot reminds us, *depends* on such a presupposition being in play. Rather than a contingent feature of moral discourse, or a feature common to all discourse, this is an essential fact about the attitudes of approval or disapproval. To reiterate a point made above, what makes moral approval and disapproval possible is “a shared assumption about the determination of conduct within a certain area.” Thus this account can explain, not only why we make moral claims and want people to accept them, but also why this is such an essential feature of moral discourse. When we accept someone’s claim of approval or disapproval, we are assenting to the assumption that we will listen to and consider that person’s views, claims, or wishes. We are assenting to the idea that the person counts or is to be “taken account of,” as Foot puts it.

The analysis proposed here is in many ways quite similar to that proposed by Stevenson, who also analyzes claims of moral rightness in terms of the attitude of approval, and is careful to distinguish this attitude from liking: “When a person likes something, he is pleased when it prospers and disappointed when it doesn’t. When a person *morally approves* of something, he experiences a rich feeling of security when it prospers, and is indignant... when it doesn’t.” But whereas Stevenson maintains that moral disagreement is always disagreement in interest, on this view it is, at least sometimes, also about “belief about interests.” (1997: 79) It is about a belief about what interests one has and shares with others.

2.2 How to be a Moral Relativist and a Moral Objectivist

A few more clarifications are in order before we can begin to see how this kind of relativism provides a solution to the puzzle described above. It is important to note that any possible worlds proposition can be transformed into a centered-worlds proposition by adding an index for an individual; in most cases, this will in no way affect the truth value that proposition gets assigned at a possible world. Any possible worlds proposition that is true in the actual world *simpliciter* will be true at (actual, t, i) for any time t and individual i we plug in. Call these 'boring centered worlds propositions'-- possible worlds propositions that can be analyzed as centered worlds proposition, but whose truth value is not altered by adopting either the former (coarse-grained) analysis or the latter, more finely-grained analysis.

This may seem like a fairly trivial observation, but appreciating this point is a crucial first step towards seeing why this view is able to handle the puzzle discussed above without forcing us into either variantism or incoherence. On a self-locating relativist view, we can accommodate the observation that some moral facts are objective and others are relative. We can accommodate intuitions that disagreements about abortion or euthanasia are faultless, as well as the intuition that any reasonable person would agree that torturing babies is just *wrong*-- no matter the agent, or the circumstances. And we can remain agnostic about the extent of moral relativism versus moral objectivity-- this account leaves the question of which moral judgments we will or would converge upon open and amenable to empirical investigation.³⁵ (This is similar to Jackson's (1998: 117) discussion of 'mature folk morality'-- on his view, whether there is

³⁵ See Chapter 2 for an argument to the effect that such convergence is not forthcoming.

an objective moral subject matter depends on the truths on which we would converge at something like 'the end of inquiry;' if there is no such convergence, moral relativism is true. But, one might respond, from our current standpoint, both outcomes are possible, so we ought not adopt a theory that rules out either possibility a priori.) We can deny that objectivity is required for morality to have a subject matter, without precluding it or ruling it out a priori. Indeed, as I will argue, with a few modifications, this kind of view can accommodate all of the data that motivate mixed or variantist metaethics while incurring none of the costs.

Let me illustrate how. Suppose that for some act ϕ , it's the case that everyone's informed, rational self would approve of ϕ -ing. Then it's the case that everyone would be correct to self-attribute the property, 'being someone whose ideally rational self approves of ϕ -ing.' Now suppose A and B are arguing over whether it's right to ϕ , with A insisting it is and B insisting it isn't. A is correctly self-attributing the property, B is mistakenly denying that he has the property; and so we have a case *not just* where two people are disagreeing and it just so happens that one is mistaken, but where if any *two* people disagree over whether ϕ is morally right, one of them must be mistaken. So it looks like we have something that is behaving an awful lot like an objective moral fact; ' ϕ -ing is morally right' is true, regardless of who says it. In other words, it's a boring centered-worlds proposition. Even those who have been impressed by the extent of moral diversity should be willing to grant the possibility that there are facts like this (and in the next chapter I will examine the evidence for thinking that, actually, there are claims like this); perhaps torturing babies for fun is something that, just by virtue of being human and having the kind of psychology we do, no one's ideally rational self would fail to

disapprove of. So for some action ϕ , it's the case that everyone would be correct to self-attribute approval of ϕ -- the truth-value of the proposition remains constant regardless of which agent we fill in the individual parameter with. Therefore, if any two people disagree over whether ϕ -ing is right, one of them is mistaken. To reiterate: this is starting to look a lot like moral objectivity.

