View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks@UMass Amherst

University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Doctoral Dissertations

Dissertations and Theses

July 2019

The Epistemic Dimensions of Moral Responsibility and Respect

John Robison

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Epistemology Commons, and the Ethics and Political Philosophy Commons

Recommended Citation

Robison, John, "The Epistemic Dimensions of Moral Responsibility and Respect" (2019). *Doctoral Dissertations*. 1654.
https://scholarworks.umass.edu/dissertations_2/1654

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

THE EPISTEMIC DIMENSIONS OF MORAL RESPONSIBILITY AND RESPECT

A Dissertation Presented

by

JOHN W. ROBISON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2019

Philosophy

© Copyright by John W. Robison 2019

All Rights Reserved

THE EPISTEMIC DIMENSIONS OF MORAL RESPONSIBILITY AND RESPECT

A Dissertation Presented

By

JOHN WILLIAM ROBISON

Approved as to style and content by:

____
Hilary Kornblith, Chair

____
Sophie Horowitz, Member

____
Katia Vavova, Member

____
Timothy Pachirat, Member

____
Joseph Levine, Department Chair
Philosophy

DEDICATION

For my father, Curt Robison, and my twin brother, Chris Robison

ACKNOWLEDGMENTS

Many people contributed toward making it possible and meaningful for me to write this dissertation, and I am deeply grateful to them all.

I begin by noting some of my professors at University of Rochester. Richard Feldman and Ralf Meerbote played major roles in making philosophy seem exciting and worth thinking about. Alexis Luko was the first professor I'd had who sincerely invited me to begin to think of myself as a scholar of sorts. I thank the three of them for their inspiration and guidance.

At UMass, I had the great fortune of learning from wonderful and generous professors. Though he entered into retirement shortly after my first year in graduate school, Fred Feldman significantly shaped the early phase of my graduate career. I learned a lot from Fred about how to pursue a line of questions with clarity and integrity. To this day, when I edit work, I will sometimes get mental imagery of Fred bewilderedly or incredulously reading aloud a claim I've made, and this helps me know that I am badly relying on an obscure idea. So, what I am saying is that Fred is a daimonion for me.

I also had very many wonderful experiences learning how to think, write, navigate the profession, and/or teach from Phil Bricker, Vanessa de Harven, Pete Graham, Gary Ostertag, and Alejandro Pérez Carballo. Vanessa, in particular, gave me much needed encouragement and advice at several junctures in my graduate career. I am importantly indebted to each of them in various ways, and I thank them all. I also thank Timothy Pachirat, the outside member of my dissertation committee, for his contributions to my dissertation and for helping me think about ways to share my research with a general audience. I thank Joe Levine, who was Chair of the department for most of my time at

UMass: I felt more than supported by the department, and this helped tremendously with the writing of my dissertation. And I also thank Beth Grybko, without whom it would have been super complicated to be a graduate student.

I owe an enormous thank you to Eileen O'Neill, whom I miss very dearly. Eileen was a maximally giving and inspiring presence in my life, and no professor played a bigger role in making me feel like I belonged in a philosophical community. I am overwhelmed when I reflect on her dedication to her students, to her craft, and to doing philosophy that matters. I feel incredibly fortunate that I got to know her and learn from her.

I had wonderful colleagues and friends in the Department of Philosophy at UMass—the entire graduate student community, in some form or other, made it deeply rewarding to work on philosophy here. For their kindness, friendship, encouragement, and philosophical exchange, I would like to thank in particular: Bob Gruber, Tricia Magalotti, Andréa Wilson, Thomas Shea, Tim Juvshik, Dallas Amico, Justis Koon, Emelia Miller, Daniel Hadad, Luis Oliveira, Eunjin Lee, Miles Tucker, Keehyuk Nahm, Lu Chen, Melanie Ludwig, and Haoying Lee. And I especially thank Molly O'Rourke-Friel for her deep friendship, her support, and her perspective.

My friends were enormously supportive during my entire graduate career. Here and in virtually all other contexts, I owe so much to Neil Chaudhary, Michael Pontacoloni, Kate Saxton, Justin Verrier, and Chris Zizzamia. I have had the outrageous fortune of continuously living in the same general area as Sean Elligers since I got to know him around 2003. Sean is intimately bound up with every remotely meaningful project I've ever tried to pursue, this dissertation included.

My family has been deeply encouraging, supportive, and understanding throughout my graduate career. I thank Cynthia Thomas, Lois Thomas, Roscoe Robison, Nina Robison, Delphine Robison, Peter Robison, Nolwenn Robison, and Kaelig Robison for constituting a very important network of love and support. I especially thank my dad, Curt Robison. He has been exceptionally encouraging and supportive in everything I do. It has been deeply meaningful for me to share this journey with him.

I also thank my twin brother, Chris Robison. The vast majority of people don't come into the world packaged with a best friend with whom to go through life. I do not understand how these people do it. Chris is there for me in everything I pursue. Some of the central contributions of this dissertation were developed or refined on walks with him where we would talk about our research projects and look for vegan baked goods.

Now, I wish to thank the three people who had the most profound and direct influence on my philosophical thinking and writing for the dissertation: Katia Vavova, Sophie Horowitz, and Hilary Kornblith. I cannot imagine a better group of people to work with. All of them were incredibly generous, kind, and extraordinarily helpful in all the ways one could possibly want one's dissertation committee to be helpful. Since they were all helpful in all ways, it feels a bit arbitrary to isolate a few things about each of them. But I will do it anyway.

Katia met with me several times at coffee shops and at Mount Holyoke College to discuss drafts of chapters. Her feedback on my ideas and on my prose was always extremely valuable. Moreover, every time we would meet to discuss a chapter, I would leave the conversation feeling as though there were at least three new projects at the intersection of epistemology and ethics that I immediately wanted to turn to.

Sophie taught me that there is more to writing a compelling paper than writing a paper with clear and seemingly sound arguments on an interesting topic. In my default mode, I try to motivate my papers by problematizing some popular view in the literature. Sophie helped me recognize the extent to which I often treated my own contributions as afterthoughts, and she helped me think about ways to stop doing that.

I cannot overstate my level of gratitude toward Hilary, my dissertation advisor. I am pretty sure that he has been in his office with the door open for the entirety of my graduate career. At every stage of the program, I have been able to turn to Hilary for guidance and support—not just in working through some puzzle in philosophy, but in every facet of being a person trying to navigate a graduate program. He has had an enormous influence on my approach to philosophy, and I feel extraordinarily fortunate to have been able to have countless deeply rewarding, fun, meaningful, and illuminating conversations with him.

And, finally, I thank the two friends who accompanied me throughout the entire dissertation-writing process: Andy and Lu.

ABSTRACT

THE EPISTEMIC DIMENSIONS OF MORAL RESPONSIBILITY AND RESPECT

MAY 2019

JOHN WILLIAM ROBISON

B.A., UNIVERSITY OF ROCHESTER

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

What epistemic conditions must one satisfy to be morally responsible for an action or attitude? A common worry is that robust epistemic requirements would have disastrous implications for our responsibility attributing practices: we would be unable to make epistemically justified responsibility attributions, or we would be licensed to disrespectfully excuse agents for their sincerely held beliefs. Those more optimistic about robust epistemic requirements inadvertently make them too demanding to explain the moral successes of ordinary agents. The present project shows how both the pessimists and optimists rely on instructively mistaken assumptions in epistemology, ethics, and action theory, and it culminates in a theoretical framework for responsibility for right action (or moral worth) from which well-motivated and unproblematic epistemic requirements fall out. A right action is morally worthy, I argue, just to the extent that it is explained by a reliable tie to the right that is secured through the influence one's values have (perhaps unreflectively) on one's informational access and processing. This Value-Secured Reliability framework has wide-reaching import and readily extends to a further variety of moral success: respect.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS v

ABSTRACT ix

CHAPTER

1. INTRODUCTION 1

    1.1 Introduction 1
    1.2 Questions about Responsibility 2
    1.3 Questions about Respect 3
    1.4 The Plan 7

PART I: VINDICATING EPISTEMIC REQUIREMENTS ON RESPONSIBILITY 13

2. SKEPTICISM ABOUT SKEPTICISM ABOUT MORAL RESPONSIBILITY 14

    2.1 Introduction 14
    2.2 Rosen's Argument 17

        2.2.1 A Necessary Condition of Blameworthy Acts 18
        2.2.2 The Skeptical Argument 22

    2.3 Skepticism about Judgments of Praiseworthiness 26
    2.4 Skepticism about Judgments of Doxastic Justification 32
    2.5 Justified Judgments about Doxastic Justification and Blameworthiness 36

3. WHEN AND WHY IS IT DISRESPECTFUL TO EXCUSE AN ATTITUDE? 48

    3.1 Introduction 48
    3.2 Clarification and the DISRESPECT THESIS 50
    3.3 Rationales for the DISRESPECT THESIS 54

        3.3.1 The Incapacity Argument 54
        3.3.2 The Rational-Relations Argument 56

    3.4 Respect and Proper Responsiveness 71
    3.5 Conclusion 76

PART II: THE VALUE-SECURED RELIABILITY THEORY 78

4. MORAL WORTH AND CONSCIOUSNESS: IN DEFENSE OF A VALUE-SECURED RELIABILITY THEORY 79

4.1 Introduction 79
4.2 Clarification, Disambiguation, and Desiderata on a Minimal Consciousness Condition 83
4.3 Consciousness of the Right-Making Facts 87
4.4 The Value-Secured Reliability Theory of Moral Worth 105
4.5 Conclusion 115

5. EPISTEMIC RESPECT AND CREDIBILITY EXCESS 118

5.1 Introduction 118
5.2 The Dominant Perspective on the (In)significance of Credibility Excess 120
5.3 Targeting Credibility Excess as Disrespectful 122
5.4 The Value-Secured Reliability Theory of Respect 128
5.5 Conclusion 138

6. CONCLUSION 141

BIBLIOGRAPHY 149

CHAPTER 1

INTRODUCTION

1.1 Introduction

At the largest level, this collection of papers seeks to explicate the epistemic or cognitive requirements on moral responsibility and respect. It asks how facts about how one accesses and processes information bear upon 1) whether one is responsible for some action or attitude and upon 2) whether some judgment is respectful.

Philosophers as far back as Aristotle (*Nicomachean Ethics*, 1110a-1111b4)[1] have recognized that moral responsibility has epistemic requirements of some form, though these requirements have received much less theoretical attention than (say) control requirements have. That *respect* has epistemic requirements has received even less in the way of direct and sustained treatment.[2] I believe that investigations into the possible epistemic requirements on each of these moral phenomena will prove mutually illuminating. Indeed, I ultimately argue that moral responsibility for right action (or moral worth) and respect have quite parallel theoretical structures, and both crucially involve an

---

[1]While Aristotle explicitly discusses epistemic requirements on responsibility, Plato at least makes an implicit suggestion about them: if I do not know that I am corrupting the youth, then I should not be punished for so corrupting them (*Apology*, 26a).

[2]Stephen Darwall's (1977) discussion of "recognition respect" and Robin Dillon's (1992) discussion of "care respect," which she takes to be a species of Darwall's "recognition respect" (Dillon, 1992: 112), loosely touch upon epistemic requirements on respect. Both characterize respect as centrally involving one person recognizing another as the agent she is. However, much is left unresolved in these discussions. Does respect require that one *factively* recognize the other as the agent she is? Does it require that one's judgments satisfy some epistemic standards? Which ones?

agent's accessing and processing information in the right sort of way. But a complicated network of problems emerges when we try to clarify just what "the right sort of way" is. The central aim of this project is to target and (I hope) solve many of these problems, thereby elucidating the epistemic dimensions of moral responsibility and respect.

1.2 Questions about Responsibility

Let me make the problems more vivid by pointing to some of the cases and questions that animate the broader project. I begin with questions about the epistemic dimensions of moral responsibility, and then I target the questions about respect.

It is intuitive that ignorance sometimes counts as a legitimate excuse for behavior with bad consequences. But when and why, precisely, does ignorance excuse? If I know more or less what I am doing but fail to know that that sort of thing is morally wrong, can such fundamental moral ignorance ever be an excuse for my wrongdoing? Under what circumstances? Must one *know*, at some level, that one's act is wrong to be responsible for one's wrongdoing? And what about responsibility for morally right action? Must one be in any sense *aware* of the facts that make one's action right to be responsible for the right action?

The above questions concern the relationship between one's awareness of or beliefs about one's own action, on one hand, and one's moral responsibility for that action, on the other. Such questions ultimately prompt another set of questions about responsibility not for actions but for various mental states, like ignorance and objectionable attitudes. For, as many see it, I am responsible for unwitting wrongdoing only if I am responsible for the ignorance from which my wrongdoing stems. Is such a view correct? And what, if anything, ever makes me responsible for ignorance? More

generally, when am I morally responsible for a belief or other attitude? Is it only indirectly, because of some causal connection between the attitude and some prior voluntary choices that led to its formation (I voluntarily chose not to read some literal or metaphorical warning label)? Or can I be directly responsible for an attitude because of the kind of relationship it has to my values, irrespective of facts about my voluntary choices? Can one's epistemic circumstances—because of the society one happens to live in, say—play some role in making one's moral ignorance or other objectionable attitudes blameless? Why, if ever?

At the largest level, we may summarize our question about the epistemic dimensions of moral responsibility as follows: how do facts about how one has (or hasn't) accessed and processed information bear upon whether one is morally responsible for one's action or attitude? Answering this question should tell us something about the general nature of moral responsibility and about how easy (or difficult) it is to satisfy its conditions. It should also tell us something about what our responsibility-attributing practices ought to look like. For, however we answer this question, we should wonder how well-positioned we are to make warranted judgments in particular cases that some person has accessed and processed information in the way that would make him responsible. Thus, our question about the epistemic dimensions of moral responsibility has wide-reaching theoretical and practical significance.

1.3 Questions about Respect

Let us now consider some of the questions about epistemic requirements on *respect*. At first glance, it may seem that we are entering a new and unrelated conversation. However, as it turns out, the literature on some of the above questions

about responsibility for attitudes directly prompts questions about the nature of respectful judgment. Here is how.

To many, it seems both plausible and compassionate to believe that a person's unfortunate epistemic circumstances can make some of his moral ignorance and objectionable attitudes blameless (or at least less blameworthy). But several philosophers (Michele Moody-Adams (1993), Christopher Bennett (2004), Angela Smith (2005, 2008), and others) have argued that it is dangerously patronizing—and, ultimately, disrespectful—to undercut an agent's relationship to her own attitude by excusing her on such grounds (consider: "you can't blame Donna for her beliefs about abortion—that's just how she was raised"). Indeed, these philosophers take this point to generate a practical argument against theories of responsibility according to which sane agents can rightfully be excused for their moral ignorance and other objectionable attitudes. Such theories, it is suggested, are in tension with basic requirements of respect and so ought to be rejected. Thus, as the dialectic has unfolded, figuring out whether/why an agent can be responsible for ignorance requires resolving questions about what respect requires in our responsibility-attributions.[3] And to resolve *those* questions, we need some sense of what respect requires in our judgments about persons more generally.

It is surprising, then, that work on responsibility for attitudes has had very little if anything to say about what makes a judgment respectful or disrespectful. What *does* respect require in our judgments about persons? Implicit in the above charge about

[3]That practical concerns might bear upon theoretical questions about responsibility has been a popular suggestion at least since P.F. Strawson's famous essay, "Freedom and Resentment."

disrespectful excusings is the thought that respect in judgment is fundamentally about not undercutting agents in various respects: unless we have strong evidence that some person is literally incapable of responding to reasons, respect requires that we not judge that the person had some diminished level of agency over their belief, even if the belief is objectionable.

As it happens, this suggestion strongly parallels a popular view in the literature concerning "epistemic respect," or respect for persons in their roles as knowers. In her influential book, *Epistemic Injustice*, Miranda Fricker argues that while I might disrespect you *qua* knower by giving you too little credibility, I can never disrespect you *qua* knower by giving you *more* credibility than is warranted by my evidence (2007: 20). These two views—about what respect requires in our responsibility attributions and in our credibility judgments—share a common idea. According to both, respect does not centrally involve proportioning one's judgments about persons to the evidence one has about them in epistemically good ways. Rather, respect simply places negative requirements on our judgments, commanding us not to *under*sell agents in various respects. On this view, respect requires that one attribute *enough* agency to another over her belief, that one give another *enough* credibility, and so on.

I believe that this targeting of respect is importantly misguided. What it misses, I believe, are the epistemic dimensions of respect. Respect of one kind—in particular, the kind antithetical to alienation and objectification—requires not esteem, trust, or praise; it

requires that one's judgments be properly responsive to persons as the particular agents they are.[4] Consider.

In some cases, I may be perfectly well justified in believing that, though Donna's objectionable belief about (say) eating meat was arrived at rationally, the belief does not reflect some stable, evaluative judgment about what is of value. The belief might be better explained by the fact that, relative to Donna's limited evidence, Donna had good, internal reasons to focus her moral attention elsewhere than on animal ethics. In such a case, there is no sense in which it is disrespectful when I judge, in accordance with my evidence, that the objectionable features of Donna's belief do not reflect stable, objectionable values (in the way that might make her morally responsible for the objectionable features of her belief). When my judgment is properly responsive to the information I have about Donna as an agent, there is no failure of respect, even if this judgment implies that there are some features of her belief over which she has some diminished level of agency (features that are better explained by the society she happens to live in rather than by her values). Moreover, when my judgment *fails* to be properly responsive to an agent, this marks an important failure of respect, even if this failure of proper responsiveness involves attributing something apparently positive to the agent. *Contra* Fricker, if I give you excessive credibility because some "positive prejudice" about your social identity makes me resistant to incoming evidence about you as the particular agent you are, my judgment about you exhibits a significant form of alienation

[4]This general targeting of respect sits in the tradition that includes Stephen Darwall's (1977) discussion of "recognition respect" and Robin Dillon's (1992) discussion of "care respect," which she takes to be a species of Darwall's "recognition respect" (Dillon, 1992: 112).

that is incompatible with respect. Collectively, these cases show that whether some judgment about a person is (dis)respectful is, in part, a function of whether the judgment meets some epistemic standard(s).

So, at the largest level, we may state our question about the epistemic dimensions of respect as follows: how do facts about how one has (or hasn't) accessed and processed information bear upon whether one's judgment is (dis)respectful? I suggested above that respect requires, in part, that a judgment satisfy some epistemic standard. But what is the standard? And, presumably, a respectful judgment ought to satisfy some *moral* standard. What standard? And how ought the satisfactions of the epistemic and moral standards relate?

1.4 The Plan

Here is the plan for this project. Above, I tried to briefly motivate the thought that responsibility and respect both importantly involve an agent's accessing and processing information in the right way, though things get complicated when we try to identify "the right way" for each. The dissertation proceeds in two stages, with the aims of (first) motivating nontrivial epistemic requirements on each and on (second) revealing what these requirements are. While each chapter is written to stand in isolation, the chapters jointly work toward explicating the epistemic dimensions of moral responsibility and respect. In what follows, I preview the central arguments taken up in each chapter.

In Part I ("Vindicating Epistemic Requirements on Responsibility"), my central aim is to make epistemic requirements on responsibility less scary. A potential reason that epistemic requirements on moral responsibility have not received their due theoretical development is that many philosophers have suggested that robust epistemic

requirements on responsibility would have disastrous implications for our responsibility attributing practices. So, here, I reject two popular worries about epistemic requirements on responsibility, one about skepticism, the other about respect.

The first worry is that epistemic requirements on responsibility would yield a significant form of responsibility skepticism. To show that this worry is misplaced, I show that even on an excessively demanding conception of the epistemic requirements on responsibility, skepticism does not follow.

Here is the skeptical worry. According to Gideon Rosen (2004), Michael Zimmerman (1997), and others, the epistemic requirement on blameworthiness is full-blown akrasia: an agent is blameworthy for an action only if the agent knows at the time of performing the action that it is wrong. From this requirement, these writers argue that we can never justifiedly attribute blameworthiness: genuine akrasia is too hard to identify in the wild. Now, Rosen and Zimmerman are happy to embrace this form of skepticism. But, even many of those who believe that Rosen and Zimmerman are mistaken about the akrasia requirement grant, as though it were trivially true, that *were* there an akrasia requirement on blameworthiness, this skeptical conclusion would follow (see Harman, 2011: 443). I believe that acceptance of this fast argument for skepticism has derailed theorizing on the epistemic requirements on responsibility. In "Skepticism About Skepticism About Moral Responsibility" (Chapter 2), I show that this argument for responsibility skepticism rests on implausible epistemological principles that, if true, would generate broad skepticism about phenomena that involve causal-historical facts about mental states. Moreover, I show how, with proper background knowledge, we *can* justifiedly attribute blameworthiness, even granting this (excessively) demanding

epistemic requirement. Thus, fears of skepticism should not deter us from taking seriously the thought that responsibility has some robust epistemic requirements.

A second popular worry for robust epistemic requirements on responsibility is that they might invite us to disrespectfully excuse persons for their moral ignorance and objectionable attitudes. As mentioned above, several philosophers bring our attention to cases in which excusing an agent for her attitude is patronizing rather than compassionate. Focusing on such cases, it is argued that it is *always* disrespectful to excuse a sane agent for her sincerely held objectionable beliefs. So, the argument continues, any theory that licenses excusing rational agents for their attitudes—as any theory with robust epistemic requirements inevitably would—has a considerable mark against it.

In "When and Why is it Disrespectful to Excuse an Attitude?" (Chapter 3), I show that this argument relies on false generalizations, both about what is involved in excusing an attitude and about the nature of respect. *Contra* Smith, Bennett, Moody-Adams, and others, we can excuse someone for her attitude without thereby implying that she is literally incapable of responding to reasons and without implying that the particular attitude was produced without his having responded to reasons. I then sketch an account of respect (something conspicuously missing in the literature arguing that it is disrespectful to excuse an attitude) and use it to explain when and why it is disrespectful to excuse an attitude. The central idea to the account (more fully developed in Part II) is that respect requires that one's judgments bearing upon another *qua* rational agent meet some basic moral and epistemic standards: respect requires that we be properly responsive to persons. Using this account, I show that one can coherently (and

respectfully) excuse an attitude *even* in some cases where that attitude was produced by a responsiveness to reasons. Thus, fears about disrespectful excusings should not deter us from taking seriously the thought that responsibility has some robust epistemic requirements (of the sort that could make some moral ignorance and objectionable attitudes blameless).

Part I rejects worries about robust epistemic requirements on responsibility and gestures toward a conception of respect that emphasizes the significance of proper responsiveness to persons. The aim of Part II ("The Value-Secured Reliability Theory") is to make good on the promise to precisify "the right way" of accessing and processing information for moral responsibility and respect. I argue that respect and moral responsibility for right action (or moral worth) have parallel structures. Each is the manifestation of a moral competence, where this manifestation centrally involves two components: namely, 1) a performance that is reliably tied to some good, 2) the reliability of which is sufficiently explained by one's values. I argue for a "Value-Secured Reliability Theory" of both moral worth and respect, I show how this theory targets indispensable (and realistic) epistemic requirements, and I show how the theory can help us resolve debates about the ethics and epistemology of testimony and about the relationship between consciousness and responsibility.

The first essay in Part II is "Moral Worth and Consciousness: In Defense of a Value-Secured Reliability Theory" (Chapter 4). The central question it pursues is: what minimal role—if any—must consciousness of morally significant information play in an account of moral worth? According to one popular view, a right action is attributable to an agent and non-accidentally right as moral worth requires only when the agent is

conscious of the facts that make it right. I argue against this consciousness condition. As I show, consciousness of such facts requires much more sophistication than writers typically suggest—this condition would bar from moral worth most ordinary, intuitively morally worthy agents. Moreover, I show that satisfying this condition cannot play a significant role in securing non-accidentality anyhow, and it is not necessary for either attributability or non-accidentality.

Drawing some lessons from these problems, I introduce and defend the Value-Secured Reliability Theory of moral worth and I show how a minimal yet indispensable role for consciousness falls out from it. According to this theory, an agent's right action is morally worthy just to the extent that it is explained by the agent's value-secured reliable tie to the right, a reliable tie to the right that is secured through the influence that her person-level values have (perhaps unreflectively) on her patterns of informational access and processing. I show how, on this independently plausible theory, an action can be morally worthy even when the agent is unaware that her act is right and unaware of the facts in virtue of which her action is right. An act can be explained by a value-secured reliable tie to the right even when one has accessed and processed only quite minimal sensory cues (rather than *the facts that make one's act right*). If it is successful, the chapter supplies a compelling theory of moral worth that plainly identifies its epistemic requirements.

The final chapter, "Epistemic Respect and Credibility Excess" (Chapter 5), applies the value-secured reliability framework to *respect* and shows how the theory can explain a range of disrespectful credibility judgments that popular views in the ethics and epistemology of credibility cannot. On the dominant perspective from this literature,

credibility excess (giving someone too much credibility) either never wrongs persons *qua* knowers or, if it ever does, it is because it renders them epistemically arrogant. I show that this seriously miscaptures the potential significance of credibility excess. In cases where a person grants another excessive credibility due to "positive prejudice" toward her social identity, to self-absorption, or to some other vice, the judgment is alienating, objectifying, and, ultimately, lacking in a basic form of respect. Drawing some lessons from the cases of disrespectful credibility excess, I introduce and develop the Value-Secured Reliability Theory of respect, both of "epistemic respect"—or respect for someone in the localized role of knower—and respect more generally.

What we see from the cases of disrespectful credibility excess is that respect requires that our judgments bearing upon persons as rational agents be properly responsive to them. According to the Value-Secured Reliability Theory of respect, respect is a manifestation of a moral competence. A judgment about a person is respectful just in case it is reliably produced *because* of the influence one's value or care for the person has on how one responds to information form one's environment. This captures at a theoretical level what should be an intuitive picture of respect: respect requires that our judgments about persons be reasonably attuned to them because we value them. I argue that this account provides satisfying underpinnings for a variety of intuitive claims about respect and disrespect, and it illuminates an underappreciated epistemic dimension of respect.

In the brief conclusion, I summarize some of the key points I defend over the course of the dissertation, and I point to future problems that the Value-Secured Reliability framework can illuminate.

PART I: VINDICATING EPISTEMIC REQUIREMENTS ON RESPONSIBILITY

CHAPTER 2

SKEPTICISM ABOUT SKEPTICISM ABOUT MORAL RESPONSIBILITY[5]

2.1 Introduction

In a wonderfully compelling and clear paper, Gideon Rosen (2004) poses a novel skeptical challenge for conventional wisdom about moral responsibility. The challenge is a welcome departure from familiar discussions of free will and determinism—he does not aim to show, for instance, that no one is ever (or ever could be) morally responsible. He aims to show, rather, that "in any particular case in which we may be tempted to judge that X is responsible for doing A, this particular judgment of responsibility would be unwarranted" (Rosen, 2004: 295). Put another way, his argument is that while (for all we know) blameworthy action may be quite common, we are never epistemically justified, in any particular case, in believing that so-and-so is morally blameworthy for her action.[6] Thus, the title of his paper, "Skepticism about Moral Responsibility," is a bit misleading: Rosen is not skeptical about the existence of moral responsibility (or blameworthiness)—he is skeptical about the existence of epistemically justified attributions of blameworthiness. Rosen's paper gives us all good reason to recalibrate the confidence with which we blame ourselves and others. However, I shall argue that Rosen has not

[5]A version of this paper appears in *Pacific Philosophical Quarterly* (2018).

[6]Rosen clarifies that when he says "moral responsibility," he is referring to blameworthiness (see Rosen, 2004: 296). In fact, one point upon which Rosen insists (and one on which I will push back) is that his argument applies only against judgments of blameworthiness and not against judgments of praiseworthiness.

adequately defended his full-blown skeptical thesis and that, indeed, we ought to reject his thesis.

Before I spell out more precisely the nature of my argument, it will be helpful to contextualize my overall approach in responding to Rosen by highlighting a specific passage from his essay. Aware of the fact that he has offered what would strike many as a hopelessly radical skeptical argument, Rosen attempts to distance his position from familiar unlivable skeptical positions:

> There is a perennial tendency in philosophy to suppose that skeptical arguments can amount at best to puzzles, and this may be true when it comes to arguments that purport to show that confident judgments about the external world, or about other minds, or about the future, are never warranted. That is in part because in these cases the skeptical posture at issue—e.g., thoroughgoing suspense of judgment about the external world—is not a serious option for us. Strawson famously seeks to assimilate skepticism (or rejectionism) about responsibility to such a case. His central thought, for present purposes, is that genuine suspense of judgment on questions of responsibility would amount to adopting the "objective stance" towards other human beings—a stance in which one prescinds from a vast range of ordinary emotional responses to human action and treats other people as one normally treats animals or lunatics—as entities that require management or treatment rather than ordinary interaction. Strawson rightly argues that this revision is not only undesirable; it is impossible. And if this were the inevitable upshot of the skeptical argument, this might be enough to warrant assimilating skepticism about moral responsibility to philosophical skepticism in other areas (Rosen, 2004: 310-311).

To fend off these Strawsonian worries, Rosen quickly suggests that his argument is still "compatible with the thought that the *positive* reactive attitudes—gratitude, pride, admiration, and the like—are for the most part warranted [in an epistemic sense]," and he concludes that there is "thus ample scope within the skeptical posture here endorsed for a distinction between the attitude we adopt with animals and lunatics and an engaged stance we might adopt with normal competent adults" (Rosen, 2004: 311). The suggestion is that his argument does not cast doubt on the possibility of epistemically

justified attributions of *praiseworthiness* and that, this being the case, his position is not as radical as all that. Certainly, he would suggest, his skepticism is very different from skepticism about, say, the external world.

I hope to show that Rosen's position is hostage to a skepticism that is much more radical than he recognizes (and much more radical than anyone should be prepared to endorse). In particular, I will show, first, that Rosen's skeptical thesis about epistemically warranted attributions of blameworthiness rests on assumptions which *would* (contra Rosen) similarly generate a skeptical argument against the existence of warranted attributions of *praiseworthiness*. If I am right about this, then Rosen's skepticism is significantly more radical than he takes it to be, and Strawson's worry about the livability of such skepticism is live again. I will show, second, that Rosen's argument rests on assumptions which also generate further skeptical theses outside of the domain of moral responsibility, where skepticism in those domains is intuitively unacceptable. More specifically, I show that underlying Rosen's argument are assumptions which would imply that no one is ever justified, in any particular case, in believing that some belief is doxastically justified. The bulk of my paper attempts to show that Rosen's position is hostage to these radical and broad forms of skepticism. If I am right about this, then, by Rosen's own lights, there are strong reasons to reject the thesis. In the brief final section of the paper, I rely on a discussion of the conditions under which one can be justified in believing that some belief is doxastically justified to help motivate a picture on which we *can*, in fact, make some epistemically justified attributions of blameworthiness. My thesis is, thus, perhaps best conceived of as a disjunction: either A) Rosen's skeptical thesis is correct, but he has not offered a problem specifically for attributions of blameworthiness,

for the key assumptions in his argument generate skeptical conclusions in a broad range of areas, or B) Rosen has not shown a genuine problem for justified attributions of blameworthiness (at all).

In Section 2, I will briefly present and explain Rosen's argument. Section 3 will show, contra Rosen, that his argument applies equally well to attributions of *praiseworthiness* and is, thus, considerably more radical than he takes it to be. Section 4 continues this theme by showing that his argument rests on assumptions that would also generate an unacceptable skepticism about the existence of justified beliefs about particular beliefs being doxastically justified. Finally, Section 5 expands upon the discussion of beliefs about doxastic justification to offer some positive reasons to think that, under certain conditions, we *can* make warranted attributions of blameworthiness.

2.2 Rosen’s Argument

Rosen's argument comes in two stages. The first stage of his argument identifies a necessary condition for blameworthy acts. The second stage suggests that since no one is ever justified in believing, in any particular case, that the specified necessary condition is satisfied, no one is ever justified in believing that an act is blameworthy. While I will be challenging only the second stage of Rosen's argument, it will be helpful to say something about the first. I briefly explain the two stages of Rosen's argument in what follows. My focus is on the second stage of Rosen's argument, and I will not put pressure on the controversial parts of the first stage.

2.2.1 A Necessary Condition of Blameworthy Acts

The first stage of Rosen's argument concludes that an act is blameworthy *only* if it is an akratic act or is the upshot of some prior akratic act—"every culpable bad action must be the causal upshot of a genuinely akratic act or omission," says Rosen (2004: 307). Here is how he comes to this conclusion. Take any morally wrong act. Either that act was performed in ignorance of some morally relevant facts or it was performed without ignorance of any morally relevant facts. In the latter case, the act is, of course, akratic.[7] Thus, any blameworthy act in this latter camp will, by default, be akratic (since all of the acts in this camp are akratic). The interesting part of Rosen's argument says that, among the bad acts performed in *ignorance* of some morally relevant facts, only those that are the upshot of an akratic act are blameworthy. If Rosen is right on this point, it will follow that all blameworthy acts are the upshot of an akratic act. So, let us look at his argument about wrong acts performed in ignorance.

Rosen uses a simple pair of examples to help motivate the following principle:

> *If X does A from ignorance, then X is culpable for the act only if he is culpable for the ignorance from which he acts* (Rosen, 2004: 300).

Suppose Alex1 asks Barry1 for sugar in his tea. Unfortunately, a malicious stranger broke into Barry1's home the night before, put arsenic in Barry1's sugar bowl, and covered up all evidence of his break in. Barry1 puts a spoonful of the white powder from the sugar bowl into Alex1's cup of tea, ignorant of the fact that it is arsenic in the sugar bowl, and

[7] I will follow Rosen's way of understanding 'akrasia', where an akratic act is an act such that the agent knows the pertinent facts about his act, knows that his act is wrong, and knows that in the circumstances, all things considered, he should not do it (see Rosen, 2004: 307).

Alex1 dies from poisoning. Alternatively, suppose Alex2 asks Barry2 for sugar in his tea. The case begins in parallel fashion to the first: Barry2 puts a spoonful of the white powder from the sugar bowl into Alex2's cup of tea, ignorant of the fact that it is arsenic in the sugar bowl, and Alex2 dies from poisoning. However, suppose Barry2 keeps both sugar and arsenic around the house in very similar looking and unlabeled bowls. It should be clear that Barry1's ignorance about the substance in the sugar bowl was not his fault and that his ignorance counts as a legitimate excuse for his bad action—Barry1 is not blameworthy for poisoning Alex1. But, it should also be clear that Barry2's ignorance is no excuse for his bad action (at least, as the case is currently described). Even though Barry2 was ignorant of the facts in virtue of which his act was wrong, it was Barry2's *fault* that he was ignorant in this way—Barry2 may well be blameworthy for poisoning Alex2. Rosen concludes from cases like these that an act done from ignorance is blameworthy only if the ignorance is itself blameworthy.

Having established the principle above, Rosen next has to say something about when *ignorance* is blameworthy. He arrives at the following principle:

> *X is culpable for failing to know that P only if his ignorance is the upshot of some prior culpable act or omission* (Rosen, 2004: 301).

The culpable acts or omissions Rosen has in mind have to do with what he calls our "procedural epistemic obligations." Roughly, one's procedural epistemic obligations are one's moral obligations to see to it that she is sufficiently informed about her situation so as not to negligently harm herself or others. Rosen provides the following heuristic for determining whether someone has abided by his procedural epistemic obligation: "in any given case we can ask whether the agent's ignorance derives from a failure to do what any reasonably prudent person in his circumstances would have done in order to see to it that

he was adequately informed" (Rosen, 2004: 301). From the examples above, Barry2 has, of course, failed to abide by his procedural epistemic obligations; his ignorance of the fact that he was posing a deadly threat to Alex2 was the upshot of his prior culpable act or omission—perhaps his failure to make his arsenic and sugar bowls sufficiently distinguishable.[8] Alternatively, Barry1's ignorance was not the upshot of some prior culpable act or omission—we can stipulate that Barry1 took all of the precautions a reasonably prudent person would have taken before putting the arsenic in Alex1's tea. When one has failed to abide by her procedural epistemic obligation, then, supposing that her failure was not the result of some further ignorance, she may well be culpable for her ignorance and, in turn, for the act performed in that ignorance.

Once Rosen has established both that an act done from ignorance is blameworthy only if the ignorance is blameworthy and that ignorance is blameworthy only if it is the upshot of a blameworthy act or omission, it is perhaps clear how akrasia will enter the picture. Consider the act or omission that led to Barry2's ignorance of the fact that his act was wrong. Barry2 failed to make his arsenic and sugar sufficiently distinguishable, and this led to his ignorance of the fact that he was putting arsenic in Alex2's tea (and, one might think, to his ignorance of the fact that this was *wrong*). But suppose his failure to make his arsenic and sugar bowls sufficiently distinguishable was *itself* an act (or omission) from ignorance. Perhaps Barry2 was ignorant of the fact that arsenic is poisonous, and his ignorance on this matter partially explains why he made no effort to make his sugar and arsenic bowls readily distinguishable. Since his failure to make the

[8]Of course, perhaps this failure was, in turn, the product of some ignorance. This will be addressed in just a moment.

bowls distinguishable was 'performed' from ignorance, we now must ask whether this further ignorance is blameworthy in order to determine whether he is blameworthy for failing to make his bowls readily distinguishable. To Rosen, “this looks like a recipe for a regress” (Rosen, 2004: 303). To terminate the regress in such a way that Barry2 is blameworthy for poisoning Alex2, it will have to be the case that somewhere in the causal chain leading to Barry2's poisoning Alex2 is a bad act or omission that is not itself an act from ignorance—and such an act, being that it is not done from ignorance, would be akratic.

So, by appealing to two independently plausible principles about action done from ignorance, Rosen takes himself to have shown that an act is blameworthy *only* if it is either akratic or the upshot of some akratic act. I will not challenge this stage of his argument. Before moving to the skeptical stage of Rosen's argument, it is worth noting how broad (on Rosen's view) the range of cases of ignorance that can potentially excuse is (a discussion of the range will help to reinforce why Rosen thinks genuine *akrasia* is required for blameworthiness). It is not merely ignorance of, say, descriptive facts about one's environment (like Barry1's ignorance of the fact that the powder in the sugar bowl is arsenic) that can excuse. Rosen also suggest that there are many cases of exculpatory ignorance of some *moral* fact (that A is wrong) where this ignorance is not rooted in ignorance of some descriptive fact about the world.[9] He further suggests that there can be cases of exculpatory *normative* ignorance—in such cases, an agent knows that her act is

[9]See Rosen (2004: 304; 2003: Section IV). He considers the ancient slave holder in a society wherein it was taken for granted by everybody (including the slaves and all the smart people the slave holder knows) that, say, beating a slave is morally permissible.

morally wrong but she is ignorant of the fact that her moral reasons outweigh, say, her self-interested reasons.[10] If we can be ignorant of moral and normative facts while still abiding by our procedural epistemic obligations (and Rosen thinks we can), then such forms of ignorance can excuse in the same way that ignorance of descriptive facts about one's environment can excuse.[11] Ignorance of descriptive, moral, or normative facts about one's act prevent one from knowing that she has decisive reason not to be doing what she is doing. On Rosen's view, genuinely blameworthy behavior must be (or be the upshot of) a bad act that is performed in full knowledge of the fact that one has decisive reason to be doing something other than what she is doing—that is, it must be (or be the upshot of) an akratic act.

2.2.2 The Skeptical Argument

The portion of Rosen's argument that argues for akrasia as a necessary condition of blameworthy behavior takes up the vast majority of his essay. It is only in the brief final section of the essay that Rosen moves to the skeptical stage of his argument. Here, Rosen summarizes:

> The culpable bad actions have a distinctive sort of causal history—an *inculpating* history—in which the act either is, or derives from, an episode of genuine akrasia. There may be no obstacle in principle to identifying such episodes. God could manage it; a super-psychologist might manage it. But given the real limitations on our access to the causal histories of human actions and to the states of knowledge and opinion that underlie them, I claim that as a matter of fact we are never

---

[10]See Rosen, 2004: 305. He considers Bill, who, knowing it would be morally wrong for him to tell the self-serving lie in his particular situation, tells the lie while mistakenly believing (as a result of some peculiar education) that *this* is one of those times when self-interested reasons outweigh the moral reasons.

[11]For an argument that moral ignorance does not exculpate, see Harman (2011).

> entitled to any significant confidence that the bad act under consideration satisfies the necessary condition we have identified (Rosen, 2004: 309).

Another passage may help to reinforce Rosen's worry:

> What must you think in order to judge that Bill, for example, is responsible for lying to his wife? You must think that at the time of action, either he knew that he had decisive reason not to lie, or if he did not know this, that his ignorance was the upshot of some prior bad action done in full knowledge of every pertinent fact or norm. You must think, in other words, that his bad action either is, or derives from, an episode of genuine, full-strength akrasia (Rosen, 2004: 308).

In order to have a warranted belief that such-and-such act is blameworthy, one must have a warranted belief that the act is either akratic or the upshot of some prior akratic act. But *no one* ever has a warranted belief that such-and-such act is either akratic or the upshot of some prior akratic act, and, thus, no one ever has a warranted belief that such-and-such act is blameworthy.

Why is it, on Rosen's view, that no one is ever justified in believing that such-and-such is either akratic or the upshot of some prior akratic act? Simply put, we are just no good at distinguishing truly akratic behavior from various close cousins of akratic behavior. As Rosen puts it:

> The agent is culpable for his bad action only if that bad action is, or derives from, an episode of genuine akrasia. But genuine akrasia in this sense is extremely difficult to identify. The reason is that it is not readily distinguishable from an impostor: ordinary weakness of will (Rosen, 2004: 309).

In a case of non-akratic 'ordinary weakness of will' (as Rosen understands it), an agent might know that A is wrong at time t1, become momentarily persuaded by t2 (the time of her performance of A) that it is okay to perform A in this case, and return to knowing that A is wrong by t3 (see Rosen, 2004: 309). In another iteration, the agent might—at the crucial moment of performance—*suspend judgment* about whether A is the thing to do. The worry is that genuinely akratic acts will *look* and *feel* virtually identical to these non-

akratic varieties of weakness of will. “Given the opacity of mind” (308), we should believe that the limited evidence available to us when we encounter wrongdoing (whether one's own or somebody else's) is insufficient to justify an attribution of akrasia or akratically derived action.

Rosen's case is purely anecdotal and “based mainly on reflection” (308): he writes that he cannot, with any confidence, identify as akratic any act of his own or of any close acquaintance. There is no explicit appeal to any general epistemological principle in the argument, but his remarks suggest the following two principles. First,

> *DISTINGUISH: If Z is the distinguishing feature between Xs and Ys (Xs and Ys are alike except Xs have feature Z and Ys do not have feature Z) and it is sufficiently plausible that A could be either an X or a Y, then in order to justifiedly believe that A is an X, one must justifiedly believe that A has feature Z.*

In many cases of wrongdoing, the hypotheses *that the wrongdoing was genuinely akratic* and (alternatively) *that the wrongdoing was non-akratically weak-willed* will each be (alone) a sufficiently plausible[12] hypothesis relative to the available evidence. The distinguishing feature between the akratic and the non-akratically weak-willed is that the former involves the normative belief (at the time of performance) that one ought to be doing otherwise. DISTINGUISH tells us, then, that in order to justifiedly believe that such-and-such action is blameworthy, one must justifiedly believe that the action is akratic or akratically derived (as opposed to non-akratically weak-willed), and to

---

[12]I have nothing very precise to say about *how* sufficiently plausible is sufficiently plausible. To be sufficiently plausible, the hypothesis must (of course) at least be *consistent* with the available evidence, but the standard of sufficiency should probably be a bit more restrictive. In the typical case, the hypothesis that *that thing is a barn facade* is not sufficiently plausible (in the sense intended), so DISTINGUISH does not require that I justifiedly believe that *that thing has the distinguishing feature between barns and barn facades* in order to justifiedly believe that *that thing is a barn*.

justifiedly believe that the action is akratic or akratically derived (as opposed to non-akratically weak-willed), one must justifiedly believe that the agent held the normative belief (at the time of performance) that she ought to be doing otherwise (see Rosen's comment above on what one must think in order to judge that Bill the liar is responsible).

Rosen's central claims to motivate his skeptical worry—his claims that genuine akrasia is not "readily distinguishable" (309) from ordinary weakness of will and that we do not have sufficient "access" (309) to the relevant causal-historical facts about mental states—suggest a second principle. The idea seems to be something like the following:

> *DETECTION: When an agent's A-ing has no perceptually or introspectively detectable features which would reliably indicate that fact F about her mental states obtains, one cannot justifiedly believe that F is true of the agent at the time of her A-ing.*

If we set aside full-blown skepticism about mental states, it seems clear that lots of actions have detectable features which are reliable indicators of various facts about the agent's mental states. When (say) Derrick angrily argues with someone, there may be some pretty reliable indicators of the fact that he acts in anger (his raised voice, his facial expression, et cetera) which a third-party can pick up on through perception, and the *feeling* associated with one's own anger is something which agents can (except in rare cases of self-deception) generally pick up on through introspection. There is a reasonably distinctive *look* and *feel* to acts in anger. Thus, DETECTION does not give rise to a skeptical argument against justified attributions of anger. But, whether or not you held the relevant normative belief at the time of your action which would be necessary for your akrasia and blameworthiness (or whether or not your action was the causal upshot of some prior akratic act), your action may *look* the same to me—your action may have no detectable features which would reliably indicate to a third-party that you held the

relevant normative belief at the time of action. Moreover, in the first-person case, the agent cannot introspect and *see* that she has (or had, as the case may be) the relevant normative belief at the relevant time (309). God or a super-psychologist (what with their access to our whole stock of beliefs at various moments before, during, and after our acts) may be able to identify the akratic acts out there, but we mere mortals simply do not have the resources to pick them out with any warranted confidence.

In appealing to principles along the lines of DISTINGUISH and DETECTION, Rosen is able to get his skeptical argument off the ground. The thought is that justifiedly attributing blameworthiness requires making a justified judgment which distinguishes between two options which *look* and *feel* indistinguishable from each other, and, as Rosen sees it, this requirement makes justified attributions of blameworthiness impossible—we simply cannot make the justified judgments about the 'causal histories of human actions and … the states of knowledge and opinion that underlie them' which we would need to make in order to justifiedly attribute blameworthiness.

So much for the skeptical argument. In Sections 3 and 4, I show that if *this* is all it takes to generate Rosen's skeptical argument, then we can mount parallel arguments against justified attributions of praiseworthiness and against justified attributions of doxastic justification. Section 5 will give us reason to reconsider the above epistemological principles.

2.3 Skepticism about Judgments of Praiseworthiness

Recall that Rosen attempts to downplay how radical his skeptical thesis is by suggesting that his thesis is fully compatible with the thought that attributions of *praiseworthiness* are often epistemically justified. In this section, I will suggest that the

same considerations Rosen uses to cast doubt upon the existence of justified attributions of blameworthiness would (if true) equally cast doubt upon the existence of justified attributions of praiseworthiness. If I am correct, then Rosen's skepticism is considerably less modest than he takes it to be, and we have grounds for doubting, with Strawson (and by Rosen's own lights), that this skepticism is a live option for us. In what follows, I identify, in schematic form, a necessary condition which any plausible account of praiseworthiness must include. I then suggest that Rosen's epistemological principles and assumptions will imply that no one is ever justified in believing that this condition holds and, in turn, will imply that no one is ever justified in believing of an act that it is praiseworthy.

Here is the stipulated necessary condition: X's act A is praiseworthy only if A is caused *in the right way* by X's having *the right kind* of motivation to perform A. I do not aim, here, to flesh out the details underlying being caused 'in the right way' and being a motivation 'of the right kind,' in part because I want the discussion to remain neutral on the question of what, precisely, makes for praiseworthiness. But it seems uncontroversial that any account of praiseworthiness will have to include at least these kinds of causal-historical facts about motivation. However, as I will illustrate, if we accept Rosen's implicit epistemological principles and assumptions, it will follow that we can never be justified in believing of some act that it satisfies the above necessary condition (and, thus, on Rosen's view, we cannot be justified in believing that it is praiseworthy).

My Rosenian skeptical argument against warranted attributions of praiseworthiness will turn on the difficulty of ruling out the possibility that when X performed A, either 1) X did not have the right kind of motivation to perform A or 2) X's

motivation (though it was of the right kind) did not cause A in the right way. I introduce the argument by way of a familiar example. Consider Donna. Donna is invited to a party that is happening next week. She knows that the guests at the party speak well of those in her socioeconomic class who donate to humanitarian organizations and poorly of those in her socioeconomic class who do not. A few days before the party, Donna donates to Oxfam.

Now, one way we can fill in the details of Donna's case is like this. Before donating, Donna considers whether to donate. She is “put in touch with” and comes to understand the force of the “moral reasons” (whatever they are) to donate, and these reasons move her to donate (which she does). Let us call this Case 1, and let us stipulate of Case 1 that *this* is a paradigm case of an agent's act being caused in the right way by the agent's having the motivation of the right kind. In Case 1, Donna is praiseworthy for donating.

Here is a different way the details of Donna's case can be filled out. Before donating, Donna considers whether to donate. She knows, in some abstract sense, that there are compelling moral reasons to donate. Indeed, she's known for quite some time that there are moral reasons she should donate money to Oxfam, but these reasons have never moved her to donate—Donna is just not that kind of person. She ends up donating, but simply because it is in her self-interest, what with the upcoming party. Let us call this Case 2, and let us stipulate that *this* is a paradigm case of an agent's act *not* being caused in the right way by the agent's having the motivation of the right kind. We might say that Donna has, in some weak sense, the right kind of motivation, but that her act is not caused by this motivation in the right way (for, her act was *primarily* or *really* caused by

her motivation to do what is in her self-interest), or we might say, more intuitively, that Donna just did not have the right kind of motivation in the first place. I find it intuitive to say that Donna is not praiseworthy for donating in Case 2. We might well suppose that she would never have donated had it not been in her self-interest—it is, in a relevant sense, a mere coincidence that her act so closely resembles a praiseworthy act.

Now, suppose a third-party observer, Thurston, were trying to determine whether Donna is praiseworthy for donating. Using reasoning parallel to that found in Rosen's skeptical argument against justified attributions of blameworthiness, we can say that if Thurston's judgment that Donna is praiseworthy is to be justified, Thurston would have to justifiedly judge that Donna's donating was caused in the right way by the right kind of motivation (see DISTINGUISH). But, of course, whether or not Donna's action was caused in the right way by the right kind of motivation, her act of donating may *look* exactly the same to Thurston. If all Thurston has to go on in making his judgment is his isolated observation of Donna donating, then he will have no more reason to believe that something like Case 1 is the accurate story than that something like Case 2 is the accurate story—he will not be able to justifiedly believe that Case 1 is the accurate story (see DETECTION). So, according to Rosen's implicit epistemological principles, Thurston cannot possibly be justified in believing that Donna is praiseworthy: being so justified would require his making a justified judgment that distinguishes between two perceptually indistinguishable options, each of which is a sufficiently plausible explanation of Donna's behavior in this isolated instance, and Rosen appears to think this is impossible.

It is worth noting that Donna herself will be in a very similar epistemic position to Thurston (and perhaps in a worse one). Suppose Donna wonders whether she is praiseworthy for donating. She wonders whether, at the time of her donating, her act was caused in the right way by the right kind of motivation. It is easy to imagine that even if Case 2 were the accurate description of Donna's story, Donna might well, on reflection, sincerely conclude: 'yes, I was moved to donate out of moral considerations of the relevant sort.' The possibility of self-deception is great—after all, Donna may take herself to be a morally conscientious person, and it would be quite jarring for her to believe that she values and is moved by self-interested reasons to the extent that she truly is.[13] Whether Donna's act was caused by the right kind of motivation is not the sort of thing that introspection will reliably pick up on—there are no introspectively detectable signs that will flag for her that her act was caused in the right way by the right kind of motivation. So, according to Rosen's implicit epistemological principles, *Donna* also cannot possibly be justified in believing that she is praiseworthy: being so justified would require her making a justified judgment that distinguishes between two introspectively indistinguishable options, each of which could be a plausible explanation of her behavior in this isolated instance, and Rosen appears to think this is impossible.

So, in Donna's case, Rosen's epistemological principles will suggest that no one can be justified in believing that she is praiseworthy. The DISTINGUISH principle says that in order to be justified in believing that Donna is praiseworthy, one must justifiedly believe that her donating was caused (in the right way) by her having the right kind of

[13]For a good discussion of the extent to which we confabulate and misread our affective attitudes, see Carruthers (2011: 119-155).

motivation. But, the scenario in which Donna's act is caused in the right way by a motivation of the right kind is neither perceptually nor introspectively distinguishable from the scenario in which it is not, and thus, according to DETECTION, no one can justifiedly judge that Donna's act was caused in the right way by a motivation of the right kind. Moreover, while I introduced the discussion of justified attributions of praiseworthiness with the perhaps 'hard' case of Donna so as to make skeptical worries especially salient, it should be clear that in *any* particular case of so-and-so doing what superficially resembles a praiseworthy act, *that* her act has the kind of causal history for being praiseworthy is not a fact which we can come to justifiedly believe by merely *observing* the performance of her act (and she herself cannot come to justifiedly believe it through mere introspection). Her act (in isolation) may look and feel exactly the same whether it has the right kind of causal history or not. *Were* it true that being justified in believing that X is praiseworthy for performing A required arriving at a justified judgment that A was caused in the right way by X's having the right kind of motivation (and *all* one could go on was the perceptually and introspectively available evidence afforded by the isolated instance), then we should doubt that anyone ever has such a justified belief. So, I suggest that Rosen's argument will apply equally well to judgments of praiseworthiness as it will to judgments of blameworthiness. To borrow from the Rosen passage quoted earlier: “given the real limitations on our access to the causal histories of human actions and to the states of knowledge and opinion that underlie them” (and, presumably, to the motivational states that underlie them), we should claim that “as a matter of fact we are never entitled to any significant confidence that the … act under consideration satisfies the necessary condition we have identified.”

2.4 Skepticism about Judgments of Doxastic Justification[14]

Rosen's move to distance his skeptical thesis from familiar unlivable ones was to suggest that nothing in his argument casts doubt on there being warranted attributions of praiseworthiness. By Rosen's own lights, the fact that his argument applies equally well against attributions of praiseworthiness ought to go some way toward assimilating his position to the unlivable skeptical theses outside of moral responsibility. In this section, I suggest that there is still further reason to assimilate Rosen's skeptical thesis to other unlivable ones. I will show that Rosen's epistemological principles will generate the following skeptical position: no one is ever justified in believing of any particular belief that it is doxastically justified.[15] Such a position is obviously unacceptable and unlivable—indeed, when coupled with Rosen's DISTINGUISH principle, this position may yield the result that we are never justified in making *knowledge* attributions (since doxastic justification is a distinguishing feature of knowledge). Such a position will not do. Needless to say, our social interactions and intellectual practices would be unrecognizably alien in a world wherein we have no warranted beliefs about who knows

---

[14]The discussion here is inspired by an argument found in BonJour (2003). BonJour argues that because one does not have direct access to the states of affairs which make for epistemic justification on an externalist view, one can, at most, make conditional claims about whether one (including oneself) has a justified belief if an externalist theory were true.

[15]Setiya (2012) offers a clear and tolerably uncontroversial statement (for present purposes) of the distinction between propositional and doxastic justification: "[w]hat one is propositionally justified in believing is what one's evidence supports, or makes epistemically probable, whether one believes it or not. To be doxastically justified by evidence is to believe *p* on the basis of that evidence, in the right sort of way" (2012: 60). What is characteristic of doxastic justification is that it has a basing requirement, and it is necessary for knowledge.

and who does not. If Rosen's epistemological assumptions generate *this* kind of skepticism, there is further reason to think that something must be going wrong in his skeptical argument.

Suppose for a moment that a crude form of process reliabilism is the correct account of doxastic justification. Nothing about the argument of this section depends on this being true, but process reliabilism will be a useful toy model. So, suppose that a belief is doxastically justified if and only if it is produced by a reliable process.