In other cases, the action we plug in for ϕ might be approved by some ideally rational agents and disapproved of by others. Here, it's entirely possible to end up with cases of faultless disagreement, with one party correctly self-attributing a property and the other party denying they have that property. Here we get variation in truth value, depending on which agent is plugged in to the individual parameter. In the case of properties like this, we *can* have disagreements where one party is at fault-- perhaps they haven't thought fully about the issue, or there is some incoherence in their beliefs, or a factual error, etc-- but insofar as faultless disagreement is possible here, it looks like these are not objective moral facts/judgments.

This account can handle the motivations for metaethical variantism nicely. It accommodates the observation that we will be more objective about some moral violations than others. And it gives a good explanation of why this is so: when it comes to some acts, the story might go, it doesn't matter where you start from, or who you are, or what other beliefs or desires you have-- the act is one that any rational person would disapprove of. In any society, group, or culture in which people's views are taken account of, the act will be disapproved. And since shared social practice is a precondition for approval and disapproval, there just could not be a rational agent who

approved such an act.³⁶ In other cases, we do think the contingencies of personal identity make a difference, and so we are more likely to treat those judgments as relative or nonobjective. Candidates for the former kind of violation are cases involving grave and gratuitous harm and suffering; the latter types of violations concern sex, or purity, or respect. We can also explain individual differences in judgments of objectivity in terms of differences between speakers in their beliefs/assumptions about similarity: speakers who are treating moral questions objectively are assuming a kind of similarity between all agents, whereas those who are more inclined to relativism are allowing that there may be a great deal of diversity, even once we have idealized the agents in question.³⁷

3. Objections and Replies

3.1 Variantism, Redux

We've seen how a self-locating relativism can reconcile our seemingly inconsistent intuitions about objectivity. But recall that, according to Sinnott-Armstrong's definition, variantism is committed to the claim that "some uses of moral language definitely fit expressivism" (Sinnott-Armstrong 2009: 239). And the relativist view described above cannot accommodate a commitment to both cognitivism and non-cognitivism. Is this a problem? It might be, if there were empirical evidence suggesting that this claim were true. But, unfortunately for variantism, the experiments discussed

³⁶ Though there could be an agent who had a pro-attitude towards the act. But some acts are so outrageous that someone who was for them would be 'out,' as it were. Foot notes that "the presumption that account [of someone's views] will be taken is only a presumption, and may be destroyed," and in such a case, the person really will not be able to approve the act, since he has forfeited claims to consideration by being for such an outrageous act. Rationality, here, means that one is (or one's judgment is) entitled to consideration and is prepared to grant such consideration to others (or to others' judgments).

³⁷ See Chapter 2.

above fail to provide any evidence that there are, in fact, uses that definitely fit expressivism. So while we have compelling empirical evidence for variation in intuitions about objectivity—that our discourse includes claims to and against objectivity—we have no evidence for the claim that moral discourse includes claims to and against cognitivism. Here's why.

Recall that in Goodwin and Darley's experiments, the nonobjectivist answer was measured by whether the subjects rated the statement of the transgression as describing "an opinion or attitude." However, the difference between an opinion and an attitude, while perhaps dispensable for purposes of measuring only objectivism, is crucial to detecting the difference between an expressivist and relativist view. A response of "opinion or attitude" could indicate either that the statement in question is not a fact because different people might have equally justified, or equally true beliefs about it, *or* that it is not a fact because attitudes aren't the kinds of things that are true or false. So this experiment provides no support for variantism as defined by Sinnott-Armstrong, though it does provide strong support for a weaker kind of variantism, a version akin to Loeb's (2008) claim that our moral discourse includes commitments both to and against objectivity—an observation that, as we've seen, is explained on a self-locating relativist view.

Likewise, in Nichols' experiments, giving a nonobjectivist answer is no indication of non-cognitivism. If anything, it is support for a kind of relativism. To wit: those who give the nonobjectivist answer are agreeing to a statement that emphasizes the 'unqualified' nature of the claim and goes on to state that 'different cultures *believe* different things...it is not *absolutely true* that it is okay to hit people...'