Consider Reilly. Reilly is at a preschool playground. A child Reilly has never met before approaches him and says "four plus seven equals eleven" (and let's suppose, for simplicity, that it is somehow obvious that the child genuinely believes what she says). According to the stipulated correct theory of doxastic justification, the child's belief is doxastically justified if and only if it was produced by a reliable process.

Now, recall that, in the case of attributions of blameworthiness, Rosen suggests that one must justifiedly judge that such-and-such either is or is the upshot of an akratic act in order to justifiedly judge that such-and-such is a blameworthy act—this follows from DISTINGUISH. Applying DISTINGUISH to Reilly's case, we should hold that Reilly must justifiedly believe that the child's belief was produced by a reliable process in order to be justified in believing that the child's belief is doxastically justified (Reilly must rule out that the child's belief is some "impostor" of doxastically justified belief).

Recall, also, that the skeptical fuel for Rosen's argument was that akratic acts have neither perceptually nor introspectively detectable marks that identify them as such (see DETECTION). Simply put, we have no direct access to the mental states that would determine whether some act is akratic. It should be obvious that reliably produced beliefs

have neither perceptually nor introspectively detectable marks that identify them as such, and if *this* is all it takes to generate a skeptical argument, then we should be skeptics about the possibility of warranted beliefs that some belief is doxastically justified. Consider, first, Reilly's judgment about the child's belief, then consider the child's judgment about her own belief.

Certainly, Reilly cannot, through merely observing this isolated bit of behavior, *see* which process the child uses to arrive at her belief and thereby determine that her belief was reliably produced. Whether the child believes that four plus seven equals eleven as a result of some reliable process (perhaps from testimony or from some conceptual understanding) or as a result of some unreliable process (perhaps she believes it on the basis of the fact that "four plus seven equals eleven" rhymes), her behavior may look exactly the same to Reilly. If Reilly has to arrive at the justified judgment that the child's belief was reliably produced in order to be justified in believing that the child's belief is doxastically justified (DISTINGUISH), and all he has to go on is the perceptual evidence afforded in this isolated instance, it should be clear that he cannot possibly become justified (DETECTION).

Similar skeptical considerations apply to the child's attempt to determine whether her own belief that "four plus seven equals eleven" is doxastically justified. (We can assume for a moment that the child possesses the sophisticated concept of something close enough to doxastic justification—the reader can also swap in a cognitively "normal" human adult for the child at this point). The child will not be able to introspect and *see* what process she used to arrive at her belief and thereby determine whether her belief was reliably produced. One's belief forming processes—which are the very things

which determine whether a belief is doxastically justified—are just not the sort of thing that one has access to via introspection. Thus, if the child has to arrive at the justified judgment that her belief was reliably produced in order to believe that her belief is doxastically justified (DISTINGUISH), and all she has to go on is the introspective evidence afforded in this isolated instance, it should be clear that she cannot possibly become justified (DETECTION).

There is nothing special about this particular case involving Reilly and the child—if in order to be justified in believing that such-and-such belief is doxastically justified, one must justifiedly believe that such-and-such belief was reliably produced, and all one has to go on in arriving at this judgment is (roughly) the evidence afforded by immediate behavioral observation and introspection within the isolated instance, then no one can ever be justified in believing that such-and-such belief is doxastically justified. And this is just to say that the epistemological assumptions that generate Rosen's skeptical thesis will similarly generate a skeptical thesis about judgments of doxastic justification.

We can put the point like this. Doxastic justification (unlike propositional justification) has a basing requirement. This being so, whether a belief is doxastically justified will (on any plausible account of doxastic justification) be a function of the very same sort of causal-historical facts about our mental lives which are at play in Rosen's account of blameworthiness.[16] To the extent that our limited access to these sort of facts

[16]It is worth emphasizing that nothing about my argument here turns on my using a process reliabilist or externalist account of doxastic justification. Even paradigmatic internalists about propositional justification like Feldman and Conee include in their attempted account of doxastic justification the requirement that the relevant belief be held [emphasis my own] "*on the basis of* some body of evidence *e*" (where *e* satisfies their conditions for propositional justification) (see Feldman and Conee, 1985: 24). Their term for doxastic justification is 'well-foundedness'.

generates a skeptical argument for attributions of blameworthiness, our limited access to these sort of facts will also generate a skeptical argument for attributions of doxastically justified belief.

2.5 Justified Judgments about Doxastic Justification and Blameworthiness

If Rosen's argument applies equally well against attributions of doxastic justification, I do not take the upshot to be that we ought to be skeptical in that domain, too. Rather, I suggest (and argue in this section) that Rosen's argument rests on epistemological assumptions that are simply mistaken.

Return, for a moment, to Reilly at the preschool playground attempting to figure out whether the child's belief that "four plus seven equals eleven" is doxastically justified. It's true that Reilly cannot arrive at a justified judgment that the child's belief is reliably produced by merely observing her behavior in this isolated instance. But there is no reason to suppose that Reilly must arrive at this judgment in this way to justifiedly believe that the child's belief is doxastically justified. Reilly has other options for arriving at a justified judgment about the child's belief beyond simply trying to figure out through immediate behavioral observation whether (and if so, how) the child's belief satisfies the necessary conditions of doxastic justification. He can collect a bit more evidence about the child. Perhaps he can ask her to answer a few other basic addition problems—if she gets most of them right, he can presumably eventually become justified in believing that she is doxastically justified when she forms beliefs about simple addition problems. He could also ask the child's teacher about the child's progress with addition—does she demonstrate understanding in her work (et cetera)? When Reilly is supplied further background information about the child's beliefs about addition, there is no reason to

suppose that he cannot come to be justified in believing that her present belief that *four plus seven equals eleven* is doxastically justified. Of course, the child's present behavior would look exactly the same to Reilly whether her belief was reliably produced or not, and Reilly *could* be mistaken about whether she is doxastically justified even once he has acquired a good deal of background knowledge about the child, but that's because justified beliefs are fallible.

Similar considerations apply to the child's own belief about whether she is doxastically justified. Introspection may not be an especially helpful guide for determining whether one's beliefs were reliably produced. Everyone has had what *feels* like an “aha!” moment of understanding, only to realize shortly thereafter that the supposed understanding was the result of some serious confusion or misguided thinking. However, the child needn't rely on mere introspection to figure out whether her belief is doxastically justified. If, for instance, the child knows that she has a good track record with these kinds of addition problems, then she may be well equipped to have a justified belief that her current mathematical belief is doxastically justified. In both the first and third person cases, one's background knowledge about the agent and her history with the relevant kinds of beliefs is central to one's justification for whether she has a doxastically justified belief. Mere immediate behavioral observation and introspection (without relevant background knowledge) will not be reliable guides here, but we do not need them to be. We should, thus, reject DETECTION—it is false that, when an agent's behavior has no perceptually or introspectively detectable features which would reliably indicate that her attendant belief was reliably produced, we cannot justifiedly judge that her belief was reliably produced.

With this in mind, let us consider how we might come to be justified in believing that some particular act is blameworthy. It will be useful to consider, first, William FitzPatrick's (2008) response to Rosen. As will come out in a moment, while FitzPatrick's remarks go some distance toward an adequate response to Rosen, his cannot be the full story—the full story, as we will see, draws from some of the general upshots from the above discussion on justified attributions of doxastic justification.

FitzPatrick suggests, quite rightly (to my mind), that Rosen 'exaggerates' the epistemic difficulties of identifying akratic and akratically derived acts (see FitzPatrick, 2008: 593). He considers four different kinds of judgments that Rosen claims we can never justifiedly make, and he then identifies a number of signs which he thinks *can* count as perfectly good (even if fallible) evidence to justify a judgment of each of the four kinds. He considers (a) the case of judging that one's own act is akratic, (b) the case of judging that someone else's act is akratic, (c) the case of judging that one's own act is the upshot of an akratic act, and (d) the case of judging that someone else's act is the upshot of an akratic act. Concerning (a), FitzPatrick suggests that "the experience of guilt or shame we sometimes feel even while acting" can count as good evidence that one is acting akratically (FitzPatrick, 2008: 595). Concerning (b), he notes that "sometimes people are honest about their failings and will come out and admit that they knew better, even at the time of acting, and felt ashamed even as they acted," and that "absent any special reason to doubt their sincerity, their testimony and accompanying emotions can constitute perfectly good evidence that they acted akratically" (FitzPatrick, 2008: 596-597). For (c) and (d), he borrows one of Rosen's cases: a doctor, ignorant of her patient's blood type, gives a transfusion with the wrong blood type, and her ignorance was the

result of her failing to have someone remind her to double-check her patient's blood type before the procedure (note: on Rosen's view, whether the doctor is blameworthy for her wrong act will depend on whether she is blameworthy for having failed to have some someone remind her to double-check). Concerning (c), FitzPatrick suggests that in many cases of this sort, one (in this case, the doctor) "will know perfectly well that she failed to set up this safety arrangement because she was embarrassed or was trying to cut corners and gambled that it wouldn't result in disaster" (FitzPatrick, 2008: 598)—the thought is that one's own sense of embarrassment can justify the belief that one had been acting akratically. And, concerning (d), FitzPatrick points to other behavioral cues which could provide reasonable evidence that a third-party's ignorance was the upshot of an akratic act: in the case with the doctor, for instance, a third-party might have seen the doctor taking active measures to conceal the fact that she wasn't setting up a safety arrangement, and, often, such "concealing behavior is good evidence that [the doctor] knew what she was doing was wrong, and so was acting akratically" (598). While FitzPatrick, I am sure, thinks there are additional ways that one might be justified in making a judgment of sort (a), (b), (c), or (d), the above discussion covers the general flavor of evidence which FitzPatrick identifies as relevant.

Rather than moving beyond the perceptually and introspectively detectable signs present in an agent's behavior, FitzPatrick suggests that many actions *do* come with detectable signs—one's own feeling of embarrassment or guilt, the familiar behavioral manifestations of a third-party's embarrassment or guilt, the behavior characteristic of concealment, et cetera—which are reasonably reliable indicators of akratic or akratically derived action. The thought is that there *are* features of actions to which we have

perceptual or introspective access and which are such that, when we pick up on them, we can justifiedly believe that the relevant action is akratic or akratically derived.

FitzPatrick helpfully reminds us that we are not altogether in the dark as to what features of an action might be suggestive of akrasia. However, in limiting the discussion to detectable features of actions by which we might distinguish genuine akrasia from its impostors,[17] FitzPatrick's response fails to meet Rosen's skeptical challenge. It should be clear that evidence about whether an agent felt guilty at the time of acting is not, by itself, good evidence of akrasia. The "lapsed Catholic" whose affective states are "lagging behind their beliefs" (Fricker, 2007: 37) may feel guilty (and show signs of feeling guilty) while performing actions they no longer believe to be wrong. Even where my sense of guilt is really a product of my believing that I am acting wrongly, when I reflect and consider whether I was acting akratically, it is not hard to imagine that I may—through some bit of self-serving but non-conscious rationalization—come to believe that my sense of guilt was tied to something unrelated to the action in question, or come to sincerely disbelieve that I felt guilty at all. Moreover, a third-party's concealing behavior, in isolation, gives me no more reason to believe that she has acted akratically than it does to believe that she did not want others to misunderstand and interfere with what she took

[17]FitzPatrick *does* discuss the evidence made available when a third-party—after her act—comes out and sincerely tells me that she was acting akratically. If I attribute akrasia to the third-party based on that kind of evidence, then I am not simply detecting some feature of the agent's original action when I attribute the akrasia. However, as I will show in the next paragraph, FitzPatrick's discussion of this kind of evidence is of no help given the dialectic—for one who is already worried about whether individuals can introspect and detect their own akratic acts, it will not help to point out (without first giving a convincing story about how agents can justifiedly self-attribute akrasia) that agents sometimes sincerely claim that they were acting akratically.

to be a permissible action. By themselves, the signs FitzPatrick points to do not give away the akratic and akratically-derived actions as such.

FitzPatrick is the first to point out that the evidence he identifies is fallible and defeasible, but his followup to this point is—given the dialectic—of no help in responding to Rosen. Given that a central worry for Rosen is that one cannot tell through introspection when she has acted akratically, it will not do to simply insist that (emphasis my own) "the presence of guilt and shame at the time of acting can often provide strong evidence of akrasia, since *it will often be clear* that … one does endorse standards that condemn the action, and that the guilt or shame in so acting stems precisely from that" (FitzPatrick, 2008: 596). After all, the worry is that introspection will *not* make these facts clear (and that apparent clarity can be illusory). It will also not do to simply note that agents sometimes come out and *say* that they knew they were acting wrongly and to insist that, 'absent any special reason to doubt their sincerity, their testimony and accompanying emotions can constitute perfectly good evidence that they acted akratically' (596-597). I may be absolutely *certain* that you are being *sincere* when you tell me that you knew you were acting wrongly at the time of action, but this, alone, does not justify my belief that you *were* acting akratically. (One way to bring this out is to revisit the case of Reilly and the preschooler: it should be clear that Reilly is no closer to holding a justified belief that the child's belief that *four plus seven equals eleven* was reliably produced when the child adds—however sincerely—that she is good at this sort of thing, that her belief was reliably produced, et cetera). For someone who is already skeptical about our ability to introspect and detect our *own* akratic acts, it will be of no help to simply point out that third-parties sometimes sincerely claim to know they were

acting akratically—the third-party may have it wrong. To adequately respond to Rosen, we need to move beyond a list of potentially distinguishing features of actions—we cannot trust our ability to introspect and *see* that (say) our guilt is a product of akrasia, and, until there is more to the story about how one might reliably identify one's own act as akratic, we cannot assume that a third-party's sincere say-so gives us sufficient reason to believe that she has acted akratically.

Return for a moment to the case of Reilly and the preschooler. There, it turned out that the justification for believing that the preschooler's belief was reliably produced did not depend on identifying a distinctive mark of reliably produced belief (there was nothing distinctive about the preschooler's mental state or behavior in reporting the belief which she or a third-party could pick up on and thereby identify her belief as reliably produced). Rather, one's background information about the relevant believer was central to one's justification for attributing reliably produced belief (and, thus, for attributing doxastically justified belief). And I think we can say something quite similar in the case of attributions of akrasia and akratically-derived action (and, thus, of blameworthiness). In particular, I will argue that it is only when one has adequate background information about an agent's standing moral beliefs and patterns of behavior and emotion that one can justifiedly attribute akrasia.

Suppose I know Paul pretty well but that Mary is a complete stranger to me. Suppose I see them commit the same morally wrong act A. Stipulate that, over the course of my knowing Paul, I have developed a reasonably good sense of Paul's patterns of behavior. I know a thing or two about his standing moral beliefs which would be pertinent to act A—I know, for instance, that he has a standing belief that A-ing (or

performing acts roughly like A) is wrong. I know something about the kinds of situations in which he is especially likely to act contrary to his standing moral beliefs, and I know that Paul performed A in precisely one such kind of situation. Moreover, I know which of Paul's behaviors and expressions signal that he is feeling guilty (it is worth noting that it can often take quite a bit of exposure to an agent before one gets a sense of the behavioral cues associated with his feelings of guilt). I know something about his patterns of feeling guilty: I know from past experience that he is especially likely to feel guilty when he violates his own standing moral standards pertaining to acts like A, and I know enough about the kinds of situations in which he is prone to feel guilty to know that he is not prone to irrational feelings of guilt like Fricker's “lapsed Catholic,” at least when it comes to acts like A. Let us stipulate that as he performed A (an action which I know to violate his own standing moral standards), Paul exhibited the kinds of behavior I've come to recognize as being indicative of his feeling guilty (and I picked up on this behavior).

Now, a third-party observer who knows *nothing* about *either* Paul or Mary may have no good grounds for believing that either one has acted ackratically (or is blameworthy), and the third-party observer may be equally unjustified when he believes that Paul is blameworthy as when he believes that Mary is. But I'm not like the ignorant third party observer. I may have no good grounds for believing that *Mary* is blameworthy, but, given everything I know about Paul, I have (at least) *reasonable* grounds for believing that he acted akratically and that he is blameworthy. I do not judge that Paul acted akratically and is blameworthy by looking for some distinguishing feature of akratic action—my judgment that he acted akratically and is blameworthy is the result of an inference to an explanation based on information about the observable features of

Paul's action (many of the sort which FitzPatrick identifies) set against loads of information I've accumulated about his standing moral beliefs and his patterns of behavior and emotion. There is no reason to believe that it is *impossible* for me to have warranted confidence in my belief that he is blameworthy, even if it might be impossible for a third-party observer who does not know him as well. (This should not be taken too far. I doubt that we must, say, know someone on a personal level to make a justified attribution of blameworthiness. I might have good background knowledge about the training of doctors that would give me decent grounds for making the judgment that Dr. So-and-so, whom I'm meeting for the first time, knew better than to have done such-and-such or that it is (in some way I am unable to identify) Dr. So-and-so's fault that she doesn't know better). Of course, I might be mistaken in my belief that Paul is blameworthy, but being mistaken is fully compatible with being fallibly justified.

Similar considerations apply to Paul's own case when he tries to determine whether he has acted blameworthily. If he tries to make this judgment just by searching for some feature of akratic action that would reveal itself to introspection, then he will not be able to arrive at a warranted belief that he has acted blameworthily. However, if he takes the information made available from introspection and sets it against background information he has about his own standing moral beliefs and his patterns of behavior and emotion—and perhaps he gets insights from close acquaintances so as to avoid giving himself a biased evaluation—there is no reason to suppose that it would be *impossible* for him to have a warranted belief that *this* was one of those times when he either knew better or it was his fault that he didn't know better.

Now, Rosen might respond that I am under-appreciating the force of his skeptical challenge. Even though I have lots of relevant background information about Paul's moral character and dispositions (his standing moral beliefs, his patterns of behavior and emotion, et cetera), how will this be of any help when I am trying to figure out which story is true of Paul *at the precise moment of action*: 1) that he maintained his standing belief that it is not okay to A, 2) that, through various self-serving and non-conscious psychological pressures, he very temporarily came to think that A-ing is okay, or 3) that he *suspended judgment* about whether it is okay to A?[18]

In responding to this worry, I point out that, though it was a mistake to think (with FitzPatrick) that the way to justifiedly attribute akrasia is just to see whether the agent's act has any of the distinctive marks of akratic acts, this does not mean that evidence of some of the features FitzPatrick identified—such as the behavior indicative of a feeling of guilt—cannot be used as *one* bit of evidence in a broader inference to an explanation that attributes akrasia. When set against the background information I have about Paul (that he is not prone to irrational feelings of guilt like Fricker's 'lapsed Catholic,' that he tends to feel guilty when he violates his own moral standards, that he has a standing belief that A-ing is wrong), the fact that Paul—as he performed A—showed signs I've come to recognize as indicative of his feeling guilty just makes explanations (2) and (3) much less plausible than explanation (1). The best explanation of Paul's signs of feeling guilty (when set against the background information I have about him) would *not* be that he underwent some self-serving and non-conscious rationalization to avoid cognitive

---

[18]See Rosen, 2004: 309. I thank an anonymous referee at *Pacific Philosophical Quarterly* for pressing me on this point and for suggesting a general line of response.

dissonance—after all, here is Paul seeming to be in *agony* as he acts—but, rather, would be that he feels guilty for performing this act he takes to be wrong. *Were* Paul prone to various irrational feelings of guilt, or *were* Paul not to have a standing moral belief that A-ing is wrong, then it might be more difficult to justifiedly label his act akratic, and his signs of feeling guilty might then play no role in the attempt to figure out whether his act is akratic. But, certainly, we *sometimes* have the kind of background information about agents which I have stipulated I have of Paul, and when we do, it will *sometimes* be the case that the best explanation of some agent's particular bit of behavior—given everything one knows about her—is that she was acting akratically. And this is just to say that we can sometimes justifiedly believe that some action is akratic rather than non-akratically weak-willed. We may need more background information to identify akrasia than we do to identify (say) acts in anger, but there is no reason to think that we never possess the kind of background information under discussion. Thus, the suggestion that we must be able to justifiedly attribute akrasia in order to justifiedly attribute blameworthiness presents no obstacle to our *sometimes* being able to justifiedly attribute blameworthiness.

We can agree with Rosen that one must be able to distinguish akratic and akratically-derived acts from non-akratically weak-willed acts in order to justifiedly attribute blameworthiness, and we can agree that akratic acts have no perceptually or introspectively detectable features which give them away as akratic. These considerations generate skeptical worries only if one is attracted to a principle like DETECTION. But there are good reasons to believe that DETECTION is false. It is not the case that, to justifiedly attribute akrasia, I must be able to identify and locate some perceptually or

introspectively detectable distinctive feature of akratic acts. When I am equipped with sufficient background information about an agent's standing moral beliefs and her patterns of behavior and emotion, there is no reason to believe that I cannot find myself with a body of evidence such that the best explanation of that evidence is that she has acted akratically. At any rate, if Rosen *insists* on some principle like DETECTION, then it will turn out that he has offered us not *just* a skeptical thesis about justified attributions of *blameworthiness*—rather, he will have offered us the thesis that we never make any justified judgments about *a whole mess* of phenomena which essentially involve causal-historical facts about mental states. Such phenomena include attributions of praiseworthiness and doxastic justification, and (by Rosen's own lights) it is at this point that Strawson reminds us that Rosen's thesis is not a serious option for us.

Rosen's paper may help to emphasize that perhaps we are qualified to judge someone's act as blameworthy only when we have reasonably good background information about the person's standing moral beliefs and her patterns of behavior and emotion. We shouldn't go around blaming "all willy nilly." But he gives us no reason to embrace a full-blown skeptical view about warranted attributions of blameworthiness, and, indeed, as I have tried to show, the only skeptical view on offer in Rosen's paper is one that is far more skeptical than he acknowledges.[19]

---

[19]I thank Hilary Kornblith and Peter A. Graham for helpful discussions of the material in this paper and for commenting on multiple drafts of it. I also thank an anonymous referee at *Pacific Philosophical Quarterly* for useful comments and for pressing me on important lines of argument.

CHAPTER 3

WHEN AND WHY IS IT DISRESPECTFUL TO EXCUSE AN ATTITUDE?[20]

3.1 Introduction

It is intuitive that, under certain circumstances, it can be disrespectful to excuse someone for an attitude (even for an attitude one finds objectionable).[21] Some attempts at understanding or explanation strike us as patronizing rather than compassionate: "you can't blame Donna for her belief about abortion—it's just how she was raised." In so excusing Donna's attitude, one may thereby undercut her status as an agent (as someone who is in some sense *active* with respect to the formation and sustaining of many of her attitudes).

While it is easy enough to find instances where it seems disrespectful to excuse an attitude, matters are complicated. When and why, precisely, is it disrespectful to judge that someone is not responsible for his attitude?

At present, the philosophical literature on disrespectful excusings is underdeveloped and overgeneralized. The few writers[22] who address the topic offer no account of respect with which to guide their discussions, and the discussions often proceed as though there were *always* something disrespectful about excusing a sane, rational agent for her attitude.

---

[20]A version of this paper is forthcoming in *Philosophical Studies*.

[21]Throughout this paper, I use "S excuses R for her attitude," "S excuses R's attitude," and "S judges that R is not responsible for her attitude" interchangeably.

[22]See, in particular, Angela Smith (2005, 2008, 2015), Christopher Bennett (2004), and Michele Moody-Adams (1993, 1994).

My aim in this paper is to give a more systematic treatment of this question on what respect requires in our judgments about responsibility for attitudes. Here is how things proceed. In Section 2, I clarify the central question of this paper: namely, when and why is it disrespectful to judge that someone is not responsible for his attitude? To motivate my answer, I first survey the arguments represented in or suggested by the literature purporting to show that there is always something disrespectful about denying a sane[23] agent responsibility for his attitude (Section 3). I show that these arguments are unsuccessful: they rely on false generalizations about what is involved in excusing an attitude. In Section 4, I sketch an account of respect with which to adjudicate when and why excusing an attitude is disrespectful. The position, in short, is that a judgment about S (including a judgment about S's responsibility for some attitude) respects S when and only when it is formed in a way that is properly responsive to who S is as an agent (to S's character, values, and, more broadly, her rational activity). I show how, on the correct account of respect, there can be cases where excusing someone for his attitude is not disrespectful (including some cases where the agent's attitude was produced by a reasons-responsive mechanism), and I offer a principled explanation for why excusing an attitude *is* disrespectful when it is. I conclude with some brief remarks on how this discussion can inform debates about responsibility for moral ignorance and objectionable attitudes. Our question about what respect requires in judgments about responsibility for attitudes has deep interpersonal significance, and it is overdue for sustained treatment.

[23]Here and elsewhere, I use "sane" in Susan Wolf's (1987) sense of having the capacity to understand and respond to moral reasons.

3.2 Clarification and the DISRESPECT THESIS

The central question of this paper is: when and why is it disrespectful to judge that someone is not responsible for his attitude? It is worth making a few clarificatory remarks upfront.

First, our question is about disrespect manifested in *judgment* rather than disrespect manifested in one's behavior toward another (we ask: "when is it disrespectful to judge, perhaps privately, that Todd is not responsible for his attitude?", not: "when is it disrespectful to *treat* Todd as though he were not responsible for his attitude?"). An assumption of this paper is that judgments can themselves be disrespectful (perhaps by objectifying another or by affronting her in her status as an agent).

Second, our question is about when excusing an attitude is disrespectful *to the one being excused.* There are important questions about what respect for the moral community or for victims requires in our judgments about responsibility (see Franklin (2013)), but this paper does not address them.

Third, our question about respect is different from the question that asks: when and why *is* an agent responsible for an attitude? Some philosophers suggest that one is *always* responsible for one's objectionable attitudes (see Robert Adams (1985)). Notice that establishing this position would go no distance toward answering our original question about respect. For, we can ask: "when someone *is* responsible for an objectionable attitude, when and why is it *disrespectful* (rather than, say, merely *mistaken*) to judge that he is not?". It may be that theoretical details about responsibility for attitudes will play an important role in answering the question about disrespect, but the mere fact that someone is responsible for some attitude does not straightforwardly

imply that it would be disrespectful to judge that he is not (and it certainly does not *explain* why this would be disrespectful).

Lastly, I note that our question assumes no particular theory of moral responsibility (on the staggering array of competing accounts, Michael Zimmerman aptly summarizes: "it's a mess" (2015: 45)). Angela Smith's theory-neutral targeting of responsibility, on which to say that S is responsible for x is to say that x reflects on her morally in some non-trivial way, will do for present purposes (2007: 468). Since Smith is one of the key figures to advance the view that it is disrespectful to excuse an attitude, special consideration is given to the implications of Smith's rational-relations/answerability view for the question of disrespect.[24]

Having clarified our question, let me locate some of the answers (or partial answers) on offer in the literature. Angela Smith briefly addresses our question in her influential articles on responsibility. She urges us to resist making a distinction between a "morally bad" agent and a "morally blameworthy" agent (between an agent with an objectionable attitude for which he is not responsible and an agent with an objectionable attitude for which he *is* responsible) partially on grounds of respect:

> This distinction itself would require us to regard some agents as the passive victims of their faulty judgments, as I was the passive victim of my faulty hearing. I think this is a dangerously patronizing and disrespectful stance to take toward another human being, one that we should be very reluctant to resort to in practice (2008: 390).
>
> To say that a person's judgmental activity is bad but that he is not responsible for it is, in effect, to say that he is not to be regarded as someone to be reasoned with, but merely as someone to be understood, treated, managed, or controlled. It is to regard a person as we would regard a vicious dog or a bratty toddler, someone to

[24]For presentations of the rational-relations view, see Smith (2005, 2008, 2015) and, for a related view, see Scanlon (1998, 2008).

> be avoided and/or trained, if possible, but not someone with whom it is possible to enter into relationships of mutual respect and recognition (2008: 388).

Elsewhere, she writes that

> To say that our grandparents are not responsible for their racist and sexist attitudes because of the conditions under which they were formed implies that they no longer have the capacity to recognize, assess, and respond to reasons bearing upon the justifiability of those attitudes. It is to treat these attitudes as fixed and non-reason-responsive features of their mental lives, rather than as judgment-sensitive states that essentially depend upon their own critical evaluation of the reasons they take to count in favour of them (2015: 125).

This echoes a point from her 2005 article, as well as points made by Christopher Bennett and Michele Moody-Adams:

> Indeed, to deny someone responsibility in a case of this sort strikes me as somewhat patronizing, insofar as it suggests that the person, because of her upbringing, is literally incapable of appreciating and responding to rational criticism directed at her evaluative judgments (Smith, 2005: 268-269).

> For me to fail to hold Grandfather responsible for his racism is to view him as less than my moral equal. I treat him as lacking certain fundamental moral capacities—the capacity to understand certain plain-as-day moral truths, for instance. … [I]n regarding someone as seriously (and not just temporarily) morally incapable, however understandably, we fail to pay them a sort of respect that we pay to those whose moral views we take seriously (Bennett, 2004: 8).

> [T]o exempt someone from responsibility and blame for wrongdoing is simply to deny that person's humanity (Moody-Adams, 1993: 125).

These passages gesture toward an answer to the "why" component of our question: the reason excusing an attitude is disrespectful is that it (for reasons in need of further clarification) involves a failure to recognize the person as an agent. What is striking about these passages is that they appear to suggest, in response to the "when" component of our question, that it is *always* disrespectful to deny someone responsibility for an objectionable attitude—note the unqualified language in each passage. Indeed, Angela Smith notes that the "morally bad" and "morally blameworthy" distinction "is certainly

relevant when it comes to the assessment of non-rational animals and young children" as well as "human beings who are subject to 'transient mental illnesses'" (2008: 388) but, when it comes to sane adults, she insists that

> [i]f I assess an (otherwise normal) adult human being as "cruel" I am not merely attributing a trait to him for which he may or may not be responsible (as when I say that a dog has a "vicious disposition"). I am (in part) making a demand of him, a demand that he justify the objectionable judgments his actions and attitudes expresses concerning the moral status of others. This demand *by its very nature* implies responsibility, for it is directed at his judgmental activity, activity for which we must regard him as responsible if we are to regard him as a moral agent in any sense (2008: 388).

And, here, Smith states without qualification that she takes her account of responsibility to imply that the *very holding of an objectionable attitude* makes one responsible for the attitude and that to excuse a person for an objectionable attitude is to deny that she is a moral agent:

> [T]o say that an attitude or an action is "morally bad" on this view is to say that an agent has *judged badly*, which is an assessment that implies both responsibility and blameworthiness. … [T]o regard a person as "morally bad" while refusing to see her as "morally blameworthy" is to deny her basic standing as a moral agent (2008: 391).

So, as Smith sees it, it is not only *true* that sane agents are always responsible for objectionable attitudes—it would be *disrespectful* toward these persons as moral agents to believe otherwise.

Let us refer to the general thesis given voice in the above passages as the "DISRESPECT THESIS."

> DISRESPECT THESIS: it is always disrespectful to judge of some sane individual that he is not responsible for his attitude.

In the following section, I precisify the rationales these writers offer for the DISRESPECT THESIS, and I show that these rationales fail (even when we weaken the thesis a bit).

3.3 Rationales for the DISRESPECT THESIS

3.3.1 The Incapacity Argument

On one rationale for the DISRESPECT THESIS, what makes it disrespectful to excuse an attitude is that, in so excusing, one implies that the person—the *agent*—is literally incapable of weighing moral reasons and rationally revising his attitude. Again, Smith writes that excusing an attitude is "patronizing insofar as it suggests that the person … is literally incapable of appreciating and responding to rational criticism directed at her evaluative judgments" (2015: 269), and Bennett writes that, in so excusing an agent, "I treat him as lacking certain fundamental moral capacities—the capacity to understand certain plain-as-day moral truths, for instance" (2004: 8). So, insomuch as excusing some sane agent's attitude involves judging of some sane agent that he is literally incapable of appreciating relevant moral facts, it is always disrespectful to excuse a sane agent for his attitude. Call this line of argument the "Incapacity Argument."

Unfortunately, those who advance the Incapacity Argument do little to motivate the suggestion that excusing an attitude invariably involves this implication about an agent's capacities. I grant that the following conditional is true: *were* one to judge of some sane individual that he is literally incapable of responding to evidence and rationally revising his attitude when one had perfectly adequate evidence of the agent's sanity, this would certainly disrespect the agent. However, there is no reason to think that

when I excuse the grandparent with the racist or sexist attitude, I thereby imply that he is *literally incapable* of understanding and rationally responding to the relevant reasons. After all, incapacity is but one among a variety of excusing conditions, so it hardly follows from the fact that I excuse him for his attitude that I imply *anything* about his capacities. Indeed, it is possible that when I excuse the grandparent, I judge that while there may have been *some* rational process by which he could have held an attitude other than the one he holds, we could not reasonably expect him to have believed otherwise given his upbringing and available resources (see FitzPatrick (2008)). It is simply false that excusing an attitude inevitably involves a judgment that the agent is literally incapable of responding to moral reasons. Thus, the Incapacity Argument fails to provide a rationale for the DISRESPECT THESIS.

Another way to reinforce what is wrong with the Incapacity Argument is to note that *one* coherent (and seemingly respectful) response to Grandfather's objectionable attitude could involve simultaneously 1) judging that Grandfather is not responsible for the objectionable attitude on the basis that, given his cultural circumstances, Grandfather's objectionable attitude does not reflect any particular lack of concern for morality (see Slote (1982) and Rosen (2002)), and 2) nevertheless judging that, since Grandfather *is* capable of understanding and responding to moral reasons after all, it would be good to bring to his attention all sorts of morally significant points that have probably never occurred to him so that he might revise his attitude. To the extent that this can be a coherent and respectful response to Grandfather's attitude (and surely it can), the Incapacity Argument fails to support the DISRESPECT THESIS.

3.3.2 The Rational-Relations Argument

The Incapacity Argument was mistaken in suggesting that excusing an agent's attitude invariably implies that the agent is literally incapable of responding to moral reasons. Here is a different rationale for the DISRESPECT THESIS. Rather than suggesting that excusing an attitude is disrespectful because of what it implies (more globally) about the agent and her capacities, we might suggest that excusing an attitude is disrespectful because of what it implies (more locally) about the agent's relationship to the very attitude for which she is being excused. According to Smith, excusing an agent's objectionable attitudes is to view those attitudes as "non-reason-responsive features of their mental lives, rather than as judgment-sensitive states that essentially depend upon their own critical evaluation of the reasons they take to count in favour of them" (2015: 125). She also writes that

> [States for which we are responsible] are "judgment-dependent" in the sense that they generally reflect and are sensitive to our (sometimes hasty, mistaken, or incomplete) judgments about what reasons we have, and they are generally responsive to changes in these judgments. We are "responsible for" these things, therefore, because they reflect rational assessments for which we are appropriately regarded as answerable" (2008: 370).

Perhaps what makes excusing an attitude disrespectful is that, in so excusing, one implies of some agent's sincerely held attitude that it is not connected to her judgmental activity in the way that is required for moral responsibility—excusing the attitude is to imply that it was not produced by the agent's having responded (even badly) to reasons.

This is a more promising way to motivate the DISRESPECT THESIS, and it sits nicely with Smith's own views about responsibility and what is involved in excusing an attitude. As Smith sees it, her account implies that anytime I judge of some agent that he is not responsible for some attitude, I thereby imply that his attitude was not produced by

his having responded (even badly) to reasons—I imply that his (implicit or explicit) assessments about reasons are completely unrelated to the fact that he holds this attitude. Put another way: in excusing the agent's attitude, I judge that the attitude is not in anyway expressive of the agent's rational agency but is, rather, something that *happens* in him like a headache. Since this seems like a disrespectful stance to take toward a sane agent's relationship to his sincerely held attitude, this suggestion appears to supply a rationale for the DISRESPECT THESIS. Let us call this argument the "Rational Relations Argument."

Right away, it should be clear that the Rational Relations Argument can—at most—supply a rationale for a restricted version of the DISRESPECT THESIS. For, at least sometimes, one can have perfectly adequate evidence that some particular attitude held by a generally sane agent was not produced by a reasons-responsive mechanism, and, surely, it would not be disrespectful to judge accurately and in accordance with one's evidence in such a case. David Eagleman (2011: 154-155) points to a case in which some person's deeply out of character interest in child pornography appears, disappears, and reappears in correlation with the growth, removal, and reappearance of a brain tumor in his orbitofrontal cortex. It seems, to me, that this person is not responsible for the objectionable attitudes toward child pornography—they do not reflect *who he is* as an agent, as the presence of the brain tumor appears to be a better explanation for his holding these attitudes than his explicit or implicit judgments about reasons. Surely, it would not be disrespectful for someone who knows all the relevant details about this

character to judge that he is not responsible for these objectionable attitudes.[25] So, let us consider a more restricted version of the DISRESPECT THESIS:

> DISRESPECT THESIS*: it is always disrespectful to judge of some sane individual S that he is not responsible for his attitude, except in cases where one has compelling evidence that S's attitude was not produced by his having responded to reasons.

Here, then, is the argument under consideration. The Rational Relations Argument suggests, first, that excusing S's attitude invariably involves a judgment that S arrived at this attitude without responding (even badly) to reasons. Such a judgment seems disrespectful, except when one has compelling evidence that S is insane or that S did not arrive at this attitude by responding to reasons. Thus, the argument concludes, it is always disrespectful to judge of some sane individual that he is not responsible for his attitude, except in cases where one has compelling evidence that his attitude was not produced by his having responded to reasons.

While the Rational Relations Argument may appear to provide a compelling rationale for the DISRESPECT THESIS*, this appearance is illusory. In the remainder of this section, I show that it is implausible that excusing an attitude invariably involves denying that the attitude was produced by a reasons-responsive mechanism; in fact, *contra* Smith, this suggestion does not even follow from her own account of moral responsibility. One can coherently judge that a person arrived at his objectionable attitude by way of responding to reasons while also judging that the attitude does not reflect on

---

[25]Indeed, Smith (2005) suggests that an agent would not be responsible for attitudes implanted by a mad scientist (with the caveat that, if the agent "becomes aware of these attitudes and shows no tendency to revise or reject them in light of her other beliefs and commitments, we may eventually conclude that these attitudes do accurately reflect her judgment" in a way that implies responsibility (261)).

him morally in the way relevant to responsibility. If I am right, then the Rational Relations Argument fails to provide a successful rationale for the DISRESPECT THESIS*, even if we grant Smith's account of responsibility.

This would be a significant result for Smith's overall project. For, Smith takes her account of moral responsibility and the DISRESPECT THESIS* to be mutually reinforcing.[26] Again, she writes that, on her view, holding an objectionable attitude *just is* to be responsible and blameworthy for it:

> I have … conceded that [my] view does not leave much space for the distinction between "bad agents" and "blameworthy agents," because to say that an attitude or an action is "morally bad" on [my] view is to say that an agent has *judged badly*, which is an assessment that implies both responsibility and blameworthiness (2008: 391).

And, to quiet theoretical objections one might have to this implication for her view, Smith reminds us that it is disrespectful to excuse an attitude:

> However, I have tried to show that this is a distinction we should not be eager to place weight on in our moral practices anyway, for to regard a person as "morally bad" while refusing to see her as "morally blameworthy" is to deny her basic standing as a moral agent. At the end of the day, I submit, it is much more respectful to be blamed for our moral faults than to be pitied for them (2008: 391).

So, as Smith sees it, her account of moral responsibility reveals what is involved in excusing an attitude such that it is disrespectful (thereby supplying a rationale for the DISRESPECT THESIS*), and the DISRESPECT THESIS* provides a compelling

---

[26]As noted earlier in this paper, when Smith explicitly discusses disrespectful excusings, she gives voice to a more sweeping and universal version of the thesis—that is, she seems to endorse the DISRESPECT THESIS rather than the DISRESPECT THESIS*. In what follows, I will attribute the more restrictive DISRESPECT THESIS* to Smith, since this thesis is both more plausible and sits better with Smith's own account of moral responsibility.

reason not to opt for an account of moral responsibility more restrictive than her own. Let us now examine her account.

In its essence, Smith's account states that S is responsible for x (whether an attitude or an action) when and only when x rationally reflects S's evaluative judgments. What are one's "evaluative judgments," and what is the relation of "rational reflection"? Let me clarify these in turn.

For Smith, evaluative judgments

> … are not necessarily consciously held propositional beliefs, but rather tendencies to regard certain things as having evaluative significance … [T]hese are continuing and relatively stable dispositions to respond in particular ways to particular situations, and not merely onetime assessments. … These judgments, taken together, make up the basic evaluative framework through which we view the world (2005: 251-252).

One's evaluative judgments are one's relatively stable, dispositional cares or (perhaps implicit) stances on what does and does not matter—they may include one's cares about (say) one's friends (2005: 243), intellectual freedom (252), or being liked by others (252). Having the evaluative judgment just is to be disposed such that one's patterns of behavior, belief, and noticing are appropriately shaped in ways that imply a certain implicit registering of value (whether positive, negative, or neutral). For instance, when one cares about a friend, one will be disposed "to recognize and to appreciate factors which bear upon her welfare" (243) across some reasonably broad range of circumstances, and to fail to be so disposed is to have an evaluative judgment of some lower level of care for one's friend.

What is the relation of "rational reflection" in Smith's account of responsibility? According to Smith, one is responsible for an attitude when and only when, in addition to the attitude, one has a dispositional, evaluative judgment such that "if one sincerely holds

[that] particular evaluative judgment, then the mental state [or attitude] in question should … occur," where "the 'should' in question here is the should of rationality" (253). On this view, one is responsible for one's objectionable attitude about a group of persons to the extent that the attitude "rationally reflects" a reasonably stable, objectionable disposition to (at least implicitly) hold them in low regard. And one is *not* responsible for (say) one's headache, since one's evaluative judgment that "headaches are bad" has no rational connection to whether the headache strikes (of course, one's evaluative judgment that "headaches are bad" has a rational connection to a range of behaviors aimed at preventing or getting rid of a headache).

Smith's account is easy to gesture toward yet difficult to precisify. Indeed, Smith uses a variety of (seemingly) substantively different turns of phrase to clarify the central relation of "rational reflection." On one hand, she suggests that the "rational reflection" relation is some sort of etiological connection: on this reading, one is responsible for an attitude when and only when the attitude was somehow formed *because* of one's evaluative judgments (as opposed to being formed because of, say, a brain tumor). Attitudes for which we are *not* responsible, she writes, are (emphasis my own) "not *based* upon the agent's own evaluative appraisal of her situation" and are formed "in a way which *bypasses* her rational capacities altogether" (2005: 261-262). And attitudes for which we *are* responsible are "*governed by* our evaluative judgments" (257). On the other hand, Smith sometimes writes as though the relation of "rational reflection" were merely a relation of rational fit between the evaluative judgment and the attitude (independent of how the attitude was actually formed). For instance, she writes that (emphasis my own) "to say that a person is responsible for an attitude, on the rational

relations view, is simply to say that that attitude is, *or should be*, sensitive to her evaluative judgments" (267). As I understand it, Smith's inclusion of "or should be" indicates that it is sufficient for responsibility that an attitude have some rational fit with one or more of the agent's evaluative judgments, where this fit is completely independent of whether the attitude's etiology involves any of those evaluative judgments.

Whatever way we spell out the details of the view, this much is true: for Smith, I am responsible for x when and only when x rationally reflects one or more of my reasonably stable, dispositional judgments, such that the presence of x is in some sense *rationally explained* by my evaluative judgment(s). For our purposes, this level of specificity will do.

Now, as mentioned above, Smith suggests that, on her account, the very holding of an objectionable attitude implies that one is responsible and blameworthy for the attitude (2008: 391). Furthermore, she suggests that, on her view, excusing an attitude invariably involves a judgment that the attitude was not produced by the agent's having responded (even badly) to reasons (2015: 125). Having clarified Smith's account, I can now show that neither of these claims follows from it—if these claims *seem* to follow from her account, it is (I suggest) only because we are running two different things together: namely, one's standing, dispositional evaluative judgments about *reasons for belief* and one's standing, dispositional evaluative judgments about what is of *moral value*. Let me explain.

Some of our evaluative judgments are about what matters morally. One may have reasonably stable dispositions to regard one's friends, honesty, or the environment as having some level of value. But other of one's reasonably stable, dispositional evaluative

judgments are about what count as good reasons for belief. One may have a reasonably stable, implicit disposition to treat [*the fact that the smart people around me suggest that p is true*] as a good reason to believe that p. When one believes that p because the smart people around one suggest that p is true, one's belief that p reflects a sensitivity to one's environment—one is believing p *for a reason* (for what is generally a pretty good reason, even). In such a case, one's belief that p is rationally related to one's standing, dispositional evaluative judgment that "[*the fact that the smart people around me suggest that p is true*] is a good reason to believe that p." One's relationship to one's belief that p in such a case is dramatically different from one's relationship to one's belief that is caused by (say) a brain tumor. We can say, with Smith, that in the former case and not the latter, one is responsible for holding the belief; in the former case and not the latter, one's belief is rationally explained by one's evaluative judgments about reasons for belief. To use Smith's turn of phrase: in the right circumstances, if "one sincerely holds [this] particular evaluative judgment [about reasons for belief], the mental state in question [the belief that p] should … occur" (2005: 253).

I am happy to say that in a case where I have perfectly adequate evidence that S's belief that p was produced because of S's implicit, evaluative judgment about the merit of testimonial evidence, it *would* be disrespectful to deny S responsibility for having formed the belief that p. However, given that some of our evaluative judgments are about reasons for belief and others are about moral value, it should be clear that there are different senses in which one may be responsible for some attitude (even on Smith's view) and that there are ways of excusing an attitude that do not involve denying that the attitude reflects the agent's judgments about reasons for belief. We can ask whether the attitude

rationally reflects the agent's evaluative judgments about *reasons for belief*, and we can also ask (as a separate question) whether the attitude rationally reflects some reasonably stable objectionable evaluative judgment(s) about *moral value*. What I want to suggest here is that, to be responsible and blameworthy *for the morally significant features* of some attitude (as opposed to simply being responsible for *holding* or *having formed* the attitude), the attitude must rationally reflect more than the agent's evaluative judgments about reasons for belief. For the agent to be responsible for the morally significant features of the attitude (that is, for the attitude to reflect on him *morally* rather than simply reflecting on him *rationally*), the attitude must rationally reflect some reasonably stable, dispositional, objectionable evaluative judgment(s) about what is or is not of moral value. If I am right, then there is room to consistently judge that S arrived at his objectionable attitude by responding to reasons for belief while also denying that S is responsible or blameworthy for the morally significant features of that belief. And if that is right, then, to the extent that the Rational Relations Argument for the DISRESPECT THESIS* relies on the assumption that excusing an attitude invariably involves a judgment that the attitude was not produced by the agent's having responded (even badly) to reasons, the Rational Relations Argument fails.

To motivate my suggestion, I need to say something about when an attitude might rationally reflect the agent's evaluative judgment(s) about reasons for belief but not some reasonably stable, objectionable evaluative judgment(s) about moral value. So, consider the following case. Stipulate that it is morally wrong to kill nonhuman animals for food and that the belief that "it's okay to kill animals for food" has morally objectionable content. Suppose that Barry lives in a society in which it is just taken for granted that

killing nonhuman animals for food is fine—for the vast majority of the society, it does not even occur to the individuals that there is a relevant moral question to ask. Suppose that Barry is (whether deliberately by a third party or just through the natural course of things) "shielded" from opportunities to seriously appreciate that there is a live moral question here. He has never seen or heard about the conditions under which the animals are raised and killed, nor has an opportunity to wonder about those conditions ever seriously arisen for him. He has never met or interacted with anyone who is morally opposed to killing nonhuman animals for food, nor is he even aware that there are such people. Moreover, all the smart and friendly people around him share and reinforce the belief that killing nonhuman animals for food is *obviously* fine.

Now, this much is clear: Barry's attitude is produced in a way that reflects a sensitivity to his environment—it rationally reflects his dispositional, implicit evaluative judgment that "[*the fact that all the smart people around me believe that p*] is a good (yet defeasible) reason to believe that p." His attitude is dependent upon and explained by his having this evaluative judgment about testimonial evidence, and the attitude reflects on him *qua* rational agent in a way that an attitude produced by a brain tumor would not. Thus, in one important sense (and in keeping with Smith's account), Barry is responsible for having formed this attitude.

However, given the nature of Barry's evidence and what he has (and has not) been exposed to, it is dubious to suppose that his attitude rationally reflects (is dependent upon, is rationally explained by) his having antecedently held some relatively stable, dispositional, objectionable evaluative judgment(s) about nonhuman animals. Put differently: surely, the best explanation of Barry's attitude does *not* involve attributing to

him some reasonably stable or moderately counterfactually robust, dispositional, objectionable evaluative judgment about the low value of nonhuman animal experience. If all one knows about Barry are his cultural circumstances and that he believes it's okay to kill animals for food, one cannot justifiedly attribute to him such a dispositional, evaluative judgment. For, one would have *no* evidence concerning a range of things at least *some* of which would need to be true if Barry is to have some *reasonably stable* or *moderately counterfactually robust*, dispositional, objectionable evaluative judgment about nonhuman animals: one would have no evidence that Barry would maintain his attitude when presented with halfway decent live opportunities to evaluate it (that he would maintain the attitude after, say, hearing about the conditions under which the animals are raised), or that he would deliberately inflict harm on animals, or that he would not treat a squirrel's capacity for good/bad experiences as a reason to avoid running one over, et cetera. Indeed, one's total body of evidence may suggest that Barry does *not* have these evaluative dispositions. There is room to judge that Barry's attitude is rationally explained by some evaluative judgment about testimony as a source of evidence but that the objectionable features of the attitude are not rationally explained by some reasonably stable, dispositional, objectionable evaluative judgment about the moral value of nonhuman animals. Thus, there is room to excuse Barry for the objectionable features of his attitude without thereby implying that his attitude was formed as though because of a brain tumor.

Now, *were* Barry to see or hear about the conditions under which nonhuman animals are raised and killed for food, or *were* he to meet people who voice moral opposition to killing nonhuman animals for food, then it may well rationally reflect a

reasonably stable, objectionable evaluative judgment about the moral value of nonhuman animal experience when he maintains this attitude. In such a case, the *persistence* of his attitude is best explained by the fact that he has at least a somewhat reasonably stable, objectionable evaluative judgment about the value of nonhuman animal experience. Nomy Arpaly (2002: 103-104) presents the case of Solomon, who grows up in a small, isolated farming community where all his evidence suggests that women are not cut out for abstract thinking—none of the women in his community discuss abstract matters, and none of the books in the small library are written by women, et cetera. Arpaly suggests that when Solomon believes that women are not cut out for abstract thinking, he "is more ignorant than irrational, and he is also, intuitively, more ignorant than morally vicious" (104). However, were Solomon to hold on to this belief after spending a year in an academic institution where he has plenty of evidence about brilliant female thinkers, "he would no longer be simply mistaken, but *prejudiced*" and "suffering from a serious moral flaw" (104).

Though Arpaly does not put it in quite these terms, I think we can usefully diagnose Solomon's case as follows. Before enrolling in the academic institution, Solomon's attitude about women rationally reflects evaluative judgment(s) about reasons for belief (he is responsible for *holding* this attitude, as he is responding to evidence), but his attitude does not rationally reflect evaluative judgment(s) with objectionable content (he is not responsible for the objectionable features of his attitude). But, when his attitude persists in the face of lots of counterevidence, we are now licensed to conclude that his attitude rationally reflects some evaluative judgment(s) with objectionable content, and he may well be responsible for the objectionable features of his attitude.

Thus, Smith is mistaken to suggest (even on her own view) that excusing someone for an attitude invariably involves a judgment that he did not arrive at his attitude by responding (even badly) to reasons—to the extent that the Rational Relations Argument for the DISRESPECT THESIS* relies on an assumption to the contrary, it fails. Of course, one *can* excuse someone on the grounds that he was brainwashed or that his attitude was produced by a brain tumor (and, when one's evidence suggests that one such story is true, it surely is not disrespectful to judge accordingly). But there is plenty of room to judge, in many other cases, that an agent arrived at his attitude through responding to reasons for belief but that, given the evidence against which the agent (rationally) arrived at this attitude, he is not responsible for the morally objectionable features of the attitude.

While I have shown the more ambitious claim that the Rational Relations Argument fails even granting Smith's own account of responsibility, we can note that, setting Smith's account aside, it is implausible that excusing an attitude invariably involves denying that it was produced by a responsiveness to reasons. First, there are plausible rival accounts of responsibility on which excusing an attitude does not imply that the attitude reflects *none* of the agent's evaluative judgments. For instance, Holly Smith (2011) persuasively argues that even if an attitude reflects *an* objectionable evaluative judgment, there is still room to judge that the agent is not responsible or blameworthy for the attitude on the grounds that it does not reflect a *sufficiently wide set* of the agent's total network of evaluative judgments. Moreover, in many of the cases in which there is a pretheoretical inclination to excuse an attitude, the inclination is largely informed by the fact that the agent *did* arrive at his attitude rationally: "yes, it's an

objectionable attitude, but—look—Barry is only believing what it was *rational* for him to believe given his limited evidence; his belief doesn't reflect any lack of concern for morality." *Contra* the Rational Relations Argument, excusing someone for the objectionable features of his attitude does not invariably involve judging that his attitude was produced without a responsiveness to reasons.

Before moving on, we can consider a brief counterargument on behalf of Angela Smith. Smith might argue that Barry's belief that "killing nonhuman animals for food is fine" implies that he *does* have some objectionable evaluative judgment about the value of nonhuman animal experience and, thus, that he *is* responsible for the objectionable features of his attitude (at least on the true account of responsibility). After all, if Barry had the *correct* evaluative judgment about nonhuman animal experience instead of whatever evaluative judgment he actually holds, then he would not believe that it is permissible to kill nonhuman animals for food just because of the testimonial evidence. I agree that Barry lacks the reasonably stable, dispositional evaluative judgment (roughly) that "nonhuman animal experience matters in a way that makes it wrong to kill them for food." But lacking this judgment is not equivalent to actively holding some reasonably stable or somewhat counterfactually robust, objectionable judgment that nonhuman animal experience *lacks* value, nor is it equivalent to having an objectionable lack of concern for nonhuman animal experience. Sure enough, there are cases of motivated ignorance, or cases where someone's objectionable motives or values make him resistant to evidence and therefore responsible for his objectionable ignorance (see Moody-Adams (1994)). But, given Barry's limited exposure to serious opportunities to see that there is even a live moral question about animal ethics, it is not the case that his attitude is best

explained by his having objectionable motives or values that make him *resistant* to evidence. Barry's moral ignorance—especially if he has a history of coming to believe the moral truth for other moral topics when previously unavailable moral considerations are made salient to him—is better explained by the fact that his total body of evidence makes it rational for him to focus his moral attention elsewhere than on the ethics of eating animals.

This point is worth emphasizing. Each of us is inevitably faced with an extraordinary range of decisions bearing on an exceptionally wide range of moral issues. But our capacity to focus attention on these issues is limited both by the pressing matters of everyday life and by the extent to which these moral matters are made salient to us. It does not show a character defect when one fails to accurately weigh *all* of the morally relevant considerations that bear on one's behaviors and beliefs. When it comes to responsibility for the objectionable features of an attitude, the question to ask, I suggest, is: is that attitude rationally explained by attributing the agent a reasonably stable objectionable evaluative judgment? And I argue that, in a case like Barry's, the answer can be "no."

The Rational Relations Argument claims, in support of the DISRESPECT THESIS*, 1) that excusing an attitude is to judge that the agent's attitude was not produced by a reasons-responsive mechanism, 2) that this is a disrespectful stance to take toward someone's relationship to her own attitude when one has no compelling evidence that the agent is insane or that the attitude was produced without the agent's having responded to reasons, and so 3) it is always disrespectful to excuse some sane agent's attitude, except when one has compelling evidence that the agent is insane or that the

attitude was produced without the agent's having responded to reasons. But I have shown that (1) is false, even when we grant Smith's account of responsibility. The Rational Relations Argument fails to provide a rationale for the DISRESPECT THESIS*.