(my emphasis). Assent to this claim hardly looks like a ringing endorsement of non-cognitivism. Rather, it looks as though the nonobjectivist subjects are saying something like, 'well, it can be true for me and false for someone else, it depends on what you and yours believe,' which looks a lot like a kind of relativist analysis. The wording of both the objectivist and the nonobjectivist answers contains an implicit assumption-- or at least insinuation-- of the truth-aptness of moral claims, regardless of the kind of truth in question. It implies that they are of a cognitive piece with beliefs. So this experiment can't possibly yield any evidence bearing on the cognitivism/noncognitivism debate.

3.2 It's Normative, but What Makes it Moral?

Another objection to the idea that metaethical analyses ought to or even can accommodate both relativist and objective moral facts runs as follows: If there are any moral facts at all, they are objective. Therefore, the purported discovery that some moral claims are relative while others are objective is actually just the claim that there are fewer moral facts than we thought. So what we've discovered is not that morality contains objective and non-objective commitments, but rather that some judgments we had taken to be moral are in fact non-moral normative judgments.³⁸ The problem with this line of argument is that it makes objectivity too cheap—by making objectivity a defining feature of morality, it rules out a priori the possibility of discovering genuine moral diversity. Because once we discover that there is widespread disagreement over the moral permissibility of incest, or of corporal punishment, we have thereby discovered that the

³⁸ For a more detailed discussion of how to identify and distinguish moral judgments and norms, with attention to empirical and cross-cultural cases, see ch. 2, where I argue for the importance of identifying standards that are non-ethnocentric and are as ecumenical as possible.

issue is not a moral one at all! And this is problematic insofar as it does not seem like moral disagreement is conceptually incoherent. The strategy of argument outlined above gets objectivity on the cheap, but at the cost of denying the possibility of genuine moral diversity.

The objection can be reiterated at the level of judgment. If subjects are not judging an act objectively wrong, the thought goes, they are not judging it *morally* wrong. Rather, they are saying it is forbidden according to some system of rules or another, but this is not the same as saying it is morally forbidden.

One instance of this argument is represented in the literature on the ‘moral/conventional distinction.’ This distinction is employed by philosophers (Nichols 2004) and psychologists (see, e.g., Turiel 1983 and Nucci 2001) alike to demarcate the moral from the social domain. According to this approach, moral norms are more serious than conventional norms, their scope is global, and they do not depend on rules or authority figures for their legitimacy—hitting would be wrong, for example, even if there were no rule against it. Conventional violations are less serious; their scope is local, and they are authority-contingent (i.e., if the teacher had no rule against chewing gum in class, it would be okay to chew gum in class). Subjects who do judge something wrong, but not globally wrong, the argument goes, are judging it conventionally, rather than morally wrong—even if their vocabulary is too confused to express this fact.

Promising though it might seem, the moral/conventional distinction is ultimately untenable unless supported by some other way of distinguishing moral from other normative violations. To see why, consider disgusting norm violations. Recall that on this account, what distinguishes moral from conventional violations is a signature

response pattern: moral violations are more serious, global and non-authority contingent. Now consider the following violation: at a dinner party, Mark spits in his water glass and then drinks from it. When subjects are given a description of Mark's behavior, they exhibit the response pattern characteristic of moral violations—they judge it as more wrong than non-disgusting etiquette violations, and they treat it as global in scope and not authority contingent (Mark can't cancel out the wrongness by fiat). So it looks like on the account we are considering, spitting in one's glass and drinking from it is a moral violation. But now things look a bit complicated. For now we find ourselves forced to treat etiquette violations as moral violations. If this seems odd to you—if you think, for example, that blowing one's nose in a napkin and then using it to wipe one's mouth is gross, sure, but not immoral—as it does to me, this shows that we are drawing on some antecedent distinction between the moral and the conventional.³⁹ We distinguish bad manners from bad morals. But the defender of the account under consideration here can't appeal to any such distinction, since for him, the distinguishing feature just is the signature response pattern. And yet philosophers who rely on the moral/conventional distinction typically do not tout the surprising discovery that the rule against spitting in a glass and then drinking from it is a moral rule; rather, they continue to talk about disgust-backed etiquette norms (see, for example, Nichols 2002). This suggests that they are relying on something other than the supposedly signature response pattern of moral judgments to distinguish moral and social rules, and therefore that judgments of objectivity are not an identifying characteristic of moral norms.

³⁹ For an argument against treating judgments of disgustingness as moral, see Chapter One.