3.4 Respect and Proper Responsiveness

None of the suggestions on offer from Smith, Bennett, or Moody-Adams supplies a successful rationale for anything as strong as the DISRESPECT THESIS*. In this section, I sketch an account of respect and use it to show that the DISRESPECT THESIS* is false and to offer, more broadly, a principled framework with which to adjudicate when and why it is disrespectful to excuse an attitude. The core idea underlying my account is that respect requires that my judgments about a person be properly responsive to my evidence about her *qua* agent.

I locate my account of respect in the tradition that includes Stephen Darwall's (1977) notion of "recognition respect" and Robin Dillon's (1992) notion of "care respect," which she takes to be a species of recognition respect (see Dillon, 1992: 112). Dillon writes that the "core of care respect … is attention to and appreciation of individual persons in the richness of their concrete particularity" (119). I agree with this general targeting of respect, though important details need clarification. In particular: what kind of attention to and appreciation of persons does respect require? What makes for deficient attention and appreciation?

I suggest that a judgment—whether a judgment about responsibility or otherwise—respects S when and only when it is properly responsive to S *qua* agent (to the facts about her rational activity), and a judgment is properly responsive to S *qua* agent when and only when the judgment about S *qua* agent is not produced by an epistemically

or morally bad inference.[27] Agents, in virtue of their natures, command a form of respect that requires that our judgments about them meet some basic epistemic and moral standards of recognition. To clarify the account, I now show how it applies to an excusing judgment (a judgment that denies someone responsibility).

On my account, whether an excusing judgment is disrespectful will depend upon its content and the evidence upon which the person forms the excusing judgment. We have seen from the previous section that an excusing judgment can have a variety of contents, as one can excuse an attitude for a variety of reasons. One can excuse S for his attitude because 1) she judges that S is literally incapable of responding to moral reasons, 2) she judges that S did not arrive at this particular attitude by responding to reasons for belief, or 3) she judges that the morally objectionable content of S's attitude does not rationally reflect some antecedent, relatively stable objectionable evaluative judgment about what is of moral value (this list is not intended to be exhaustive). My claim is that, in each case, the excusing judgment will be disrespectful only if it fails to be properly responsive to the evidence about S's rational activity.

In some cases (when an agent is not sane), one will have perfectly adequate evidence that S *is* incapable of responding to reasons, and it will not be disrespectful to judge accordingly. Yet, in a case where one's evidence suggests that S *is* sane and one

[27]This targeting of respect is adequate for the purposes of this chapter, though a complete discussion of my account of respect requires some precisification. As we will see in Chapter 5 (which focuses more explicitly on capturing the nature of respect), *proper responsiveness* requires more than that the judgment not be produced by an epistemically or morally bad inference. Rather, it requires that the judgment meet some epistemic and moral standards, where the satisfaction of those two standards is related. In particular, I argue that it requires that the judgment be reliably produced *because* of the influence one's values have on how one accesses and processes information.

judges that she is not, this would be disrespectful. However, as previously shown, one can excuse an attitude without thereby implying that the agent is literally incapable of responding to moral reasons. One might judge just that some *particular* attitude of S was not produced by a responsiveness to reasons. And when such a judgment is just what one's evidence supports, surely the judgment would not be disrespectful: one can, after all, have evidence that an attitude was produced by (say) a brain tumor. Of course, if one judges that S's attitude was not produced by a responsiveness to reasons when one's evidence does not support such a judgment, then one's judgment is disrespectful; this is what goes on in a paradigmatically disrespectful excusing, where one judges—with no sensitivity to the evidence about S—that S believes that p only because of, say, hormones. But, again, one can excuse S for some attitude without denying that the attitude was produced by a responsiveness to reasons—one might judge that the morally significant content of the attitude does not rationally reflect some reasonably stable objectionable evaluative judgment about moral value. And, here too, when this is just what one's evidence supports, surely the judgment is not disrespectful.

My framework can help explain why (*contra* the DISRESPECT THESIS*) one can excuse Barry for the objectionable content of his attitude that "killing nonhuman animals for food is fine" without disrespecting him. We can suppose that, given my background knowledge about Barry, my evidence suggests that his attitude reflects a reasonably stable, evaluative judgment that "[*the fact that all the smart people around me say that p is true*] is a good reason to believe that p." Perhaps I have seen him similarly rely on testimonial evidence in other cases, and I have no reason to think that this attitude was caused by (say) a brain tumor or a hypnotist. Here, it *would* be disrespectful for me

to judge that Barry is not responsible for having formed his belief—my judgment would be resistant to my evidence about Barry *qua* agent. But things are different when it comes to a judgment about whether Barry's attitude reflects a reasonably stable, objectionable evaluative judgment about moral value. My total body of evidence does *not* suggest that his attitude reflects a reasonably stable, objectionable evaluative judgment about nonhuman animal experience—the best explanation of his attitude does not involve attributing to him some counterfactually robust, objectionable evaluative judgment about nonhuman animal experience. So, when I judge that he is not responsible for the objectionable features of his attitude, there is no sense in which I am failing to be properly responsive to my evidence about Barry *qua* agent, and there is, thus, no sense in which my judgment is disrespectful.

It is possible, on my view, that one respectfully but mistakenly judges that someone is not responsible for an attitude. Consider Larry, who lives in the same community as Barry and who has had the same non-exposure to live opportunities to question the permissibility of killing nonhuman animals for food. I am open to the possibility that, as a matter of fact, even with his limited experiences, Larry has some reasonably stable, counterfactually robust evaluative judgment about the low value of nonhuman animal experience that plays a causal role in the formation of his attitude that it is fine to kill nonhuman animals for food. In this case, Larry's attitude *really does* rationally reflect some objectionable evaluative judgment, and he is responsible for the morally objectionable content of his attitude (at least on Smith's view). However, for a third party looking in on Larry's situation, the evidence one has about Larry simply does not support the inference that his attitude is best explained by his having some antecedent

reasonably stable evaluative judgment about the value of nonhuman animal experience. In this case, I suggest that it would not be disrespectful for the third party to judge that Larry is not responsible for the morally objectionable content of his attitude, even though this judgment is mistaken (at least, if Smith's view is correct). For, here, the third party's judgment manifests no failure to be properly responsive to who Larry is as an agent.

Of course, there is some gray area with respect to whether an agent's objectionable attitude rationally reflects an objectionable evaluative judgment, and it will not always be clear what inferences are licensed on a third party's evidence, and so it will not always be clear whether some excusing judgment is disrespectful. If Barry simply hears that there is a vegetarian somewhere in the world and he holds on to his objectionable attitude, should one then believe that Barry's attitude rationally reflects a reasonably stable, evaluative judgment about the value of nonhuman animal experience? It seems not—given the evidence about Barry, the best explanation of his attitude would still not involve attributing to him a reasonably stable, counterfactually robust evaluative judgment about nonhuman animal experience. It is not perfectly clear *how much* exposure to live opportunities to consider the wrongness of killing nonhuman animals for food Barry must have for the evidence to suggest that his attitude rationally reflects a reasonably stable evaluative judgment about nonhuman animal experience, but this should hardly be surprising: we are dealing with a complex psychological phenomenon.

I have offered an account of respect that shows that the DISRESPECT THESIS* is false and that provides a principled way to adjudicate when and why it is disrespectful to excuse an attitude (even if there are tricky cases). According to my account, a judgment about S respects S when and only when it is properly responsive to S *qua*

agent, and a judgment is properly responsive to S *qua* agent when and only when it is not produced by an epistemically or morally bad inference. Drawing on this account, I have shown the DISRESPECT THESIS* to be false by noting that, in some cases, without being resistant to the evidence about S *qua* agent, one may excuse S for the objectionable content of his attitude even while judging that the attitude was produced by a responsiveness to reasons.[28]

3.5 Conclusion

Let us take stock of what this paper has accomplished. I have shown, first, that familiar and breezy attempts in the literature at showing when and why it is disrespectful to excuse an attitude are unsuccessful—the discussions are underdeveloped, and the verdicts are overgeneralized. Along the way, I have shown that, in some cases, one can coherently (and respectfully) excuse an agent's attitude without implying that he is literally incapable of responding to reasons and without implying that the particular attitude was produced without his having responded to reasons. Indeed, I have shown

[28]A quick clarificatory remark about my targeting of respect is in order (the remark has no serious bearing upon the central uses to which I have put my account in this paper). If a person suffers from a serious cognitive disability that makes her incapable of drawing epistemically good inferences about S *qua* agent, it seems mistaken to suggest that her judgment about S is outright *disrespectful* (though it is not obviously mistaken, to me, to suggest that the judgment may be lacking in some level of respect). This is why, when I precisify the account of respect in Chapter 5, I distinguish between respect, lack of respect, and outright disrespect. Roughly: a judgment about an agent is *respectful* when it is reliably produced *because* of the influence from one's values; a judgment is *disrespectful* when it is unreliably produced because of one's values; and a judgment is lacking in respect when it is neither respectful nor disrespectful: for instance, if one's judgment fails to be reliably produced but not because of one's values but because (say) one is exceedingly hungry or because an unexpected distraction arose. These distinctions are not central to the core arguments of this chapter, though they matter in Chapter 5.

that, even on Smith's rational relations account of moral responsibility, there is—*contra* Smith—room to distinguish between an agent's being responsible for *having formed* an attitude and an agent's being responsible for *the morally significant features* of the attitude. Moreover, while the literature addressing disrespectful excusings has failed to draw from any particular account of respect, I have offered a plausible account that helps to explain when and why it is disrespectful to excuse an attitude. On my account, a judgment respects S if and only it is properly responsive to who S is as an agent. This account applies to judgments generally, including judgments that deny someone responsibility for an attitude, and it shows that the DISRESPECT THESIS* is false while also explaining why excusing an attitude is disrespectful when it is.

Smith, Bennett, and Moody-Adams suggest that when we see the extent to which it is disrespectful to excuse an attitude, we should resist accounts of responsibility on which cognitively normal agents are sometimes exculpated for false moral beliefs and other objectionable attitudes (for such accounts, see Rosen (2002, 2004, 2008), Zimmerman (1997), and Levy (2009)). This paper shows that while it certainly *can* be disrespectful to excuse an attitude, nothing like the sweeping thesis found in Smith, Bennett, and Moody-Adams is true. The facts about disrespectful excusings do nothing to motivate a categorical rejection of blameless moral ignorance and objectionable attitudes.

PART II: THE VALUE-SECURED RELIABILITY THEORY

CHAPTER 4

MORAL WORTH AND CONSCIOUSNESS: IN DEFENSE OF A VALUE-SECURED RELIABILITY THEORY

4.1 Introduction

Huckleberry Finn is considering whether to continue helping Jim escape from slavery.[29] Consciously, Huck believes that it is morally wrong to help a slave escape—his conscious belief reflects the popular and mistaken norms of his society. A ripe opportunity to turn in Jim presents itself. All the while believing that he may go to hell for doing what is wrong, Huck decides to continue helping Jim escape, even though he has no clear story he can tell himself or others about *why* he helps Jim escape.

Though Huck is unaware of the moral significance of his action, many find it intuitive that his action is not only morally right but morally worthy: non-accidentally right and attributable to *him*.[30] If this is correct (and I believe that it is), then Huck-like cases raise a puzzle about the relationship between moral worth and consciousness.[31] On

---

[29]Bennett (1974) is widely taken to be the first to bring Huck's case to the attention of moral psychologists. Arpaly and Schroeder (1999) reignited interest in the case.

[30]A quick note on terminology. Throughout this paper (and following the literature), I will use "S's action has moral worth" interchangeably with "S's action is praiseworthy" and "S is morally responsible for her right/good action."

[31]This paper largely proceeds on the assumption that it *is* a datum that there is some version of Huck's story on which his action is morally worthy. However, for those who are skeptical that Huck's action is morally worthy, the account of moral worth I offer in Section 4 can supply an argument that, on a plausible account that secures everything we might want from an account of moral worth, Huck-like actions can be morally worthy.

one hand, such cases suggest that there is some sense in which moral worth does not require awareness of the moral significance of one's action. On the other hand, consciousness must play *some* role in securing moral worth: Huck's behavior was not morally worthy if he was in a trance and altogether cognitively disconnected from the morally significant features of his situation. So, what minimal role *does* consciousness play in securing moral worth?

According to one popular view, a right action is attributable to the agent and non-accidentally right as moral worth requires only when the agent is conscious of the facts that make it right. Intuitive as this may be, I argue that this cannot be the minimal consciousness condition on moral worth. As I show, consciousness of such facts requires *much* more sophistication than writers typically suggest—this condition would bar from moral worth most ordinary, intuitively morally worthy agents. Moreover, I show that satisfying this condition cannot play a significant role in securing non-accidentality anyhow, and it is not necessary for either attributability or non-accidentality. What we need is a consciousness requirement on moral worth where 1) the requirement is not so demanding that Huck and other intuitively morally worthy agents are automatically ruled out, but where 2) it is nevertheless plausible that satisfying the requirement would play some significant and necessary role in securing attributability and non-accidentality. We want to know: what is the most minimal sense in which an agent must be conscious of the moral significance of her action if the action is to be non-accidentally tied to the right and attributable to the agent as moral worth requires? This paper shows what is mistaken about the popular response, and it offers an account of moral worth from which a minimal yet indispensable role for consciousness falls out.

Here is how things proceed. Section 2 clarifies our question and specifies some desiderata for a minimal consciousness condition. In Section 3, I show, first, that the depth of disagreement between the Anti-Consciousness Camp (those theorists who actively seek to downplay or eliminate any role for consciousness) and the Pro-Consciousness Camp (those theorists who seek to emphasize its role) has been exaggerated.[32] Across *both* camps, there is widespread commitment to the aforementioned view that an action is non-accidentally right and attributable to the agent only when the agent is conscious of the facts that make it right. The section continues by showing that this condition (no matter what the details) is poorly motivated and ultimately false: it sets unreasonably demanding standards for moral worth, satisfying it cannot play a significant role in securing non-accidentality anyhow, and it is not necessary for either non-accidentality or attributability. Drawing some lessons from these problems, Section 4 develops and defends what I call the "Value-Secured Reliability Theory of Moral Worth." On this view, an agent's action has moral worth just to the extent that its production is explained by her *value-secured reliable tie to the right*, a reliable tie to the right that is secured through the influence her person-level values have (perhaps unreflectively) on her patterns of informational access and processing.[33] As I

[32]See Arpaly (2002, 2015a, 2015b), Arpaly and Schroeder (1999, 2013), and Sher (2009) for discussions from the *Anti-Consciousness Camp*. Levy (2011a, 2011b, 2014) is among the key representatives of the *Pro-Consciousness Camp*.

[33]My notion of *value-secured reliability* is in important respects indebted to Ernest Sosa's work in virtue epistemology (2007, 2015). Sosa has long argued in some form or other that <reliability secured by something attributable to the agent> marks an important category in epistemology. Irrespective of whether Sosa is correct about the epistemology, I believe that this broad category is central to moral worth. This being said, there are important differences between me and Sosa on what makes reliability attributable to *the agent* (or something for which *the agent* is responsible), and these

show, a minimal yet indispensable role for consciousness falls out from this account, for consciousness is the integral vehicle through which an agent's action can be explained by a value-secured reliable tie to the right. Moral worth does not require that an agent be conscious of *the fact(s) in virtue of which her action is right*—it requires just that certain information was accessed and processed to produce the right action *because* of the agent's value-secured reliable tie to the right. Depending upon the agent's background values, the strength of those values, and the influence her values have had on shaping her information accessing and processing mechanisms, she can perform morally worthy actions even when conscious of nothing more robust than quite minimal sensory cues.

This account resolves our puzzle about the relationship between moral worth and consciousness: it secures both non-accidentality (*reliability*) and attributability (*value-*secured), and it gives consciousness a clear and well-motivated role. Moreover, as I suggest in the concluding remarks, the theory offers a new way of modeling

---

differences matter a great deal for the debate about moral worth and consciousness. Roughly, Sosa locates attributability to the agent in an agent's "second-order awareness" of his own reliability (2015: 79). Such "reflective competence" is the key ingredient underwriting Sosa-ian concepts such as "reflective aptness" and "aptness full well" (76): when an agent achieves the status of "reflective aptness," her belief is reliably produced *and* sustained in virtue of her appreciation of the fact that it is reliably produced. It is *the agent's* appreciation and active endorsement of her own reliability (or aptness) that makes it genuinely attributable to *her as an agent* (it is only in such cases that one's "rational nature is most fully manifest" (2015: 51)). I reject any such meta-awareness or meta-competence condition for attributability. On my Value-Secured Reliability Theory, reliability is attributable to the agent simply to the extent that it has been secured by the agent's values having shaped the inputs and outputs of the relevant cognitive and decision-making mechanisms. One's values can have this shaping effect completely unreflectively and without the agent's having any metarepresentational grasp of her own reliability. To the extent that the reliability is secured by influence from *person-level values* (as opposed to some God-hand), the reliability is properly attributable to *the agent*.

attributability that is readily extendable to a wide range of philosophically interesting phenomena.

Let us turn to clarifying the central question of this paper.

4.2 Clarification, Disambiguation, and Desiderata on a Minimal Consciousness Condition

What is the most minimal sense in which an agent must be conscious of the moral significance of her action if the action is to be morally worthy? The literature tends to approach this question by asking *whether* moral worth requires that one be conscious of the moral significance of her action. In this section, I suggest that this approach has been unhelpful, and I clarify what it is that we are looking for when we are looking for a minimal consciousness condition on moral worth.

In *Consciousness and Moral Responsibility*, Neil Levy explicitly labels Nomy Arpaly and George Sher opponents of his view that consciousness of the moral significance of one's action is necessary for moral worth (2014: 77). Moreover, Arpaly explicitly identifies herself as an opponent of Levy's view in her review of his book (2015b: 829). This might suggest that there is a well-defined dispute about whether moral worth requires consciousness of the moral significance of one's action, with the Anti-Consciousness Camp on one side and the Pro-Consciousness Camp on the other. However, the turn of phrase "S is conscious of his action's moral significance" can pick out a variety of substantively different cognitive relations between an agent and some bit(s) of information. Once we disambiguate this turn of phrase, we will see that it is far from clear that the debate between the Pro-Consciousness Camp and the Anti-Consciousness Camp is as well-defined as it might appear.

What might be meant by the claim or denial that moral worth requires "consciousness of the moral significance" of one's action? Some passages from the Anti-Consciousness Camp leave things mysterious. Arpaly writes that Huck is morally worthy but "not capable of bringing to consciousness his *nonconscious awareness*" of Jim's humanity (2002: 77). But what precisely is this relation of *nonconscious awareness*? "Conscious" and "aware" are often used synonymously in this context, so it is unclear what specific cognitive relation Arpaly has in mind. George Sher, another card-carrying member from the Anti-Consciousness Camp, similarly leaves things mysterious when he writes that "agents can satisfy responsibility's epistemic condition by accurately but unconsciously processing the *information to which they have access*" (2009: 143) and that satisfying this condition requires just that an agent have "made enough *cognitive contact*" with the evidence for an action's moral rightness (143). Given that a common referent for "conscious" is Ned Block's (1995) *access conscious* (according to which information is conscious just when an agent has the right kind of *access* to it), Sher invites confusion through his liberal use of "information to which an agent has access" while defending an Anti-Consciousness position. When situated in a debate about what bearing—if any—consciousness has upon moral worth, passages like these obscure matters.

To understand some claim or denial that moral worth requires that one be "conscious of the moral significance" of one's action, there are two things we need to know. First, we need to know what *cognitive relation* between an agent and some target information is under discussion. Is the claim about whether an agent must be *access conscious* of certain information, or about whether an agent must have *consciously and*

effortfully deliberated upon certain information, or about whether certain information must be *readily available for report*, or… ? Second, we need to know what *kind(s) of information* of which an agent can be conscious is under discussion. Whatever the relevant cognitive relation should be, is the claim about whether an agent must stand in that relation to facts about the *deontic status of his action*, or to some *non-deontic moral facts about his situation under explicitly moral concepts* (e.g., "Jim deserves respect" or "Jim is being treated unfairly"), or to the *non-moral facts upon which some moral reasons supervene* (e.g., "Jim is in pain" or "Jim is not living the life he wishes to live"), or… ?

Given the numerous candidates for both the cognitive relation and the kind(s) of information identified above, it should be clear that there are several different things a writer may have in mind in claiming or denying that moral worth requires consciousness of the moral significance of one's action. Moreover, it should be clear that some versions of this claim would be much more demanding than others. Information can pop into mind and be access conscious without the agent necessarily *consciously deliberating* upon that information. And, an agent can be aware of some *non-deontic moral facts about his situation under explicitly moral concepts* without being aware of the *deontic status of his action* (Huck might be aware that Jim is deprived of respect while being unaware that helping Jim escape is morally required). Any minimal consciousness condition should identify the most minimal combination of cognitive relation and kind(s) of information that is required for moral worth (and sufficient as far as consciousness is concerned, bracketing any other potential conditions on moral worth).

To be motivated, a minimal consciousness condition should play some nontrivial role in securing two desiderata on moral worth: namely, a non-accidental tie to the right and attributability to the person. After all, concern for these features is, I take it, what makes a consciousness condition of any sort attractive in the first place. When I inadvertently donate to some charity while unaware of what button I am pressing at the self-checkout register, my action has no moral worth because it lacks the non-accidental tie to the right. Moreover, when I am not appropriately aware of my circumstances, the morally significant features of my action are not expressive of my person-level beliefs and desires—my unwitting donation to the charity does not involve the right kind of connection to my person-level attitudes that moral worth requires. So, whatever the correct minimal consciousness condition on moral worth should turn out to be, satisfying the condition ought to play some significant role in securing non-accidentality and attributability.

In evaluating some purported minimal consciousness condition, we must ask two questions. First, we should ask: does satisfying the condition help secure non-accidentality and attributability? If the answer is "no," then we should reject it. Second, we should ask: can a consciousness condition *more* minimal than the proposed one secure non-accidentality and attributability at least as well as the purported minimal consciousness condition? If the answer is "yes," then we should reject it.

Having clarified the shape and desiderata of an adequate minimal consciousness condition, let us locate and evaluate a popular answer that shows up (surprisingly) across *both* the Anti-Consciousness Camp and Pro-Consciousness Camp.

4.3 Consciousness of the Right-Making Facts

According to this popular answer, a right action is non-accidentally tied to the right and attributable to the agent as moral worth requires only when the agent is in some sense conscious of the facts that make it right. Since such a position seems antithetical to the Anti-Consciousness Camp, it is worth taking time to show that and how representative writers from this camp really are committed to the view. The aims of this section are to locate this minimal consciousness condition across both camps, to show why no version of the requirement makes for a successful minimal consciousness condition, and to draw some lessons about non-accidentality and attributability that can inform our pursuit of a minimal consciousness condition.

Let us begin by focusing on the role given to consciousness in the work of Sher, Arpaly and Schroeder from the Anti-Consciousness Camp. It is clear that these writers reject any minimal consciousness condition with *deontic status* as the targeted information of which one must be conscious. Arpaly writes that "for an agent to be praiseworthy for an action, it is *not required* that she believe that what she does is right" (2015a: 145), and Sher also clearly denies that moral worth requires that one be conscious in any sense of the fact that one's act was morally right (2009: 143). Moreover, if we take as a datum that Huck-like actions can be morally worthy, then the minimal consciousness condition cannot have *deontic status* as the target information, as Huck is, by hypothesis, not conscious (in any sense) of this information.

It is also clear that these writers reject any minimal consciousness condition with *deliberation* as the relevant cognitive relation. In describing Huck's morally worthy action, Arpaly writes that Huck "constantly perceives data (*never deliberated upon*) that

amount to the message that Jim is a person, just like him" (2002: 76). None of these writers makes moral worth contingent upon the agent's having consciously or effortfully weighed moral reasons.

While these writers clearly reject any essential role for *deliberation* in an account of moral worth, things are less clear concerning *awareness* (understood as *wide availability of the relevant information to mechanisms in the agent*).[34] In reviewing Levy's (2014) *Moral Responsibility and Consciousness*, Arpaly writes that "some philosophers, including yours truly, have argued against the Consciousness Thesis," the thesis that "to be morally responsible for an action, one needs to be aware of those features of the action that make it good or bad" (2015b: 829). This sounds as though Arpaly is denying that moral worth requires that an agent be aware of those features of his action that make it good. However, in this same review, the cases Arpaly eventually cites as counterexamples are actually counterexamples just to *deliberative* versions of a minimal consciousness condition: they involve a jazz musician who is praiseworthy for his improvisation even though "he has no time for conscious deliberation" (2015b: 830) and a praiseworthily witty conversationalist who "does not deliberate before every funny comment" (831). Whether an agent has *deliberated* upon some information is a different matter from whether the agent was *aware* of or had *access* to that information without having consciously and effortfully deliberated upon it.

[34]Awareness in this sense roughly picks out what Ned Block (1995) refers to as "access consciousness" and what others refer to as information being "globally broadcast" in the Global Workspace model of consciousness (see Baars, 1988).

What is more, a careful look at Arpaly's descriptions of Huck suggests that awareness (albeit, not of *deontic status*) *does* play a crucial role in securing moral worth on her account. Arpaly writes (all italicizing is my own) that "while Huckleberry does not conceptualize his realization, it is [an] *awareness* of Jim's *humanity* that causes him to become emotionally incapable of turning Jim in" (2002: 10). Huck, she writes, "constantly *perceives data* (never deliberated upon) that amount to the message that Jim is a *person*, just like him" (76). And, with Schroeder, Arpaly offers the following extended interpretation of Huck's case:

> Different interpretations of the novel are possible, but one possible interpretation (not unrealistic, and one we favor) is that Huckleberry is motivated to not turn in Jim because Huckleberry intrinsically desires what is right or good via the relevant concepts, the ones that would be identified by a correct normative moral theory, and *sees that this end will be promoted by Jim's escape* … . Huckleberry *sees* that Jim's life as a slave, separated from family against his will, always forced to do what another says, and never compensated for his efforts, is lived in the absence of the *respect* he intrinsically desires everyone to enjoy. On one interpretation of the novel, *these things might all have come clearly to Huckleberry's mind*, and have weighed heavily with him emotionally because of his strong intrinsic desire that everyone be treated with *respect*. And this might well have happened without Huckleberry ever concluding that Jim's escape from slavery would be right or good—might have happened while Huckleberry self-consciously concluded that what is right or good is to return Jim to slavery (2014: 178).

A few things are worth noting about these passages. First, *awareness* is quite clearly part of the story underwriting Huck's moral worth (the passages make mention of Huck's "awareness," of the "data" he "perceives" or "sees," and of things that have "come clearly to Huckleberry's mind"). More than this, Huck is presented in such a way that he is aware of moral facts about his situation *under explicitly moral concepts*. He sees that Jim's life is lived in the absence of the *respect* he deserves, and he is aware of Jim's *humanity* or of the fact that Jim is a *person* (it is clear that "humanity" and "person" pick

out the forensic categories and not the biological categories). On this reading, Huck *is* aware of moral features of his situation under explicitly moral concepts, he just does not consciously *deliberate* upon these facts, and he never arrives at a belief that, all things considered, it is morally right to help Jim. So, the language used to capture the cognitive relation Huck has to the moral significance of his action suggests that Arpaly and Schroeder are committed not only to a role for awareness in moral worth but to a role for awareness of (non-deontic) moral facts about one's situation *under explicitly moral concepts*.

Now, I suspect that Arpaly and Schroeder would not welcome such an explicit commitment to the significance of an agent's awareness of their situation under *explicitly moral concepts*. Huck, Arpaly writes, acts "for the reasons that make [his action] right," but he "does not *know* that they are moral reasons" (2015a: 143). So, perhaps when Arpaly and Schroeder write that the facts about Jim's "humanity/personhood" or the "respect he deserves" come clearly to Huck's mind, we ought to read "respect" and "personhood" as shorthand for "the nonmoral facts upon which respect/personhood supervenes." In this case, the view on offer is that moral worth requires that one be aware of the morally relevant nonmoral facts upon which the moral rightness of one's action supervenes. Even if *this* is the view (rather than the more demanding view that requires explicitly moral conceptualizations), it turns out—somewhat surprisingly—that these representative writers from the Anti-Consciousness Camp give an important role to awareness after all.