Conclusion

Moral relativists have, in the past, conceded too much too quickly to moral realists. Relativism is not, in fact, a revisionary position. Furthermore, relativism and moral objectivity can coexist peacefully. Relativism leaves questions about the extent of moral disagreement and convergence open to empirical inquiry, as they should be; it explains the subject matter of moral discourse in a way that does not require an a priori commitment to full objectivity. It allows for variation within moral discourse, with some claims being more objective than others. And it does not preclude the possibility of either moral mistakes or moral criticism. Some may complain that the type of account described here makes morality unacceptably anthropocentric: the kind of objectivity we end up with is, in the end, an objectivity that is contingent on the attitudes of actual moral agents. This seems to me to be an asset rather than a cost, though this may, in the end, be a matter of temperament. Nonetheless, I think the account described above exposes what is at issue in many if not all of these debates about objectivity: the extent to which the minimum requirements of rationality, combined with facts about human psychology, determine a unique set of moral norms. As such, it is both naturalistically respectable and philosophically illuminating.

Works Cited

- Allport, G.W. 1954. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.
- Anonymous. 2000. *Dissoi Logoi*. In Waterfield, ed., *The First Philosophers*, 285-299. Oxford: Oxford University Press.
- Blackburn, S. 2001. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.
- Bodenhausen, G. V., Sheppard, L., & Kramer, G. P. 1994. Negative affect and Social Perception: The differential impact of anger and sadness. *European Journal of Social Psychology*, 24: 45-62.
- Boyd, R. 1988. How to Be a Moral Realist. In Sayre-McCord, ed. *Essays on Moral Realism*. Ithaca: Cornell University Press.
- Brandt, R. 1944. The Significance of Differences of Ethical Opinion for Ethical Rationalism. *Philosophy and Phenomenological Research*, 4(4): 469-495.
- Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Casasanto, D. 2009. Embodiment of Abstract Concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology* 138(3): 351-367.
- Chapman, H., Kim, D., Susskind, J., & Anderson, A. 2009. In Bad Taste: Evidence for the oral origins of moral disgust. *Science* 323: 1222-1226.
- Cosmides, L., & Tooby, J. 2005. Neurocognitive Adaptations Designed for Social Exchange. In David Buss, ed., *The Handbook of Evolutionary Psychology*. Hoboken: Wiley & Sons.
- Couzin, J. 2009. Friendship as Health Factor. *Science*, 323(5913): 454-457.
- D'Arms, J., & Jacobson, D. 2000. The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research*, 61(1): 65-90.

- Daniels, N. 1979. Wide Reflective Equilibrium and Theory Acceptance in Ethics. *Journal of Philosophy* 76(5): 256-282.
- Darwall, S. 1998. *Philosophical Ethics*. Boulder: Westview Press.
- Darwall, S. 2006. *The Second Person Standpoint*. Cambridge, MA: Harvard University Press.
- Doris, J., & Murphy, D. 2007. From My Lai to Abu Ghraib: The Moral Psychology of Atrocity. *Midwest Studies in Philosophy*, 31(1): 25-55.
- Doris, J., & Plakias, A. 2008. How to Argue about Disagreement. In Walter Sinnott-Armstrong, ed, *Moral Psychology: Volume II*, 303-331. Cambridge, MA: MIT Press.
- Dover, K.J. 1994. *Greek Popular Morality in the Time of Plato and Aristotle*. Indianapolis: Hackett.
- Dreier, J. 2006. Relativism and Nihilism. In D. Copp, ed., *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Egan, A. ms. Why Ethics is All About Me. Rutgers University, Department of Philosophy.
- Empiricus, S. 1933. *Outlines of Pyrrhonism*. R.G. Bury, trans. New York: Putnam.
- Enoch, D. 2009. How is Disagreement a Problem for Realism? *Journal of Ethics*, 13(1): 15-50.
- Fallon, A.E., & Rozin, P. 1983. The Psychological Basis of Food Rejections by Humans. *Ecology of Food and Nutrition*, 13: 15-26.
- Feldman, R. 2007. Reasonable Religious Disagreements. In Antony, L., ed. *Philosophers Without Gods*. 194-214. Oxford: Oxford University Press.
- Foot, P. 2003. Approval and Disapproval. In *Virtues and Vices: Essays in Moral Philosophy*. Oxford: Oxford University Press. 189-208.
- Frandsen, P.J. 2009. *Incestuous and Close-Kin Marriage in Ancient Egypt and Persia: An Examination of the Evidence*. Copenhagen: Museum Tusculanum Press.
- Frazer, J. 1967. *The Golden Bough*. New York: MacMillan.
- Gert, J. 2005. Neo-Sentimentalism and Disgust. *The Journal of Value Inquiry*, 39: 345-352.

- Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Oxford: Oxford University Press.
- Goodwin, G. and Darley, J. 2007. The Psychology of Meta-Ethics: Exploring Objectivism. *Cognition*, 106: 1339-1366.
- Goodwin, G. & Darley, J. ms. Why Are Some Moral Beliefs Perceived to be More Objective Than Others? Princeton University & University of Pennsylvania, Department of Psychology.
- Goody, J. 1956. A Comparative Approach to Incest and Adultery. *The British Journal of Sociology*, 7(4): 286-305.
- Greene, J. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293: 2105-2108.
- Greene, J. The Secret Joke of Kant's Soul. In Walter Sinnott-Armstrong, ed., *Moral Psychology: Volume III*, 35-80. Cambridge, MA: MIT Press.
- Grim, P., et al. 2005. Modeling Prejudice Reduction: Spatialized Game Theory and the Contact Hypothesis. *Public Affairs Quarterly* 19(2): 95-125.
- Haidt, J. 2001. The Emotional Dog and Its Rational Tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4): 814-834.
- Haidt, J. and Bjorklund, F. 2008. Social Intuitionists Answer Six Questions About Moral Psychology. In Walter Sinnott-Armstrong, ed., *Moral Psychology: Volume II*. 303-331. Cambridge, MA: MIT Press.
- Haidt, J., Koller, S., & Dias, M. 1993. Affect, Culture and Morality, or, Is it Wrong to Eat Your Dog? *Journal of Personality and Social Psychology*, 65: 613-628.
- Haidt, J., Rozin, P., McCauley, C., & Imada, S. 1997. Body, Psyche, and Culture: The relationship of disgust to morality. *Psychology and Developing Societies*, 9: 107-131.
- Harman, G. 1988. Ethics and Observation. In Sayre-McCord, ed., *Essays on Moral Realism*. 119-124. Ithaca: Cornell University Press.
- Hauskeller, M. 2006. Moral Disgust. *Ethical Perspectives* 13(4): 571-602.
- Henrich, J., Boyd, R., Bowles, S. et al. 2005. 'Economic Man' in Cross-cultural Perspective: Behavioral Experiments in 15 Small-scale Societies. *Behavioral and Brain Sciences*, 28 (6): 795-814.

- Hill, A., et. al. 2010. Emotions as Infectious Diseases in a Large Social Network. *Proceedings of the Royal British Society of Biological Sciences*, 277 (1701): 3827-3835.
- Horgan, T., & Timmons, M. 2008. What Does Moral Phenomenology Tell Us About Moral Objectivity? *Social Philosophy and Policy*, 25 (1), 267-300.
- Inbar, Y., Pizarro, D., & Bloom, P. 2009a. Conservatives Are More Easily Disgusted Than Liberals. *Cognition and Emotion*, 23, 714-725.
- Inbar, Y., Pizarro, D., Knobe, J., & Bloom, P. 2009b. Disgust Sensitivity Predicts Intuitive Disapproval of Gays. *Emotion*, 9, 435-439.
- Isen, A. 1984. Towards Understanding the Role of Affect in Cognition. In *Handbook of Social Cognition*. Wyer, R. S. & Srull, T. K. Eds.. Erlbaum: Hillsdale, NJ.
- Jackson, F. 1998. *From Metaphysics to Ethics: a Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Joyce, R. 2002. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2007. *The Evolution of Morality*. Cambridge: MIT Press.
- Kahan, D. 1998. *The Anatomy of Disgust* in Criminal Law. *Michigan Law Review* 96: 1621-1657.
- Kahane, G. 2011. Evolutionary Debunking Arguments. *Nous* 45(1): 103-125.
- Kass, L. 1997. The Wisdom of Repugnance. *New Republic*, 216(22): 17-26.
- Kekes, J. 1992. Disgust and Moral Taboos. *Philosophy*, 67(262): 431-446.
- Kelly, D. In press. *Yuck!: The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.
- Kelly, T. 2008. Disagreement, Dogmatism, and Belief Polarization. *Journal of Philosophy*, 105(10): 611-633.
- Keltner, D., Ellsworth, P., & Edwards, K. 1993. Beyond Simple Pessimism: Effects of sadness and anger on social perception. *Journal of Personality and Social Psychology*, 64(5): 740-752.
- Knapp, C. 2003. De-moralizing Disgustingness. *Philosophy and Phenomenological Research* 66(2): 253-278.