George Sher, another member of the Anti-Consciousness Camp (see his *Who Knew? Responsibility Without Awareness* (2009)), is similarly implicitly committed to a

role for awareness in the end. Sher writes that we can correctly capture the "crucial epistemic linkage" between an agent and moral reasons that moral worth requires only by "removing that linkage from the conscious realm" (2009: 143). This may sound like a rejection of *any* consciousness condition on moral worth. However, the ways that Sher gestures toward that "crucial epistemic linkage" suggest that—like Arpaly—he intends to reject only consciousness conditions that involve deliberation or information about the deontic status of an action. Consider the following passages (italicizing is my own):

> …[A]gents can satisfy responsibility's epistemic condition by accurately but unconsciously processing the information *to which they have access* (2009: 143).
>
> When someone performs an act in a way that satisfies … any other conditions for responsibility that are independent of the epistemic condition, he is responsible for his act's morally or prudentially relevant feature if … he is unaware that the act is right or prudent despite having made enough *cognitive contact* with the evidence for its rightness or prudence to enable him to perform the act on that basis (143).

What precise combination of cognitive relation and type of information underlies the "crucial epistemic linkage" required for moral worth, according to these passages? On a natural reading, Sher seems to have the following picture in mind. Certain facts count as evidence for the rightness of an act. Presumably, these facts are the nonmoral facts upon which the moral rightness of the action supervene (the right-making features of the action). When an agent has "access" to these facts (or has made "enough cognitive contact" with these facts), the information is access conscious and, thus, made widely available for nonconscious processing by various mechanisms in the agent. Of course, the information may be access conscious without the agent's having any awareness of *how*

her awareness of these facts will influence her subsequent behavior.[35] But, this nonconscious processing of the nonmoral facts upon which the moral facts supervene allows the agent to perform the morally right action *on the basis* of those reasons that make it right. If something like this is Sher's view (he never makes precise what cognitive relation he has in mind when using terms such as "access" and "cognitive contact"), then he, too, is ultimately committed to the view that moral worth requires that one be aware of the nonmoral facts upon which the moral rightness of one's action supervenes.

So, key representatives from the Anti-Consciousness Camp are committed to a nontrivial role for consciousness in a theory of moral worth after all—consciousness, it is suggested, is what enables agents to act for right-making reasons. Is the view on offer—the view that moral worth requires that the agent be aware of the nonmoral facts upon which the moral rightness of his action supervenes—a plausible minimal consciousness condition on moral worth? I suggest that it is not. Consider the following two ways of fleshing out the condition.

On one reading of the condition, when Huck's action is morally worthy, he is aware of the nonmoral facts upon which the relevant *pro tanto* moral reasons supervene (see Arpaly and Schroeder, 2014: 166). On another reading, when his action is morally worthy, he is aware of the nonmoral facts upon which the *moral rightness* of his action supervenes—he is aware of "the reasons that make it right" (Arpaly, 2015a: 143). The

[35]Sher's way of putting this—that *the agent* has access to the information—is misleading. Better to say that sub-personal mechanisms in the agent have access to the information.

former reading cannot significantly contribute to securing a non-accidental tie to the right (and thus falters on a primary motivation for any consciousness condition on moral worth), and the latter reading sets cognitive requirements that Huck and other ordinary intuitively morally worthy agents cannot reasonably be expected to meet. Let me explain these points in turn, beginning with the *pro tanto* version.

Stipulate that *helping a friend* is *pro tanto* morally important. Is Huck's action morally worthy when, aware of the fact that he can help a friend, he helps Jim? Perhaps, but perhaps not. It had better not be the case that Huck, when aware that he can help a friend, will help the friend *no matter what the cause*. If Huck's awareness of the fact that he can help a friend would, itself, cause him to help a friend execute a broad range of morally wrong actions (steal, physically harm people for fun, et cetera), then the fact that Huck is aware of and motivated by this fact does *not* make his action morally worthy—moral worth requires a more robust non-accidental tie to the right. When one is simply aware of and moved by the nonmoral facts upon which *pro tanto* moral reasons supervene, non-accidentality is not secured, so this combination of cognitive relation and relevant information cannot supply an adequate minimal consciousness condition on moral worth.[36]

The other reading of the proposed minimal consciousness condition can (potentially) overcome this worry about non-accidentality only at the cost of setting

[36]As I suggest in a moment, it will be of no help to argue that awareness of *pro tanto* moral reasons, while not *sufficient* for securing non-accidentality, is *necessary*. Once we admit that other factors (such as an agent's background values and patterns of awareness) play an important role in securing non-accidentality, it becomes an open question whether awareness of *pro tanto* moral reasons is, itself, necessary. And, in fact, I will suggest that it is not necessary.

unrealistic cognitive requirements on moral worth. On this reading, moral worth requires that the agent be aware of the morally relevant nonmoral facts upon which the *moral rightness* of his action supervenes. Were an agent aware of and guided by these facts, it is perhaps clear how his action would have a non-accidental tie to the right (though I will give reasons to doubt this in a moment). However, note that *what makes* Huck's action right is some complex conjunction of many different facts: "Helping Jim escape 1) helps a friend 2) who is not about to cause a bunch of harm 3) and who will otherwise have his freedom impinged upon by another agent 4) to whom he never non-coercedly gave his consent to be treated that way, where 5) …". The concern, now, is that we cannot expect Huck to be aware of all of this—he cannot be aware of *the nonmoral facts in virtue of which his action is right*. That set of facts is too complex for someone like Huck (as it is for the rest of us).

One way to reinforce this point is to note that, if Huck is really to be aware of *the fact in virtue of which his action is right*, it will not do for him to simply be simultaneously aware that (1) obtains, that (2) obtains, that (3) obtains, and so on. Already, awareness of each of these different facts simultaneously is out of reach for someone like Huck. But, awareness of each of these facts does not yet make Huck aware of *the fact in virtue of which his action is right*. To be aware of *that* fact, Huck needs to be aware of the *union* or *conjunction* of the facts above: he needs to be aware that "[(1) & (2) & (3) & (4) & …] obtains." That long conjunction is the fact in virtue of which his action is right. Were Huck aware of and responsive to *that* fact, then he may well be non-accidentally tied to the right (though, again, I will give reasons to doubt this in a moment). But awareness of this long conjunctive fact is surely out of reach for someone

with Huck's cognitive abilities (as it is out of reach for any of us). It is a mistake to think that we can secure non-accidentality by simply piling on facts of which the agent must be aware (at least, if we are not to be skeptics about moral worth).[37]

We might think that the lesson is this: to secure a non-accidental tie to the right in such a way that does not set unreachable cognitive requirements, the minimal consciousness condition must require that an agent be aware of at least *some* of the (most?) morally relevant features of his situation under *explicitly moral concepts*. Perhaps this revision can solve the over-sophistication worry: rather than requiring that Huck be aware of that whole conjunctive mess of nonmoral facts upon which the relevant moral facts supervene, we might just require that Huck be aware of his moral situation under some more tractable moral conceptualization (we might require that Huck be aware that *Jim is not given the respect he deserves*). This would be disappointing for moral psychologists who have aimed to downplay the significance of an agent's moral conceptualizations (Arpaly and Sher included, I take it).

However, even if this modification might allay worries about cognitive overload, the proposed minimal consciousness condition still does not secure the requisite non-accidental tie to the right. We can see this by pointing to a problem that should also renew our reasons for rejecting the minimal consciousness condition that requires awareness of the right-making *nonmoral* facts.

---

[37]Note that it will be of no help to suggest that Huck can simply be aware of his situation under familiar Utilitarian or Kantian concepts. It is implausible to suggest that Huck and other non-ethicists are morally worthy only when aware that "my act is maximizing utility" or that "I can rationally will that the maxim of my action become universal law."

Suppose Soprano is aware of his circumstances under the relevant explicitly moral concepts. He is aware 1) that he made a promise to Baritone to kill Tenor, 2) that it is *pro tanto* good to keep one's promises, 3) that killing is *pro tanto* wrong, 4) that loyalty (of some sort) is a virtue (of some sort), and 5) that Tenor has been loyal to him.[38] Suppose, in light of all this, that Soprano does the morally right thing: he does not kill Tenor. Soprano, thus, does the morally right thing *and* he is aware of the morally relevant features of his situation under explicitly moral concepts—he satisfies the minimal consciousness condition under consideration. Is his act non-accidentally tied to the right in the way that moral worth requires?

There is an obvious version of the story in which it is not. Suppose that, in the nearby possible world where Tenor had not been "loyal" to Soprano, Soprano would not have hesitated at all to keep his promise to kill Tenor. The truth of this counterfactual makes it the case that Soprano's right action in the actual world does not have moral worth, even if, in the actual world, he was aware of the right-making features of his action and acted on the basis of them. Indeed, this counterfactual would undercut Soprano's moral worth even if he were aware, in the actual world, *that his act is morally right*. Moral worth requires a robustness that we cannot capture by simply zooming in on the information of and to which an agent was aware and responded. Awareness of the morally relevant features of one's situation (whether conceptualized under explicitly

[38]If the reader does not like "keeping promises" or "killing" as *pro tanto* rights and wrongs, feel free to substitute favorite *pro tanto* rights and wrongs.

moral concepts or not) cannot be what secures the non-accidentality that moral worth requires.[39]

Let me consider an objection. One might think that my argument against this minimal consciousness condition has been too quick. "Sure, that an agent was aware of the right-making features of his right action (conceptualized morally or not) is not, by itself, *sufficient* to secure a non-accidental tie to the right. But, such awareness is *necessary* for securing non-accidentality. When an agent is equipped with an appropriate set of background desires and motives, he may be reliably disposed to respond appropriately to his awareness of various morally relevant features of his environment. When an agent with *that* set of background desires is aware of the right-making features of some action, his performance of that morally right action *is* non-accidentally tied to the right. So, awareness of the right-making features of one's action *is* an integral part of securing non-accidentally—it's just not the *only* integral part."

[39]Some might draw a different lesson. Paulina Sliwa (2016, see 401 in particular) argues that moral worth requires that the agent *know* that his action is morally right. *Knowledge*—as opposed to mere awareness—has a counterfactual robustness. One knows that p only if one's belief in p is at least somewhat secure. We might think that the problem with Soprano above is that, even if he *believes* or is *aware* that his action is morally right, he does not *know* that it is morally right (after all, he would change his mind in a very nearby possible world). We might think that Soprano's action is non-accidentally right only if he *knows* that his action is morally right. It *might* be sufficient for moral worth if I both know that my action is right *and* my action is caused in the right way by my moral knowledge. Specifying the "caused in the right way" clause would require some work, but I'll assume that this can be done in a way that would make for a plausible sufficient condition on moral worth. Still, even if this would make for a sufficient condition on moral worth, there is no reason to think that it is a *necessary* condition on moral worth. We should think this is a necessary condition on moral worth only if moral knowledge is the *only* way to secure non-accidentality and attributability, and I'll argue in Section 4 that we can secure these through the satisfaction of a more minimal consciousness condition (one that does not involve moral knowledge).

I agree that appealing to an agent's background desires and motives is an important part of explaining an agent's non-accidental tie to the right. However, once we admit that background desires and motives are part of the story for securing non-accidentality, there is an open question about whether non-accidentality can be secured with awareness of information much less robust than the *nonmoral facts upon which the relevant moral facts supervene*. It might turn out, for instance, that, against an appropriate set of background desires, an agent could reliably perform right actions through awareness of quite basic sensory cues rather than through awareness of *the nonmoral facts upon which the relevant moral facts supervene*. A good conversationalist, for instance, might reliably perform the morally right action of *waiting a beat in conversation* when aware of nothing more robust than sensory content of furrowed brows or subtle shifts in inflection, content which is then nonconsciously processed against his background desires. This agent might sincerely be unaware of *the facts in virtue of which his action is right* (he may sincerely be unaware of, say, *the fact that his interlocutor wishes to finish her point* and other relevant facts), yet his action—guided by an awareness of quite minimal sensory cues—could be non-accidentally tied to the right. I argue in the following section that something like this is, in fact, true. If I am right, then consciousness of the facts in virtue of which one's action is right cannot be a necessary condition on non-accidentality (and so concern for non-accidentality gives us no reason to accept this condition as the minimal consciousness condition).

Here is a different objection. "Okay, *maybe* an agent's action can be non-accidentally tied to the right even when the agent is not conscious of the right-making features of her action. However, moral worth requires both non-accidentality *and*

attributability. For the morally significant features of an action to be *attributable* to the agent in the way that moral worth requires, surely the agent must be conscious (in some sense) of the right-making features of her action. How else could we pin the moral significance of the action to *the agent*? Since moral worth requires attributability, the minimal consciousness condition must involve an agent being conscious (in some sense) of the right-making features of her action."

Concerns about attributability are, in fact, among the principle motivations for many in the Pro-Consciousness Camp (see, for instance, Levy (2014: 87-108)). It is utterly intuitive that the morally significant features of an action are attributable to the agent only if the agent is in some sense conscious of those features—for, in being conscious of those morally significant features, *the agent* is importantly connected to them. However, it is a mistake to suggest that this is the *only* way that an agent can be relevantly connected to the moral significance of her action. The morally significant features of an action can be perfectly well expressive of and attributable to an agent without the agent's being in any sense conscious of those features.

It will be helpful to briefly examine Neil Levy's minimal consciousness condition and his discussion of consciousness and attributability (this will also give us an opportunity to examine a consciousness condition from the Pro-Consciousness Camp). I will argue that Levy's minimal consciousness condition, like those already considered, fails *at least* because it demands excessive sophistication from ordinary, intuitively morally worthy agents. But even once we set those concerns aside, we will see that one of the central motivations for Levy's minimal consciousness condition depends on a

misguided picture of what attributability requires. Let us turn to Levy's minimal consciousness condition.

On Levy's official statement of his minimal consciousness condition (the "Consciousness Thesis"), "in order to be morally responsible for their actions, agents must be conscious of facts that explain the valence of its moral significance" (2014: 37). While this might look like just another statement of the view earlier attributed to Arpaly and Sher (namely, the view that moral worth requires that the agent be aware of the nonmoral facts in virtue of which her action is right), Levy's understanding of "conscious" changes the view a bit.

As Levy uses the term "conscious," information is conscious when it is "online" (i.e., actually guiding the agent's behavior) and "personally available"—that is, "when the agent is able to effortlessly and easily retrieve it for use in reasoning" (2014: 33). And, information is "available for easy and effortless recall if it would be recalled given a large range of ordinary cues: no special prompting (like asking a leading question) is required" (2014: 34). So, Levy's minimal consciousness condition relies on the cognitive relation of *reportability*—whether an agent's right action is morally worthy crucially depends on whether the agent can, at the time of acting, *bring to mind* the features of his action which make it right and to which he actually responds, even if the agent does not actually deliberate upon that information or subject it to occurrent focus. I may perform a rescue that is so cognitively demanding that all my attention is given to mechanical operations ("turn this dial 90 degrees now, flip switches B and C now, …"). I am not then thinking about the fact(s) in virtue of which my action is right (e.g., the fact that I am rescuing some people). However, I can satisfy Levy's minimal consciousness condition

(and my action can be morally worthy) because, at the time of performing the rescue, it is true of me that I *could* easily report the facts that guide my behavior and in virtue of which my action is right (see Levy, 2014: 34).

Levy is not without philosophical company in endorsing a reportability requirement on the kind of agency needed for moral worth. John Doris writes that

> Where the causes of her cognition or behavior would not be recognized by the actor as reasons for that cognition or behavior, were she aware of these causes at the time of performance, these causes are *defeaters*. Where defeaters obtain, the exercise of agency does not obtain (2014: 64-65).

Similarly, Fischer and Ravizza, in clarifying what it is to act for a reason in the way relevant to moral worth, approvingly reference Robert Audi (1986) and suggest that

> In order for an agent to act for a reason, r, it is not necessary that the person deliberate and formulate r as his reason for acting; roughly speaking, it is enough that he would give r as the reason for his action, if he were asked for an explanation (1998: 64).

So, on this view, the minimal consciousness condition on moral worth requires that the agent easily be able to report the morally relevant facts that actually guide her behavior and in virtue of which her action has its moral significance. Is this a plausible minimal consciousness condition?

I suggest that it is not. Like the other conditions considered, this minimal consciousness condition sets overly-sophisticated requirements on moral worth—we cannot reasonably expect ordinary, intuitively morally worthy agents to satisfy this condition. Levy writes as though it is easy for agents, even children like Huck, to satisfy this condition:

> [Huck] rejects his [explicit moral] principles altogether, but continues to guide his behavior by reference to the facts upon which moral principles genuinely supervene. *All of this he does consciously*. He lacks only the concepts to

perspicuously describe what he does; he lacks nothing in the way of *awareness of what he does (2011a: 260).[40]

The suggestion, here, is that, were we to interrupt Huck mid-action and ask him what he is doing and why, he would easily be able to report the facts to which he responds and in virtue of which his action is morally right (he just would not use explicitly moral concepts in his explanation).

But this interpretation of Huck is highly implausible. *The fact that explains the moral valence of Huck's action* (or, the fact that explains the moral rightness of his action) is, as we have already seen, some long, unwieldy, conjunctive fact. To satisfy Levy's condition, Huck would need to be capable of bringing to mind what I had previously formalized as "[(1) & (2) & (3) & (4) &…]". And, surely, we cannot expect ordinary, morally worthy agents to be able to bring a fact like this to mind. Here is another way to put the point: ethics is hard. If we want there to be ordinary agents whose morally right actions can be morally worthy, it had better not be the case that moral worth requires the ability to bring to mind the total set of nonmoral facts that jointly explain the moral rightness of an action.[41] Levy's condition cannot be the minimal consciousness condition.

---

[40]Note that "*awareness" just picks out consciousness in Levy's "personal-availability" sense (see Levy, 2011a: 247).

[41]*Psychology* is hard, too. We cannot reasonably expect ordinary, intuitively morally worthy agents to be able to bring to mind the full set of morally relevant information to which they respond. Indeed, Doris—who endorses a reportability requirement on moral worth as we see above—argues for agency skepticism precisely on the grounds that we cannot reasonably expect agents to be able to bring to mind the full set of information to which they respond (see Doris, 2014: 41-77).

Now, as mentioned earlier, one of the central motivations for a consciousness condition of any sort is that it is intuitive that consciousness plays an important role in securing attributability. And, indeed, Levy suggests that satisfying his purported minimal consciousness condition is required for attributability:

> In order for our actions to express our evaluative agency [in the way required for moral worth], *we* must be able to assess the moral significance of our actions for consistency with the beliefs, desires, goals, and commitments (and so on) that together constitute our evaluative agency … (2014: 107).
>
> When agents are aware neither of the mental states that are responsible for the moral significance of an action, nor of that moral significance itself, … *the agent* cannot assess either for consistency or conflict with their personal-level beliefs. The action therefore does not express their evaluative agency [in the way that moral worth requires]" (2014: 102).

Levy offers an intuitive picture of how attributability works. On any plausible view, if X is to be attributable to *me*, it must be the case that facts about *me*—about my person-level beliefs, desires, goals, or values—are a sufficiently significant part of the explanation about why X obtains. What Levy tells us is that one's person-level attitudes are a sufficiently significant part of the explanation for why one ended up acting *rightly* only in those cases where one can bring to mind the facts in virtue of which one's action is right. If *the person* wasn't even in a position to bring to mind and assess against her person-level values the facts in virtue of which her action was right, how (goes the thought) could facts about *the person* have played any significant role in explaining why a morally right action (rather than a neutral or wrong action) was performed?

However, reportability is not required for roping person-level attitudes into a plausible explanation of why a right action was produced. As I will elaborate in the next section, person-level attitudes can play a significant role in making certain information from one's environment widely available to mechanisms in the agent, and they can also

play a significant role in shaping how that information from one's environment is processed to output a decision, and all of this can happen *without* the agent being in any position to report that information or the influence her person-level attitudes have had on its availability and processing. If the concern about attributability is to find a way of linking person-level attitudes to an explanation of why a right action was performed, then reportability is not necessary for securing attributability, and so concern for attributability gives us no reason to accept Levy's reportability condition on moral worth.[42]

It is worth pausing to consider where things stand concerning consciousness, non-accidentality, and attributability. We have seen that a popular view represented in the Pro-Consciousness Camp and even in the Anti-Consciousness Camp is the view that securing non-accidentality and attributability requires that the agent be conscious (whether in the "awareness" sense or the "reportability" sense) of the facts in virtue of which her action is morally right. But I have shown that satisfying such a condition requires *much* more cognitive and moral sophistication than we can reasonably expect of

[42]Levy might argue that wide availability to mechanisms in the agent *implies* reportability. As Block (1995) notes, reportability is often a good heuristic for determining when information is "access conscious" or widely available to mechanisms in the agent (see Block, 1995: 231). But wide availability does *not* imply reportability. One reports information in specific contexts and in response to specific queries. Levy's reportability condition does not *simply* require that the agent be able to point at morally relevant information in her environment under *whatever* description without any understanding of how that information might guide her behavior. Rather, it requires (roughly) that the agent be able to bring to mind the morally relevant information and understand the role that that information plays in guiding her behavior *specifically* in the context of a query about what she is doing and why. It is clear that information can be widely available to mechanisms in the agent (and thus interact with person-level attitudes and play a role in guiding behavior) even when the agent would not bring the information to mind in that specific context of a query about what she is doing and why.

ordinary, intuitively morally worthy agents. Already, this oversophistication worry should makes us reject this condition as the minimal consciousness condition.

Of course, we could be steadfast in endorsing this as the minimal consciousness condition and just become skeptics about moral worth. To a degree, that is the route taken by Neil Levy (2011b), John Doris (2014), and Michael Zimmerman (1997). But we should take the skeptical road only if it really is true that non-accidentality and attributability require satisfying this deeply demanding condition. I have already presented doubts about what role the satisfaction of this condition could even play in securing non-accidentality. Moreover, I have gestured toward (and will soon expand upon) reasons for believing that non-accidentality and attributability can be secured without satisfying the condition. Bearing in mind these considerations, I now wish to offer an account of moral worth that gives a plausible and significant role to consciousness, that avoids the oversophistication worries, and that secures non-accidentality and attributability without requiring that the agent be conscious in any sense of the facts in virtue of which her action is right.

4.4 The Value-Secured Reliability Theory of Moral Worth

Here is the proposal.

The core idea is that, since moral worth requires non-accidentality and attributability, what moral worth ultimately involves is a reliable tie to the right that is sufficiently secured by person-level values. I suggest that an agent's action is morally worthy just to the extent that its production is explained by her value-secured reliable tie to the right. Call this the "Value-Secured Reliability Theory" of moral worth.

Obviously, “value-secured reliability” is a technical term. So, what is it, and why should it matter to moral worth?

An agent has a value-secured reliable tie to the right (in a context) just to the extent that

1) she is reliably disposed to perform the morally right action in situations like the one under consideration, and
2) that reliability is secured or explained by the influence her person-level values have on shaping (perhaps unreflectively) the inputs and outputs of her cognitive and decision-making systems. (That is: *because* of her person-level values, the agent is reliably fed informational inputs from her environment that, when processed, reliably output morally right action).

An agent’s *particular* action is explained by her value-secured reliable tie to the right just to the extent that

1) her awareness of the information from her environment that served as inputs to her action is explained by the fact that her person-level values have shaped her patterns of awareness to make her reliably aware of information that, when processed alongside her person-level values, reliably yields right actions in such situations; and
2) the decisional output of the processing of that information is explained by the fact that, given the agent’s person-level values, she will reliably perform right actions when *that* kind of information is processed.

In a moment, I will walk through an example to illustrate the theory and to show how it yields a minimal consciousness condition. But let me first clarify the theory and explain how it is well-equipped to secure non-accidentality and attributability.

Since the account states that an action is morally worthy *to the extent* that it is explained by a value-secured reliable tie to the right, it captures the intuitive idea that moral worth comes in degrees. Without resolving the geometry of moral worth, we can note that *how* morally worthy an action is will be a function of: the *extent* to which the action is explained by the value-secured reliable tie to the right (as opposed to being explained by self-interested motives, or by the nice smell of the baking bread[43]), of the extent of the *reliability*, of the extent to which the reliability is explained by the agent's values, and of the extent to which the reliability is tied to the *right* (as opposed to the more or less nearly right).

The account secures *non-accidentality* because of its reliability condition. It secures *attributability* because the reliability is, itself, secured or explained by the influence of person-level values (as opposed to some God-hand tinkering with the environment). The Value-Secured Reliability Theory thus has a natural way of handling the key desiderata of moral worth.

So, what minimal role—if any—is given to consciousness on the Value-Secured Reliability Theory? And what sense of "consciousness" is relevant to the theory? The sense of "conscious" that is integral to achieving value-secured reliability is roughly what Block (1995) has in mind by "access consciousness" and what Baars (1988) has in mind

[43]See Doris (2014) for a discussion of a range of influences that might undercut person-level explanations of behavior.

by "global broadcast." Information is access conscious or broadcast in this sense when it was at least momentarily attended to and thereby made widely available for consumption by sub-personal mechanisms in the agent (to a decision-making system, a judgment-making system, et cetera). Once the information is made widely available, it may be processed by mechanisms in the agent in ways that guide her behavior and judgments without the agent's being aware of *how* her behavior and judgments have been so shaped. In what follows, I illustrate the theory and show how person-level values can shape the inputs and outputs of a decision-making system in such a way that makes a right action both non-accidental and attributable to the agent, even when the agent is not aware of the right-making facts.

Note, first, that an agent's personal-level attitudes can play a significant role in determining what information is and is not made widely available. This is borne out in commonsense observations: as Arpaly notes, the person who cares about cleanliness is more likely to notice the dust than the person who does not care about cleanliness (2002: 83). It is also uncontroversial in discussions of global broadcast theories of consciousness that person-level attitudes can affect what information is broadcast (or made widely available): such top-down influences on patterns of awareness occur "when one's goals or interests direct attention to one aspect of the stream of current sensory processing rather than another" (Carruthers, 2011: 48). Here is, thus, already one meaningful interaction between an agent's person-level attitudes and the information from her environment of which she is aware.

Now, when an agent's person-level attitudes play this role in determining what information is and is not momentarily attended to and thereby made widely available, the

fact that one's person-level attitudes play this role need not itself be widely available. This point should be uncontroversial. A good conversationalist's background values (that, say, people feel heard, or that people are not made to feel unnecessarily uncomfortable) may direct her attention to various features of her environment (subtle facial cues, et cetera) even while she is altogether unaware of the role that her goals play in guiding her attention.

Once some information is attended to and made widely available, that information can be consumed by a decision-making system that has direct access to some of the agent's personal-level attitudes (her beliefs and goals) without those person-level attitudes *themselves* needing to be attended to. As Carruthers notes, "we should expect … decision-making systems to be capable of accessing some of the subject's beliefs and goals directly, without the latter needing to be reached through global broadcast" (2011: 53). Were this *not* true, an agent would need to attend to her relevant standing beliefs and goals any time she performed some action that drew on those beliefs and goals—agents would be mentally exhausted by the end of breakfast.

So, the decision-making system can output a decision by processing the access conscious information from her environment and the agent's person-level attitudes together. Insofar as the decision-making system outputs a decision by processing both the access conscious information and the agent's person-level attitudes together, the agent's person-level attitudes play an important role in determining how the agent responds to information. To use a simple example: the decision-making system may draw directly on my goal to warm up, so that, when imagistic content pertaining to the mug of green tea and the glass of ice water on the table are briefly attended to (when I see both on the

table), the decision-making system processes all of this and outputs a decision to grab the mug of tea rather than the glass of water.

Of course, *that* and *how* the decision-making system has processed some access conscious information alongside some person-level attitudes is not necessarily, itself, access conscious. In the example above where I grab the mug of tea rather than the glass of water, I need not be aware of or have attended to the fact that my goal of warming up was processed alongside sensory information pertaining to the drinks on the table. Now, in this case, I could probably tell you correctly after the fact why I grabbed the tea instead of the water: "I wanted to warm up." But, importantly for present purposes, not all outputs of the decision-making system are like that. A good conversationalist—a good listener and conversation partner—pauses at the appropriate moments, changes subjects at the appropriate moments, interjects at the appropriate moments, and so on. Moreover, some of the best listeners/conversationalists are unaware of the extent to which their background values (that, say, people feel heard, or that people are not made to feel unnecessarily uncomfortable) shape their conversational patterns. One might correctly tell Roscoe after seeing him navigate several especially emotionally loaded and complicated conversations: "wow, Roscoe—the way you do x, y, and z in conversation is perfect—I'm going to start trying to do x, y, and z." And Roscoe might sincerely respond: "huh, are x, y, and z things that I do?".

Let me continue with the example of Roscoe the Conversationalist to illustrate the joint significance of the considerations adduced thus far. Roscoe's person-level values may reliably make widely available relevant sensory cues: content pertaining to subtle facial expressions and speech inflections reliably correlated with an interlocutor's feeling

of discomfort are reliably made available to mechanisms in Roscoe in part because of his background person-level values (including, say, the desire that people are not made to feel unnecessarily uncomfortable). Moreover, when sensory cues like *those* are made widely available to Roscoe, a right action—to, say, wait a beat in the conversation—is reliably produced from the unconscious processing of those sensory cues alongside his person-level values.

When Roscoe waits a beat in conversation, his action is morally worthy: it is explained by a value-secured reliable tie to the right. It is no accident that he waits a beat at the appropriate moments, and this non-accidentality is explained to a large degree by the influence that his person-level values unreflectively have on shaping his access to and processing of information.

In this test case, Roscoe need not be conscious (in any sense) of the fact that waiting a beat right now (or at a moment like this) is the right thing to do. He may not even be conscious (in any sense) *that* he waits a beat, and he may have no especially nuanced, readily articulable beliefs about how conversations ought to go.

Moreover, Roscoe need not be conscious (in any sense) of the fact(s) in virtue of which it is morally right to wait a beat in the conversation. He may not be conscious (in any sense) of the fact that (say) Nina wishes to add a qualification to her most recent point. The wide availability of various sensory cues (furrowed brows, voice inflections, et cetera) can, itself, reliably lead to the outputting of a morally right decision (to wait a beat) given Roscoe's background values. Access consciousness played a critical role in enabling Roscoe's person-level values to guide his response to his environment, but

access consciousness of *the fact(s) in virtue of which his action is right* need not have played any role.

On the Value-Secured Reliability Theory, an agent need only be aware of (or, have momentarily attended to and thereby have widely available) *whatever* minimal information from her environment will, in virtue of being processed against her particular set of background values, enable her to reliably perform right actions. What *specific* information an agent must have widely available if her action is to be explained by a value-secured reliable tie to the right may vary depending upon her values, the strength of those values, and how morally complicated the action is. But, as we see in the case of Roscoe, this information may sometimes involve nothing more robust than basic sensory cues (ones about which the agent may be in no position to say anything concerning how they guided his behavior).

In Huck's case, it is perfectly well imaginable that his decision to keep helping Jim is explained by a value-secured reliable tie to the right. *Because* of his person-level values, in situations like this one, Huck is reliably made access conscious of information that, *when* access conscious, reliably leads to right action. When Huck's action is explained by his value-secured reliable tie to the right, his action is both non-accidentally right and attributable to him. Access consciousness plays a necessary and significant role in this, but none of this requires that Huck be aware of the set of facts in virtue of which his action is right. The Value-Secured Reliability Theory offers a successful account of moral worth that gives a motivated and delineated role to consciousness without requiring excess cognitive sophistication.

To further spell out the implications of the Value-Secured Reliability Theory, it will be useful to respond to a potential objection. Consider Kant's famous self-interested grocer. The grocer's business decisions are explained by his desire to maximize profits—it just so happens (we stipulate) that he will maximize profits if he regularly makes business decisions that are morally right. It is intuitive that the grocer's decisions are not morally worthy. This might be a problem for the Value-Secured Reliability Theory. After all, the grocer reliably makes business decisions that are morally right, *and* he reliably makes the decisions he makes because of a person-level value (the value he places on making as much money as possible). He thus appears to have a value-secured reliable tie to the right. Is the theory committed to saying that his business decisions are morally worthy?

It is not. Let me sketch a tempting but ultimately unpromising response first, and then I will offer the correct response.

It is tempting to argue that the grocer's decisions are *not* reliably tied to the right. Presumably, there are nearby possible worlds where his business interests do not align with morality, and in such worlds, the grocer would *not* make the morally right decisions. Moreover, in situations outside of business, he is perhaps unlikely to treat people fairly. The Value-Secured Reliability Theory would then yield the intuitively correct verdict that his actual business decisions are not morally worthy, since he does not have a value-secured reliable tie to the right.

But, Kant's grocer case is interesting because his behavior *does* seem to be reliably tied to the right (at least in some domain). Any theory of moral worth should allow that an agent can perform genuinely morally worthy actions in *some* domains while

having moral blind spots that prevent him from performing morally worthy actions in certain other domains (imagine the professor who performs genuinely morally worthy actions within the department but is callous with family). We do not want the result that an action is morally worthy only if the agent is reliably tied to the right across *all* domains. While there is certainly an important issue concerning how to demarcate the relevant domains, there is not the space to settle the issue here. So it would be good to have a different response to the case.

A different response draws on the role of *explanation* in the Value-Secured Reliability Theory. Recall that, on the theory, it is not enough that an agent simply *have* a value-secured reliable tie to the right. A particular action is morally worthy just to the extent that its production is *explained* by that value-secured reliable tie to the right. The grocer may well have a value-secured reliable tie to the right. But, when he gives the eight-year-old correct change on Friday as morality requires, this transaction is not *explained* by his value-secured reliable tie to the right. There is a competing and better explanation that undermines this one: namely, the transaction is explained by a value-secured reliable tie to *maximal profits*. An explanation of the grocer's transaction that references a reliable tie to the right is outstripped by (rather than supported or amplified by) an explanation that references a reliable tie to maximal profits. The moral explanation is undercut, as the tie to maximal profits is not plausibly "a part or a symptom" of the tie to the right (see Sturgeon, 1992: 100).

Contrast the case of the self-interested grocer with Roscoe, our morally worthy conversationalist. Both characters have value-secured reliable ties to the right. But *Roscoe's* particular right action (to wait a beat in conversation) is best explained by his

value-secured reliable tie to the right. Roscoe's values have shaped his patterns of informational access and processing to reliably produce right actions in situations like the one he is in—*that* is what most fully explains why he came to have access to the information that served as inputs to his decision and why the processing of that information produced a right action. We could explain Roscoe's action with nonmoral language. We could say that his action is explained by a reliable tie to actions that promote his relevant values (his values that people feel heard, or are not made to feel unnecessarily uncomfortable, or the value he places on the persons around him). But this nonmoral explanation does not undermine the moral one—it *supports* it (see, again, Sturgeon, 1992). The tie to the promotion of such values (insofar as these values are for things that matter morally) *is* plausibly a part or symptom of a tie to the right.

The Value-Secured Reliability Theory is *not* committed to saying that wherever there is a reliable tie to the right, there is moral worth. It shows, instead, that an action is morally worthy just to the extent that it is *explained* by the agent's value-secured reliable tie to the right.

4.5 Conclusion

Let us take stock of what this paper has accomplished. I have shown, first, that a popular view about the minimal consciousness condition on moral worth is mistaken. Across the Pro-Consciousness Camp and even the Anti-Consciousness Consciousness, there is widespread endorsement of the view that a right action is attributable to the agent and non-accidentally right as moral worth requires only when the agent is conscious (in *some* sense) of the facts that make it right. However, it is not clear that such consciousness can even play any significant role in securing non-accidentality. Moreover,

actually being conscious of *the fact(s) in virtue of which one's action is right* is much more difficult than writers typically suggest. *Were* it true that non-accidentality and attributability required consciousness of such facts, skepticism about moral worth should follow. But non-accidentality and attributability do not require such consciousness. Non-accidentality and attributability are secured when, *because* of the agent's value-secured reliable tie to the right, certain information from her environment was made widely available and then processed to produce a right action. Consciousness is the vehicle through which an agent's value-secured reliable tie to the right is brought to bear upon her response to her environment. As we saw with Roscoe the Conversationalist, an agent's right action can sometimes be explained by a value-secured reliable tie to the right (and, thus, morally worthy) even when the agent is conscious of neither the fact that he acts rightly nor of the fact(s) in virtue of which his action is right, and even when he is incapable of explaining how his values might have shaped the production of his right action. The role of consciousness in a theory of moral worth is no bigger nor smaller than the role it plays in making our actions explained by a value-secured reliable tie to the right.

Beyond resolving our original question about the relationship between consciousness and moral worth through an independently plausible theory of moral worth, the discussion of the Value-Secured Reliability Theory points toward a new way of modeling attributability that is readily extendable across sub-disciplines of philosophy. Epistemologists may care about when it is attributable to *me* that my belief was reliably produced (see Sosa, 2007, 2015). The philosopher of art may care about when it is attributable to *me* that my painting is a good piece of art (see Wolf, 2015). I have shown

that it is false that a right action is attributable to *me* only if I am conscious of the facts that make it right—similarly, we should think that is false that a reliably produced belief is attributable to *me* only if I am conscious of the facts that make it reliably produced, and false that the good artwork is attributable to *me* only if I am conscious of the facts that make it a good artwork. Because of the influence that an agent's person-level values have on shaping her patterns of informational access and processing, an agent might be reliably tied to the right, or to making good artistic decisions, or to forming beliefs through truth-reliable processes. When an agent's right action (or artistic decision, or reliably produced belief) is explained by such a value-secured reliable tie to the right (or to the artistic good, or to the truth-reliable), the success is both non-accidental and *attributable* to the agent. There is, of course, a legitimate question about whether and why we should care that some feature is attributable to the person. Should it matter to epistemologists whether, beyond having my belief be reliably produced, it is reliably produced *because* of the influence that my person-level values (rather than, say, evolutionary pressures) have on my patterns of informational access and processing? There is no space to resolve this question here. But the Value-Secured Reliability Theory of moral worth helps us understand how person-level values can be coopted into the explanation of a success even when the agent has limited or no conscious access to the success or to the facts that make for the success.

CHAPTER 5

EPISTEMIC RESPECT AND CREDIBILITY EXCESS

5.1 Introduction

You are at a party, and awkward milling about unfortunately lands you in conversation with Fred, a friend of a friend. Fred eagerly tells you that he has been "getting really into spirituality recently." You smile politely. He tells you about how he gave up the Catholicism on which he had been raised and how, after a few setbacks in his personal life which he is not going to get into right now, something about "the wisdom of Far East thought" started to resonate with him. He begins sharing his interpretations of his favorite passages from the Daodejing, and, for each interpretation, he asks you with maximal earnestness whether you agree with his reading. "Yeah, maybe the passage means that," you tell him while scanning the room for a way out of the conversation. Things carry on like this, and it is abundantly clear, as Fred offers his interpretation of chapter 38, that he is interested in specifically *your* affirmation of his readings—he does not check in with the other people nearby, and he quietly rejoices at each of your (disinterested and mostly monosyllabic) affirmations. You are the only person of East Asian descent at the party, and Fred—who is entirely unresponsive to the clear cues that you neither know nor care about the Daodejing—treats you as the resident Laozi expert.

Fred's exoticism and generalizing of "Far East thought" are gross. But I argue that there is also an important sense in which his judgments about *you*—about your level of expertise in Daoist literature—disrespect *you* as an agent with a rational subjectivity of your own. You command a form of respect that requires that people's judgments bearing

upon your rational activity satisfy basic moral and epistemic standards. Fred's evidence-resistant judgment fails to accord you such respect.

Cases of disrespectful credibility excess are underexplored in the literature on epistemic injustice.[44] On the dominant perspective, the significance of credibility excess is either overlooked or seriously miscaptured. It is suggested that "while credibility excess may (unusually) be disadvantageous in various ways, it does not undermine, insult, or otherwise withhold a proper respect for the speaker *qua* subject of knowledge" (Fricker, 2007: 20), or credibility excess is said to wrong its recipient *qua* knower only when it renders them epistemically arrogant (Fricker, 2007: 21; Medina, 2011: 18). But judgments involving excess credibility, when resistant to the facts about some agent *qua* knower, can be alienating and disrespectful insofar as they fail to be responsive to that person's subjectivity. The project of this paper is to develop cases like Fred's and to use them to motivate an account of both epistemic respect (respect for someone in her capacity as a knower) and respect more generally.

Here is how things proceed. Section 2 locates and rejects the dominant position on credibility excess: namely, the position that either credibility excess never wrongs, harms, or disrespects its recipient *qua* knower or, if it ever does, it is only by rendering him epistemically arrogant. Section 3 zooms in on the opening case and related cases to begin explaining credibility excess as a potential manifestation of disrespect. Drawing some

[44]Emmalon Davis (2016) is a notable exception. Though Davis does not put the issue in terms of "respect," she targets a variety of cases in which giving someone too much credibility can harm the recipient. For reasons that are developed in Section 3, I am sympathetic with Davis' discussion, though I argue that credibility excess can wrong its recipient in a broader range of cases than her explanation would suggest.

lessons from this discussion, Section 4 defends an original account of respect: the Value-Secured Reliability Theory of respect. Respect, I argue, has an underappreciated epistemic dimension that this theory is positioned to capture with a new level of precision. It requires that our judgments about persons—*because* of our values—be reliably attuned to them as the particular agents they are.

5.2 The Dominant Perspective on the (In)significance of Credibility Excess

This section clarifies the dominant perspective on credibility excess—as located in Fricker (2007) and Medina (2011)—and shows that it overlooks the capacity for credibility excess to disrespect its recipient. I begin with Fricker's discussion.

Fricker's work on testimonial injustice targets a form of undermining or disrespecting an agent *qua* knower. As she presents it, "a speaker suffers a testimonial injustice just if prejudice on the hearer's part causes him to give the speaker less credibility than he would otherwise have given" (2007: 4). Though credibility deficit can disrespect someone *qua* knower, Fricker suggests that credibility excess cannot (20). This asymmetry may be utterly intuitive. We usually hope that people take us to be credible. Moreover, we do not think that someone who receives more than his fair share of (say) money because of his social identity is, himself, a victim of economic injustice—more plausibly, his excess wealth may signal that there is *an* economic injustice, but the victims are those systemically deprived of certain economic goods.

As it turns out, Fricker qualifies her suggestion about credibility excess. She considers a case of what we might call "epistemic affluenza," where someone is given so much credibility excess that various epistemic virtues are put out of reach (2007: 20). As Fricker sees it, this agent may have been wronged *qua* knower, but she is quick to register

two points. She suggests, first, that such cases are "semi-fanciful" and rare,[45] and, second, that while constant credibility excess can, cumulatively, wrong someone *qua* knower, no individual judgment can (21).

José Medina (2011) holds a similar position, though he argues that Fricker understates the significance of one-off judgments of credibility excess. He first agrees with Fricker about what *can* make credibility excess wrong its recipient—it can make him epistemically arrogant—but he suggests that even isolated judgments of excess credibility can do this (17). He secondly suggests that Fricker overlooks the extent to which epistemically privileging one person because of his social identity is—whether conceptually or historically—inseparable from undercutting the credibility of persons without that social identity (see 17-19, and, for a similar point, see Anderson, 2012: 170). As the current project focuses on the potential for credibility excess to disrespect its *recipient*, I will ignore the latter suggestion. Let me say a bit about the former.

First, I tend to agree with Fricker that individual judgments of credibility excess—without some underlying pattern—are unlikely to render their recipients epistemically arrogant. Indeed, lots of judgments involving credibility excess go unexpressed and have no consequences for the relevant parties, and, even when expressed, such judgments in isolation are unlikely to instill epistemic vices (understood as reasonably stable features of a person's psychology).

---

[45]It is not obvious that we should agree with Fricker that such cases are fanciful or rare. Take a person in a position of great power (a CEO or politician) who is so hostile toward expressions of opinion contrary to his own that those around him simply abet his ever increasingly out of touch worldview. Such a case seems far from rare.

But, more importantly, if our aim is to determine whether credibility excess can *disrespect* (as opposed to *harm*) its recipient *qua* knower, then facts about whether it makes someone epistemically arrogant are the wrong place to look. Whether a judgment disrespects is not a function of the consequences of expressing it, but a function of something internal to it: its content and production. If I privately judge that you are worthless, this judgment is *itself* disrespectful, even if my attitude goes undiscovered and is inconsequential in your life. Even if my expression of that judgment makes your life go better (e.g., I tell you that you are worthless, and this fuels you to finish writing that book), the judgment remains disrespectful. Fricker and Medina's discussion of the potential significance of credibility excess is orthogonal to questions about its capacity to disrespect.

Let us return to our opening case to begin targeting the features that can make credibility excess disrespectful.

5.3 Targeting Credibility Excess as Disrespectful

In our opening case, Fred gives excessive credibility to your (disinterested and monosyllabic) affirmations of his Daodejing readings. His prejudice makes him utterly resistant to his evidence about you *qua* knower. Such cases are ubiquitous. The basic recipe involves one agent who—from vice—is seriously resistant to the evidence about another *qua* particular epistemic agent and consequently gives her excessive credibility. Such failures of responsiveness are seriously lacking in respect.[46]

---

[46]Emmalon Davis (2016) discusses some similar cases and illuminatingly points to various ways in which credibility excess judgments that bypass a person's subjectivity because of identity-prejudice can harm them—in particular, such judgments can harm them by making their acceptance in some knowledge-exchange contingent upon their

We might think that our negative reactions to such cases are not responses to *disrespect* but, rather, to the simple fact that (say) Fred is prejudiced. But this thought forgets the individual *victims* of prejudice. Were Fred's prejudice the only thing of moral significance here, then there would be no morally significant difference between the case as presented and some case where Fred is just sitting alone at home with prejudiced beliefs. Prejudices can prevent persons from being recognized for the agents they are—they can thus prevent agents from enjoying one basic form of respect. The significance of the opening case extends beyond Fred's being prejudiced.

Indeed, while prejudicial credibility excess may be the most salient form of disrespectful credibility excess, disrespectful credibility excess need not involve identity-prejudice (just as disrespect, more generally, need not involve prejudice). A person might just be viciously self-absorbed and, as a result, disrespectfully fail to be responsive to the agents around him. Without prejudice but through self-absorption, I may mistakenly judge (in an evidence-resistant manner) that those around me at the party have sufficient background knowledge about my narrow sub-discipline to follow my monologue. I may utterly fail to respond to the cues that they are getting bored and confused and have no familiarity with or interest in my area. We can imagine that when my interlocutors nod

---

adopting the voice of "the exotic" (490). I agree with Davis that bypassing a person's subjectivity is morally significant, though I argue that credibility excess is morally significant in a much broader range of cases than Davis' explanation might suggest. In particular: credibility excess can be a locus of epistemic *disrespect* (not merely harm, where harm is contingent upon the consequences of expressing the judgment or upon how the agent is treated (490)), its significance is not inextricably tied to identity-prejudice (failures of responsiveness due to other forms of vice—like self-absorption—also make for epistemic disrespect), and, relatedly, the marginalized/dominant relationship (490) is not essential to its moral significance.

politely, my self-absorption makes me resistant to my total evidence such that I give them excessive credibility ("Perfect! Here are folks well-versed in the literature affirming my views!"). It is a mistake to inextricably link disrespectful judgment to prejudice.[47] Disrespect is a broader phenomenon: as I argue in Section 4, it essentially involves a failure of responsiveness to persons because of vice or objectionable values (whether prejudice, self-absorption, or…).

Respecting someone *qua* knower requires a proper responsiveness to her *qua* epistemic subject. This should not be mistaken with an *accuracy* requirement. One might be perfectly justified in believing (falsely) that Sarah is an expert in astrophysics because one unwittingly stumbled upon the filmset where she plays one. But, surely, this belief would involve no failure of respect—respect does not require *accurate* judgments.[48] Moreover, accuracy is not *sufficient* for respect: if it just so happens that you *are* deeply familiar with Daodejing scholarship, there is still something disrespectful about Fred's judgment given the way in which it was produced.

I have been suggesting that credibility excess manifests epistemic disrespect when, because of vice, the judgment fails to have been produced in a way that is properly responsive to the person *qua* knower. Before providing a fuller theoretical underpinning for this claim, it will be useful to respond to some general worries about the very

---

[47]Fricker (2007: 22) suggests that identity-prejudice is a necessary ingredient for credibility deficit to undermine or disrespect someone as a knower. Similarly, Davis (2016) argues that credibility excess can sometimes count as an epistemic harm, but the explanation of when and why builds identity-prejudice directly into the potential significance of credibility excess.

[48]See Fricker's suggestion that some cases of credibility deficit are innocent errors and not instances of testimonial injustice (2007: 21-22).

suggestion that epistemically shoddy attributions of credibility excess can—themselves—amount to significant failures of respect.

One worry stems from the fact that, very often, people in positions of power are the ones receiving excessive credibility. Pick your favorite despicable person in a position of power. Call him "Ronald." Suppose that, Dale, one of Ronald's employees, routinely attributes excessive credibility to Ronald through deeply shoddy inferences. "*Really?*," one might ask; "does Dale *really* fail to respect Ronald *qua* knower, as you suggest?".

Stephen Darwall's (1977) distinction between "appraisal respect" and "recognition respect" and Robin Dillon's (1992) discussion of "care respect" (which she takes to be species of recognition respect) are helpful here. To be sure, Ronald receives no lack of *appraisal* respect—Dale appraises Ronald remarkably highly. But there is a form of respect that has little to do with esteem or high regard. More than this, it should be clear that appraisal respect is not the variety of respect that is antithetical to the forms of objectification, alienation, and estrangement on which the current project focuses. In failing to be responsive to Ronald, Dale fails to satisfy basic norms of *recognition*. I suggest that the very conditions underlying a relationship of genuine respect have eroded in this case: Dale fawns over Ronald, but his judgments about Ronald fail to manifest an important form of respect.[49] Put this way, the verdict about Dale should be less

[49]This is not to say that Dale is blameworthy for failing to accord Ronald this form of respect. Moreover, the suggestion here is that Dale fails to accord Ronald an important form of respect. Whether Dale's judgment toward Ronald is outright *dis*respectful will depend on why his judgments fail to be properly responsive to Ronald. If it is because of some vice—a vain obsession with the powerful, say—then Dale's judgment may well manifest disrespect rather than merely a lack of respect. I discuss the relationship between respect, lack of respect, and disrespect in Section 4.

surprising: the present dynamic between Dale and Ronald precludes the possibility of attuned responsiveness that (one form of) genuine respect requires.

A second worry is about whether credibility excess—even if disrespectful—can actually have the kind of significance characteristic of epistemic injustice. According to Fricker, the "original significance" of epistemic injustice is that

> [t]he subject is wrongfully excluded from the community of trusted informants, and this means that he is unable to be a participant in the sharing of knowledge… He is thus demoted from subject to object, relegated from the role of active epistemic agent… (2007: 132).[50]

With this diagnosis of the "original significance" in mind, it is tempting to think that only credibility *deficit* can have such significance. After all, can we really say that the person given too much credibility is *excluded* from some knowledge-sharing practice?

In some cases, I think we can. Emphasis might help. In credibility deficit, the agent is *excluded* from some informational exchange. But in both credibility excess and deficit, *the agent* is excluded from some informational exchange. That is, the agent—as a concrete individual with an epistemic subjectivity of her own—is perceived in such a way that alienates *her* (and the distinctive contributions *she* makes) from that exchange. It is a mistake to think that the only thing agents are owed as knowers is *enough* credibility. Rather, agents—in their role as knowers or epistemic subjects—are owed that our judgments about them be responsive to them as the particular epistemic subjects they are. We owe it to persons that their rational activity have some effect on our judgments about them, and when our judgments about them as agents fail to be minimally sensitive to their

---

[50]Marušić (2015), Moran (2005), Fricker (2006), and Craig (1990) similarly develop this suggestion that certain responses to testimony can be objectifying.

agency, this failure can mark a significant form of objectification or estrangement that is incompatible with respect. The "original significance" of epistemic injustice—namely, that it excludes, alienates, or objectifies epistemic agents—applies equally to credibility deficit and excess.

Indeed, we can say more than this. Step back from issues of respect/disrespect for a moment. While establishing the following point is not crucial for the present argument, it is plausible that, for virtually any feature of credibility deficit that might make it significant, that same feature can—in principle—be found in cases of credibility excess. A (the?) central worry about credibility deficit is that it can make both its recipient and the broader epistemic community worse off as knowers. Sure enough, when someone is repeatedly given credibility deficit, this may well damage their confidence, and this may well in turn make the person a worse knower. And, if people are not being believed when they ought to be believed, this may damage the entire epistemic community insofar as the community would be failing to pick up on the information that various testifiers have to offer. But, of course, credibility excess can also make its recipient and the broader epistemic community worse off as knowers, for reasons that include—but are not exhausted by—the fact that credibility excess can foster epistemic arrogance. For, when our credibility judgments are not responsive to facts about persons as the particular epistemic subjects they are, this surely worsens the dynamics of interpersonal informational exchange and makes it less conducive to transfers of knowledge. In educational contexts, credibility excess may prevent professors from helping their students learn and improve as knowers. We can imagine the lecturer who fails to adjust his pedagogy in response to the students in the room, who judges that their slow nodding

signifies that they *must* understand the content that he has been talking at them about for an hour and a half, even when his total evidence suggests that the students are not grasping the material. When the lecturer gives excessive credibility to the students, this judgment may play a crucial role in hampering the exchange of knowledge in the classroom. And, to return to the topic of respect, if the reason the lecturer gives excessive credibility to the nodding students has to do with a vice of sorts, then the lecturer's attributions of excessive credibility may well manifest a significant form of disrespect to those students.

All of this is to say that it is unclear that credibility deficit has some significance that cannot—in principle—be found in cases of credibility excess. A theory of epistemic respect must explain the capacity for both credibility deficit and excess to disrespect. Below is a theory that does.

5.4 The Value-Secured Reliability Theory of Respect

Consideration of disrespectful credibility excess brings out more generally what respect in judgment[51] is about. Respect of one kind—in particular, the kind antithetical to alienation and objectification—requires not esteem, trust, or praise; it requires proper responsiveness to persons as the particular agents they are.[52] This points to an

---

[51]Respect in judgment is not the only kind of respect there is. Respect may require certain behaviors. Perhaps respect requires that I *take steps* to inform myself about you as an epistemic agent (that I ask you questions or effortfully pay closer attention in our conversation). But I suggest that respect also directly governs judgments—this paper focuses on respect governing *judgments* as opposed to behaviors (though there is good reason to believe that the two have largely parallel structures).

[52]My general account sits in the tradition that includes Stephen Darwall's (1977) discussion of "recognition respect" and Robin Dillon's (1992) discussion of "care

underappreciated epistemic dimension of respect: whether some judgment is respectful depends, in part, on how information was accessed and processed in its production. In this final section, I sketch a theory of respect—of *epistemic* respect and respect in general—that can explain the cases of disrespectful credibility excess while also capturing this epistemic dimension of respect.

Fleshing out the theory requires precisifying the aforementioned *proper responsiveness*. Before I do this, note that this targeting of respect—as requiring proper responsiveness to the agent—identifies a genus of which *epistemic* respect is a species. Respect in its general form requires proper responsiveness to facts about the agent (including facts about his judgment-sensitive desires or emotions).[53] *Epistemic* respect is just a more localized form of respect for persons specifically *qua* epistemic subjects.[54]

So, what is it to say that a judgment must be *properly responsive* to the agent? It is to say that a judgment bearing upon an agent's rational activity must meet some basic epistemic and moral standards, where the satisfaction of the standards is connected. When my judgment accords you respect, it is a success of sorts—it manifests a form of moral competence. It is not *just* that my judgment is responsive to your agency, but, rather, my responsiveness to you is somehow expressive of or explained by my valuing you. This is the core idea underlying my theory of respect, the "Value-Secured Reliability

---

respect," which she takes to be a species of Darwall's "recognition respect" (Dillon, 1992: 112).