- Laham, S., Alter, A., & Goodwin, G. Easy on the Mind, Easy on the Wrongdoer: Discrepantly easy to process violations are deemed less morally wrong. *Cognition*, 112(3): 462-466.
- Lane, H. 1976. *The Wild Boy of Aveyron*. Cambridge: Harvard University Press.
- Larmore, C. 1987. *Patterns of Moral Complexity*. Cambridge: Cambridge University Press.
- Leiter, B. 2009. Moral Skepticism and Moral Disagreement in Nietzsche. University of Chicago, Public Law Working Paper no. 257. Available at SSRN: <http://ssrn.com/abstract=1315061>
- Levine, L., & Pizarro, D. 2006. Emotional Valence, Discrete Emotions, and Memory. In B. Utl, N. Ohta, & A.L. Siegenthaler, eds., *Memory: Interdisciplinary Perspectives*. Blackwell Publishing.
- Levi-Strauss, C. 1966. *The Savage Mind*. Chicago: University of Chicago Press.
- Lewis, D. 1979. Attitudes De Dicto and De Se. *Philosophical Review* 88(4): 514-543.
- Loeb, D. 2003. Gastronomic Realism- A Cautionary Tale. *Journal of Theoretical and Philosophical Psychology*. 23(1): 30-49.
- Loeb, D. 2008. How to Pull a Metaphysical Rabbit out of a Semantic Hat. In Sinnott-Armstrong, ed. *Moral Psychology: Volume II*, 355-385. Cambridge: MIT Press.
- Machery, E., Kelly, D., and Stich, S. 2005. Moral Realism and Cross-Cultural Normative Diversity: Comments on 'Economic Man' in Cross-cultural Perspective: Behavioral Experiments in 15 Small-scale Societies. *Behavioral and Brain Sciences*, 28(6): 830-831.
- MacIntyre, A. 1981. *After Virtue*. Notre Dame: University of Notre Dame Press.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.
- McGrath, S. 2008. Moral Disagreement and Moral Expertise. In Shafer-Landau, ed. *Oxford Studies in Metaethics*, v 3: 87-107.
- Miller, William. 1997. *The Anatomy of Disgust*. Cambridge: Harvard University Press.
- Moody-Adams, M. 1997. *Fieldwork in Familiar Places*. Cambridge: Harvard University Press.
- Murphy, D. 2010. Philosophy of Psychiatry. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.)

- Nichols, S. 2002. "On the Genealogy of Norms: A Case for the Role of Emotion in Moral Judgment." *Philosophy of Science*, 69.
- Nichols, S. 2004a. After Objectivity: An Empirical Study of Moral Judgment. *Philosophical Psychology*, 17: 5-28.
- Nichols, S. 2004b. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- Nichols, S. 2007. The Rise of Compatibilism: A Case Study in the Quantitative History of Philosophy. *Midwest Studies in Philosophy*, 31(1): 260-270.
- Nisbett, R.E., & Cohen, D. 1996. *Culture of Honor: The Psychology of Violence in the South*. Boulder: Westview Press.
- Nisbett, R. E. 2003. *The Geography of Thought: How Asians and Westerners Think Differently... and Why*. New York: Free Press.
- Nordgren, L., and McDonnell, M. 2010. The Scope-Severity Paradox: Why doing more harm is judged to be less harmful. *Social Psychological and Personality Science*, available online only.
- Nucci, L. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- Nussbaum, M. 1994. Platonic Love and Colorado Law: The Relevance of Ancient Greek Norms to Modern Sexual Controversies. *Virginia Law Review*, 80(7): 1515-1651.
- Nussbaum, M. 2004. *Hiding from Humanity: Disgust, shame and the law*. Princeton: Princeton University Press.
- Park, J. & Banaji, M. 2000. Mood and Heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology*, 78: 1005-1023.
- Phillips, D. 1974. The Influence of Suggestion on Suicide: Substantive and Theoretical implications of the Werther Effect. *American Sociological Review*, 39: 340-354.
- Pizarro, D., Inbar, Y., & Helion, C. In press. On Disgust and Moral Judgment. *Emotion Review*.
- Plakias, A., & Doris, J. 2008. How to Find a Moral Disagreement. In Walter Sinnott-Armstrong, ed, *Moral Psychology: Volume II*, 345-353. Cambridge: MIT Press.

- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Rochat et al. 2009. Fairness in Distributive Justice by 3- and 5-Year Olds Across 7 Cultures. *Journal of Cross-Cultural Psychology*, 40(3): 416-442.
- Rozin, P., Haidt, J., & McCauley, C. R. 2000. Disgust. In Lewis & Haviland, eds. *Handbook of emotions*, 2nd edition, (pp.637-653). New York: Guilford Press.
- Sarkissian, H., et al. In press. Folk Moral Relativism. *Mind and Language*.
- Scheidel, W. 1996. Brother-Sister and Parent-Child Marriage Outside Royal Families in Ancient Egypt and Iran: A Challenge to the Sociobiological View of Incest Avoidance? *Ethology and Sociobiology*, 17(5): 319-340.
- Schnall, S., Haidt, J., Clore, G., & Jordan, A. 2008. Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin*, 34: 1096-1109.
- Schwarz, N. 2002. Situated Cognition and the Wisdom of Feelings: Cognitive tuning. In L. Feldman Barrett & P. Salovey (Eds.), *The Wisdom in Feelings*. New York: Guilford.
- Schwarz, N. 2004. Metacognitive Experiences in Consumer Judgment and Decision Making. *Journal of Consumer Psychology*, 14(4), 332-348.
- Shafer-Landau, R. 1994. Ethical Disagreement, Ethical Objectivism and Moral Disagreement. *Philosophy and Phenomenological Research*, 54(2): 331-344.
- Shafer-Landau, R. *Moral Realism: A Defence*. Oxford: Clarendon Press.
- Shepher, J. 1971. Mate Selection Among Second Generation Adolescents and Adults: Incest Avoidance and Negative Imprinting. *Archives of Sexual Behavior*, 1(4): 293-307.
- Singer, P. 2005. Ethics and Intuitions. *Journal of Ethics*, 9(3-4): 331-352.
- Sinnott-Armstrong, W. 2009. Mixed-up Metaethics. *Philosophical Issues* 19: 235-255.
- Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
- Smith, M. 1991. Realism. In P. Singer, ed., *A Companion to Ethics*, 399-410. Oxford: Blackwell.
- Song, H., & Schwarz, N. 2010. If it's Easy to Read, it's Easy to Do, Pretty, Good, and True. *The Psychologist*, 23(2): 108-111.

- Sommers, T. 2009. The Two Faces of Revenge: Moral Responsibility and the Culture of Honor. *Biology and Philosophy*, 24 (1): 35-50.
- Strohminger, N., et al. In press. Divergent Effects of Different Positive Emotions on Moral Judgment. *Cognition*.
- Stevenson, C.L. 1997. The Emotive Meaning of Ethical Terms. In Darwall, Gibbard and Railton, eds. *Moral Discourse and Practice*, 71-82. Oxford: Oxford University Press.
- Street, S. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127 (1): 109-166.
- Thibodeau, P.H., & Boroditsky, L. 2011. Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE* 6 (2): e16782. doi:10.1371/journal.pone.0016782
- Tucker, H. 2011. *Blood Work: A Tale of Medicine and Murder in the Scientific Revolution*. New York: W.W. Norton.
- 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 three functional domains of disgust. *Journal of Personality and Social Psychology*, 97: 103-122.
- Valdesolo, P. & DeSteno, D. 2006. Manipulations of emotional context shape moral judgment. *Psychological Science*, 17: 476-477.
- Wedgwood, R. 2010. The Moral Evil Demons. In Feldman & Warfield, eds, *Disagreement*, 216-246. Oxford: Clarendon Press.
- Westermarck, E. 1891. *The History of Human Marriage*. New York: MacMillan.
- Wheatley, T., & Haidt, J. 2005. Hypnotically Induced Disgust Makes Moral Judgments More Severe. *Psychological Science*, 16: 780-784.
- Williams, B. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
- Williams, N., Connolly, K., Cisler, J. et. al. 2009. Disgust: A cognitive approach. In B. Olatunji & D. McKay, eds, *Disgust and its Disorders*. Washington, D.C.: American Psychological Association.

Zimbardo, P. 2007. *The Lucifer Effect*. New York: Random House.