[53]For a discussion of judgment-sensitive desires and emotions, see Angela Smith (2005).

[54]Toward the end of this paper, I respond to worries about the extent to which we should think of *epistemic respect* as occupying its own category.

Theory." Respect requires that my judgment about you be reliably produced *because* of the influence my value or care for you has on how I respond to information from my environment.[55]

What is it for a reliably produced judgment to be, in this sense, *value-secured*? Our values clearly shape our patterns of informational access and processing, often unreflectively. The person who cares about cleanliness is more likely to notice the dust (Arpaly, 2002: 83). The person watching *Au Hazard Balthazar* who values animal welfare is more likely to form the judgment that "this film must have been awful on the mule." Since one's values can shape how one accesses and processes information, one's values can sometimes explain why one underwent some particular reliable belief-forming process in some situation. When one's judgment about an agent is reliably produced, *and* one's value for them explains[56] why one underwent that particular process in that particular situation (why certain information was made available and processed in the way it was), the judgment satisfies the requirements of the Value-Secured Reliability Theory of respect. This captures at a theoretical level what—at an intuitive level—is at

---

[55]For a variety of reasons (some less central to the purposes of this paper, some more so), I prefer to use the category of *reliably produced* judgments when discussing my account of respect. If the reader would prefer to substitute a more internalist notion (e.g., the evidentialist notion of well-foundedness, which involves a basing relation), much of what is central to this current project remains intact.

[56]Explanation, of course, admits of degrees. In no case will a person's judgment be produced *solely* because of her value for persons—other factors will play a causal role in the production of the judgment, too. The Value-Secured Reliability Theory requires that one's value for persons have played some non-trivial role in shaping the production of the judgment.

the core of respect: respect requires that our judgments about persons be reasonably attuned to them because we value them.[57]

Two interrelated clarificatory remarks are in order. First, valuing a person is not the same as *liking* the person. Rather, valuing a person is to (implicitly or explicitly) attach a certain value or worth to them in such a way that ought (rationally) to influence what one notices about them and how one processes what one notices about them (see Angela Smith, 2005: 242-246 for a nice discussion of how our values ought to affect our patterns of noticing). One can respect someone that one seriously dislikes by continuing to have one's judgments be properly responsive to him as an agent—if one's implicit or explicit appreciation for the other's worth as a person influences one's patterns of awareness and processing such that one's judgments about him as an agent continue to be formed in reliable ways, this is a minimal way of continuing to respect someone whom one seriously dislikes. Sometimes, we "write someone off" such that our negative judgments about them will no longer be updated by relevant evidence about them—in such cases, we stop respecting this person.

---

[57]My notion of *value-secured reliability* is in important respects indebted to Ernest Sosa's work in virtue epistemology (2007, 2015). Sosa has long argued that <reliability secured by something attributable to the agent> marks an important category in epistemology. Irrespective of whether Sosa is correct about the epistemology, I believe that this broad category is central to varieties of moral success, like respect. This being said, for reasons that are not pressing for present purposes, my Value-Secured Reliability Theory offers a much less cognitively demanding model for securing attributability to the agent than Sosa's theories of "reflective aptness" and "aptness full well" do (76). Roughly: where Sosa makes reliability attributable to the agent depend on the agent's appreciation of her own reliability, I make reliability attributable to the agent depend just on the agent's person-level values shaping (perhaps unreflectively) her patterns of informational access and processing.

The second clarificatory remark is that valuing a person ought not to be thought of as excessively demanding. Indeed, we can value strangers we encounter for the first time. As far as judgments are concerned, valuing or caring about a person roughly entails that, *if/when* one forms a judgment about that person, one notices and appropriately processes a range of available cues pertaining to them as the particular agent they are. Respect does not require that we drop everything we are doing and distribute surveys to the strangers around us so that we can form maximally reliable judgments about them. Rather, respect requires that *when we do* form judgments about persons, they are formed by reliable processes because we basically care about being attuned to persons when we make judgments about them. For strangers, respect will often require that we withhold judgments about them or that our judgments about them be held with minimal confidence—relative to our limited evidence about strangers, withholding judgment (or placing minimal confidence in our judgments) will often be the only attitude that could result from reliable belief-forming processes.

The Value-Secured Reliability Theory readily explains both the cases of disrespectful credibility excess and deficit. In both cases, one agent forms a judgment about another *qua* rational agent (*qua* knower) through some shoddy, unreliable process due to vice. We owe it to persons that our judgments about them not be so recklessly formed. Beyond explaining epistemic disrespect, the account provides satisfying theoretical underpinnings for intuitive claims about respect. In what follows, I will say a bit about some of the advantages of the theory, and then I will clarify the theory by responding to some potential objections.

One virtue of the Value-Secured Reliability Theory is that it supplies a natural framework with which to draw necessary distinctions between *lack* of respect and outright *dis*respect. If someone forms a judgment about you *qua* agent through some shoddy inference but only because she has a cognitive disability that prevents her from being more attuned to you, surely her judgment is not disrespectful. The judgment is perhaps (faultlessly) *lacking* some level of respect, but not disrespectful.

Such cases highlight a need to distinguish respect, lack of respect (without disrespect), and disrespect. And the structure of the Value-Secured Reliability Theory points to a plausible way of doing this. We can draw these distinctions by appealing to (1) the (un)reliability of the process that produced the judgment and to (2) the explanation of why the agent underwent that particular process. My judgment about S is *respectful* when, due to my valuing or caring about S, the judgment was reliably produced—this is the success case. When my judgment about S fails to be reliably produced, it may be *merely* lacking in respect, or it may be disrespectful. What is the difference? I suggest that, when my failure of responsiveness is not explained by vice or objectionable values, the judgment is merely lacking in respect. Such cases might include failures of responsiveness due to cognitive incapacity, preoccupation with a tragedy, extraordinary hunger, or (arguably) mundane unexpected events that momentarily distract us. However, when my failure of responsiveness *is* explained by vice or objectionable values (prejudice, self-absorption, and the like), the relevant judgment escalates to the level of

disrespect. The Value-Secured Reliability Theory's general structure positions us to capture and explain these distinctions.[58]

Another virtue of the theory is that it can help us grapple with the thorny question about when, if ever, relying on a stereotype is compatible with the respect that a person is owed. I will not pretend to settle the issue here. But the Value-Secured Reliability Theory helps us identify the relevant considerations. In the opening case of this paper, Fred forms a judgment about you through an evidence-resistant prejudice that makes his belief unreliably produced. In more complicated cases, one might form a judgment about some agent by relying on a stereotype that is, as a matter of empirical fact, statistically grounded. The Value-Secured Reliability Theory shows us that whether such a judgment is compatible with respect depends, first, on whether the judgment *really is* reliably produced (is the reliance on the stereotype so strong that it makes the agent resistant to information?). If it is not reliably produced, then the judgment is not respectful. If it *is* reliably produced, then whether it is compatible with respect depends on *why* the agent underwent this reliable process in this situation. Perhaps it is because the agent cares about being attuned to the person that, when he formed this judgment about this person, he accessed and processed information in this reliable way. Alternatively, perhaps he relied on this (by hypothesis) reliable inference because of an eagerness to draw unflattering conclusions about persons with some particular social identity (it is not obvious that this kind of case is well described: for, if the judgment is largely motivated

[58]More complicated are cases in which a judgment about an agent is reliably produced, but the production of the judgment is largely explained by the agent's objectionable attitudes or vices. I discuss such cases in a moment.

in this way, the judgment is perhaps not produced by a reliable process). Plausibly, only the former and not the latter case is compatible with respect for the agent. In real cases, it will sometimes be hard to know whether some judgment was produced by a process that was both reliable *and* relevantly explained by a value for the relevant person. This difficulty simply underscores the complexity of the ethics and epistemology of stereotyping.

Having discussed the Value-Secured Reliability Theory and some of its theoretical virtues, let me respond to what may seem like an obvious objection to the view. Consider the following case:

> Obsession: Max forms lots of reliably produced judgments about Fiona *qua* agent. Moreover, he undergoes these reliable belief-forming processes because he values Fiona. But that's just the problem. Max is so obsessed with Fiona that he is constantly spying on her in ways that clearly flout requirements of respect.

Surely, Max fails to accord Fiona an important form of respect. However, on first glance, it looks as though the Value-Secured Reliability Theory is committed to saying that this is a paradigm case of respect! After all, Max's judgments are reliably produced, and his value or care for Fiona seems to explain why his judgments are reliably produced. So, how can the theory handle this case?

Contra first appearances, Max's judgments do not satisfy the conditions of the Value-Secured Reliability Theory. The case is badly described: though Max's judgments are reliably produced, this fact is not plausibly explained by his value for Fiona as a person. Rather, though things may not seem this way from his first-person perspective, a more plausible explanation for why information was accessed and processed as it was is that Max is fascinated in Fiona purely or primarily as an object of investigation—Fiona, for whatever unfortunate reason, is simply a stand-in for satisfying Max's strange

obsession. In this case, it is not Max's caring about Fiona as a person that shapes his informational access and processing—it is his interest in such objects of investigation. Such objectification is incompatible with respect, and the Value-Secured Reliability Theory correctly rules out Max's judgments as respectful.

Now, so far, we have established that Max's reliably produced judgments are not the success case—they are not expressive of or explained by a value for persons, and this being so, they are not manifestations of respect. Are the reliably produced judgments outright disrespectful? Of course, all the spying behavior rested on attitudes that are outright disrespectful. The attitudes underwriting the spying behavior fail to be responsive to facts about Fiona as an autonomous agent with standing preferences not to be treated in these ways, and Max's failure to be so responsive is presumably explained by objectionable attitudes or vice (unless his obsession is so pathological that the case seems better explained by chemical imbalances rather than by objectionable attitudes or vice). So, outright disrespect can be found somewhere in the causal chain leading to Max's reliably produced judgments about Fiona. But are the reliably produced judgments *themselves* disrespectful?

Ultimately, I think we should say the judgments are significantly lacking respect, though they are not outright disrespectful (registering, of course, that the spying behavior is bound up with attitudes that *are* outright disrespectful). Take the moment right after Max has gone through the spying, when he is landed with some judgments about Fiona. Vices or objectionable attitudes are surely a large part of the story about how Max arrived at these judgments, for vice explains how Max acquired his evidence in the first place. But, once Max has acquired all this evidence, it would seem that the proper functioning

of his belief-forming mechanisms is what best explains why he formed these particular reliably produced judgments about Fiona. The judgments are not explained by or expressive of a care for Fiona as a person, and that is why the judgments are not respectful. But insofar as the reliably produced judgments (once Max has acquired the evidence) are best explained by the proper functioning of his belief-forming mechanisms, it seems mistaken to say that these judgments are, themselves, outright disrespectful. Max believes what anybody ought to believe if confronted with his evidence.

Cases like Max's show us why respect cannot involve *merely* reliably produced judgments. The reliably produced judgments must be *value-secured*, or else the judgments do not manifest the kind of moral competence central to respect. (In a case even more worrisome than Max's, we can imagine one person seeing to it that their judgments about another are reliably produced only to ideally position themselves to manipulate or deceive the other—here, too, we have reliably produced judgments that are not properly rooted in value for the person, and these judgments, thus, fail to manifest the kind of moral competence central to respect). While such cases show that respect in judgment requires that the judgment be expressive of a value for some person, one might wonder about the *reliability* component of my account. Can a judgment be respectful so long as it is sufficiently explained by a value for the person, irrespective of whether it is reliably produced?

It cannot. That is because when a judgment about a person is not reliably produced, the problems of alienation and estrangement rearise. When my judgments about you as an agent are not formed reliably—when they fail to be properly responsive to the information I have about your agency—I am not attuned to you as respect requires,

even if my judgments are largely explained by my valuing or caring for you. There is an important lack of respect if, even in valuing you, I begin to attribute features to you as an agent in ways that are unresponsive to your agency—this is so even if what I attribute to you are things that are usually seen as desirable, like virtues or high levels of credibility. Such badly grounded attributions make for a special and underappreciated form of alienation. Again, respect in judgment is a manifestation of a certain moral competence: it is only when a judgment about a person is reliably produced because of one's value for them that the judgment accords the person the respect they command as a concrete agent.

5.5 Conclusion

I have now sketched and motivated the Value-Secured Reliability Theory of respect. As I have tried to show, cases of intuitively disrespectful credibility excess help us appreciate how respect crucially involves a proper responsiveness to agents in their particularity. Epistemic respect—or, respecting someone as a knower—is not centrally about making sure one gives an agent enough credibility; rather, it essentially involves being attuned to the agent as the particular knower that they are because one basically values them. The unifying explanation of disrespectful credibility deficit and excess is that both involve failures of responsiveness due to vice. A theory of respect—whether for persons *qua* knowers or more generally—should be responsive to the value of responsiveness.

Before closing, I wish to consider a concern about whether *epistemic respect* really marks a distinctly *epistemic* phenomenon. There is good reason to be somewhat wary of the proliferation of "epistemic x" categories in the literature, and I am not eager to heap on more taxonomy. Moreover, if we are to accept *epistemic respect* as a category,

what is to stop us from identifying different forms of respect for so many other conceivable roles an agent might occupy? Do we really want to go in for *baker respect*, *gardener respect*, and so on? Let me address these issues for a moment.

At a certain point, it is not especially important for the broader aims of this paper that we collectively accept *epistemic respect* as a distinct category or concept. The Value-Secured Reliability Theory provides a well-motivated and general theory of respect governing our judgments about persons as agents, it reveals an underappreciated epistemic dimension of respect, and it provides theoretical underpinnings for a variety of intuitive claims about respect. More than this, there are interesting questions about what respect requires in our judgments about persons as knowers, and my theory helps us answer them. Ultimately, it does not much matter whether we want to call this form of respect "epistemic respect" or just "respect" (noting in a separate breath that respect bears upon our judgments about persons as knowers). On a similar note, whether tackling questions about epistemic injustice and related topics is "more the business of epistemologists" or "more the business of ethicists" is not, to my mind, a terribly interesting question.

That being said, I think there are reasons to give special attention to the requirements of respect for persons in their roles as knowers or epistemic subjects, and these reasons do not apply to persons in *whatever* roles they happen to occupy. Moreover, some of these reasons ought to be of particular interest to epistemologists. First, the role of knower or epistemic subject is an integral role for *any* agent. Facts about us as epistemic subjects are intimately bound up with who we are as particular agents—imagine trying to characterize *who someone is* without appealing to facts about what the

person knows and doesn't know, reasonably believes and unreasonably believes, and so on. If respecting persons requires that our judgments about them as agents be properly responsive to them as the particular agents they are, then how we respond to persons as *epistemic* subjects (as opposed to as, say, *bakers*) will be a central and quite universal part of respect.

This does not yet provide a reason for epistemologists (*qua* epistemologists) to care about epistemic respect. But there are reasons. One reason is that there are questions in the epistemology of testimony concerning our practices of using persons as sources of information, and resolving these questions requires that we resolve questions about what respect requires in our judgments about persons as knowers or epistemic subjects. Richard Moran (2005), Elizabeth Fricker (2006), Berislav Marušić (2015), Sarah Stroud (2006) and others suggest in some form that what we rationally ought to believe upon the receipt of another's testimony should somehow be informed or constrained by considerations of respect for the speaker as an agent. If something like this is at all plausible (indeed, if we are to *find out* whether something like this is at all plausible), then we need to know what it is to respect someone specifically in their role as an epistemic subject. The Value-Secured Reliability Theory gives us an account with which to resolve these questions in the epistemology of testimony while also supplying a well-motivated general theory of respect.

CHAPTER 6

CONCLUSION

In this brief conclusion, I wish to summarize some of the core positions defended in the preceding chapters and to gesture toward some avenues for future research that are opened up by or otherwise continuous with this project.

The first central claim—defended in Part I—is that the existence of robust epistemic or cognitive requirements on moral responsibility would not have the disastrous consequences for our moral practices and interpersonal relationships that philosophers have suggested. Even one of the more demanding construals of the epistemic requirement—namely, that moral culpability requires full-blown akrasia—does not, itself, motivate an argument for responsibility skepticism. Moreover, there is no reason to believe that excusing a person for some objectionable attitude and associated wrongdoing because of his epistemic circumstances is necessarily disrespectful. If the worry with epistemic requirements on moral responsibility is rooted in concerns about skepticism and disrespect, then this worry is misplaced. These consequences simply do not follow from allowing that a person's epistemic situation is sometimes exculpating.

That being said, there are serious problems with the ways philosophers have tried to capture the epistemic requirements on responsibility. We see this vividly in discussions of responsibility for right action. As we have seen, a popular view—among both those who seek to play up and play down the significance of consciousness for responsibility—is that moral worth requires that one be aware of the facts in virtue of which one's action is morally right. But satisfying such a condition is much more difficult than these

philosophers recognize: were this actually a requirement on moral worth, then most intuitively morally worthy actions would fail to be morally worthy. The solution is not to abandon epistemic/cognitive requirements on moral worth—after all, the person who performs a right action in a trance, entirely isolated from the morally significant features of her action, is surely not morally responsible for her right action. Rather, the solution lies in recognizing the false assumption underlying the popular view on moral worth's consciousness requirement.

That assumption is that, if my morally right response to my environment is to be attributable to *me* and not just an accident, I must be aware of the facts that make my response the right one. Only then (goes the assumption) are *my* values sufficiently brought to bear upon my response. Yet, as I have argued, one's values can shape one's unconscious processing of quite minimal information such that one's right action is both non-accidental and correctly attributable to the agent *without* the agent being aware of the right-making facts. This brings us to the second central contribution of this project.

In Part II, I defended an original theoretical framework of moral responsibility for right action: the Value-Secured Reliability Theory of moral worth. According to this theory, one's right action is morally worthy just to the extent that its production is explained by one's value-secured reliable tie to the right, a reliable tie to the right that is, itself, explained by the influence one's values have on shaping one's informational access and processing. When an action is explained by a value-secured reliable tie to the right, consciousness is an integral part of the story, as it is by the (perhaps unreflective) influence one's values have on one's informational access and processing that an action can be explained by a value-secured reliable tie to the right. But none of this requires

consciousness of *the facts in virtue of which one's action is right*—depending upon one's background values and the strength of those values, one's action can be explained by a value-secured reliable tie to the right even when one is conscious of nothing more robust than quite minimal sensory cues. The Value-Secured Reliability Theory offers a compelling account of moral worth that at the same time identifies the role consciousness plays in securing moral worth.

The third central claim of this project has been to show that *respect*, too, has an underappreciated epistemic dimension and that, indeed, it has a parallel theoretical structure to responsibility for right action. A certain variety of respect, I argue, requires that one's judgments bearing upon someone *qua* rational agent be reliably produced *because* of the influence one's values have on how one access and processes information about them. Many philosophers have suggested that respect instead requires that our judgments simply not undersell agents in various respects (that we not attribute too little agency to them, that we not give them too little credibility, and so on). But this construal of respect—in failing to recognize the epistemic dimension of respect—cannot make sense of varieties of estrangement that are antithetical to respect. The Value-Secured Reliability Theory of respect correctly captures the intuitive idea that respect centrally involves that we be properly attuned to the persons around us because we value them.

Beyond responsibility and respect, I believe that the Value-Secured Reliability framework can usefully illuminate a wide variety of other philosophically interesting phenomena. In Chapter 4, I mention that it may make sense of epistemic attributability and artistic attributability: it may be able to explain what it is for knowledge to be attributable to me and for some good artistic choice to be attributable to me. Indeed,

insofar as success essentially involves a performance whose non-accidental tie to a good is attributable to the agent, I believe the Value-Secured Reliability framework can readily be applied to any variety of success.

One perhaps unexpected area of inquiry the Value-Secured Reliability framework may inform concerns the natures of love and friendship. Consider Phillip Pettit's (2015) work on these subjects. Pettit discusses what he labels the "robust goods" of love and friendship, characterizing each as a manifestation of a disposition to provide some thin good (care, in the case of love, and favor, in the case of friendship) in a robust range of possible worlds (2015). In a critical review, Paul Hurley (2016) suggests that Pettit's accounts of friendship and love fail to explain their distinctive value. He writes:

> …if some stranger were permanently conditioned by a Clockwork Orange-like procedure to be disposed to provide me robustly with the thin good of favor, perhaps with even greater reliability than my friends do, although I may value such robust favor, I would not take myself to have gained another friend. Nor would I consider the stranger to be "giving me" the good of friendship that I receive from my friends. Such considerations suggest that perhaps what is valuable about friendship, love, justice, mutual respect, etc. is not simply the dispositions to provide corresponding thin benefits robustly; indeed, that the value of such dispositions taken by themselves fails to capture the importance of such rich goods in human life.[59]

Now, Hurley drops the discussion where this quotation ends. But here is a proposal about what distinguishes the Clockwork Orange "friend" from the true friend. Both supply some thin good reliably, but only the latter supplies the thin good reliably *because* of the influence that his values have on how he accesses and processes information. This suggests that the Value-Secured Reliability framework may be able to supply compelling

[59]Hurley, Paul. (2016). "Review. Phillip Pettit: *The Robust Demands of the Good*," *Notre Dame Philosophical Reviews* 201601.

accounts of “robust goods” like friendship and love that accommodate some of Pettit’s key insights while not losing sight of what makes these relationships meaningful.

Let me end by noting a few of the questions that—though they are part of the core set of questions that animate the present project—I have not fully answered.

First, and perhaps most obviously: though I have supplied a framework for moral responsibility for *right* action and specified its epistemic/cognitive requirements, I have not taken a position on responsibility for *wrong* action—I have merely defended the possibility of exculpatory ignorance against worries about skepticism and disrespect. (In Chapter 3, I provide partial explanations of cases where an agent who arrived at some objectionable attitude rationally may not be morally responsible for the objectionable features of his attitude, but I offered nothing like an account of moral responsibility for wrong action nor a precise statement of its epistemic requirements). It is not immediately clear how/whether the Value-Secured Reliability framework can make sense of responsibility for wrong action (or blameworthiness). Let me consider two interpretive options and then point to the larger problem that any option faces.

Suppose we say that one is blameworthy for some wrong action just to the extent that it *fails* to be explained by a value-secured reliable tie to the right. This surely will not do. If I unexpectedly have an uncontrollable muscle spasm and break your fancy mug, my action fails to be explained by a value-secured reliable tie to the right, but I am not responsible for the action.

A different possibility is that one is blameworthy for some wrong action just to the extent that it is explained by a value-secured reliable tie to the *wrong*, a reliable tie to the wrong that is secured by the influence one’s person-level values have on one’s

informational access and processing. Call this the "Value-Secured Reliability Theory of blameworthiness." Such an account would not imply that a basically good person could never perform a blameworthy action. If something about a basically good person's value-set is such that, in a circumstance *just like this one*, he will reliably perform the wrong action because of his values, he can perform a blameworthy action. But, there are potential problems for this account that speak to more fundamental puzzles about the (a)symmetry of praiseworthiness and blameworthiness.

To see this, consider Susan Wolf's (1987) famous case of JoJo.

> JoJo is the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father's special feelings for the boy, JoJo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison or to death or to torture chambers on the basis of whim. He is not *coerced* to do these things, he acts according to his own desires. Moreover, these are desires he wholly *wants* to have. When he steps back and asks, "Do I really want to be this sort of person?" his answer is resoundingly "Yes," for this way of life expresses a crazy sort of power that forms part of his deepest ideal (53-54).

On the face of it, JoJo performs *lots* of morally wrong actions that are explained by a value-secured reliable tie to the wrong. Yet, for many (myself and Wolf included), there is a strong intuition that someone with JoJo's upbringing—if we spell out the case carefully—is not morally responsible for these wrong actions. How could we reasonably expect JoJo to have done anything other than what he did? Famously, Wolf believes that the conditions for praiseworthiness and blameworthiness are asymmetrical: blameworthiness requires that the agent have had a reasonable opportunity to avoid performing the wrong action, but praiseworthiness does not require that the agent have had a reasonable opportunity to avoid performing the right action (1980, 1987). If

praiseworthiness and blameworthiness do, in fact, have asymmetrical conditions, then perhaps we should not expect the Value-Secured Reliability Theory to provide an account of blameworthiness—blameworthiness, and not praiseworthiness, may have some further "reasonable avoidability" condition.

That said, there is perhaps reason to believe that the Value-Secured Reliability Theory of blameworthiness *can* correctly imply that JoJo is not responsible for his wrong actions. In discussing cases somewhat similar to JoJo's (cases of moral ignorance reinforced by cultural circumstances), Michael Slote suggests that we cannot "sensibly suppose that the failure of the ancient world to attain and act upon a correct moral view of slavery merely reflects personal weaknesses or deficient methods of moral training that, unluckily, happened to be fairly universal in those days" (1982: 72). We might extend Slote's suggestion by suggesting that, in the case of the reliable dispositions to perform morally atrocious actions that we find in the "ancient world" and in JoJo, these reliable dispositions are not well explained by the agents' values. Facts about the agents' upbringings and cultural circumstances far better explain these agents' reliable ties to the wrong. Since these agents' actions are not explained by *value-secured* reliable ties to the wrong, they are not blameworthy.

It is not clear to me whether the above is a line of inquiry worthy of further development. Given that JoJo's values fit with his actions, why can't we say that JoJo's values (as opposed to his upbringing) explain his reliable tie to the wrong? How, in principle, might we demarcate the explanatory roles of an agent's values as opposed to his cultural circumstances in securing a reliable tie to the wrong? Perhaps we can defend the Value-Secured Reliability Theory of blameworthiness by noting that we *should* feel

conflicted about whether JoJo is responsible: with JoJo, we should not be confident about the explanatory role his values play in securing his reliable tie to the wrong. At any rate, I do not aim to resolve these issues here.

Concerning respect, there are some unresolved questions about how to weigh respect against (potentially) competing demands of friendship and trust. I have suggested that respect requires, in part, that one's judgments bearing upon a person *qua* agent be reliably produced. Yet, as some see it, friendship and trust often require forming beliefs in ways that are unreliable or not well proportioned to one's evidence (see, for instance, Marušić, 2015, and Stroud, 2006). Now, I believe that these views about friendship and trust are mistaken, and mistaken, in part, because they demand violations of respect. But showing that the goods of friendship and trust do not conflict with my account of respect will have to wait for another time.

While there are certainly remaining questions, I hope 1) to have shown that facts about how we access and process information play significant explanatory roles in the correct theories of responsibility and respect, 2) to have shown that this does not generate unacceptable consequences for our moral practices, and 3) to have offered a compelling new theoretical framework that explains and motivates the epistemic requirements on these (and other) varieties of success.

BIBLIOGRAPHY

Adams, Robert Merrihew. (1985) “Involuntary Sins.” *Philosophical Review* 94 (1):3-31.

Anderson, Elizabeth. (2012) “Epistemic Justice as a Virtue of Social Institutions.” *Social Epistemology* 26 (2):186-219.

Aristotle. (1999) *Nicomachean Ethics*, Terence H. Irwin (ed. trans.). Hackett Publishing Co.

Arpaly, Nomy. (2002) *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.

—. (2015a) “Huckeberry Finn Revisited: Inverse Akrasia and Moral Ignorance.” In Randolph Clarke, Michael McKenna, & Angela M. Smith (eds.), *The Nature of Moral Responsibility*. Oxford University Press.

—. (2015b) “Consciousness and Moral Responsibility, by Levy, Neil.” *Australasian Journal of Philosophy* 93 (4):829-831.

Arpaly, N. and Schroeder, T. (1999) “Praise, Blame, and the Whole Self.” *Philosophical Studies* 93 (2):161-188.

—. (2013) *In Praise of Desire*. Oxford University Press.

Audi, Robert. (1986) “Acting for Reasons.” *Philosophical Review* 95 (4):511-546.

Baars, Bernard J. (1988). *A Cognitive Theory of Consciousness.* Cambridge University Press.

Bennett, Christopher. (2004) “The Limits of Mercy.” *Ratio* 17 (1):1-11.

Bennett, Jonathan. (1974) “The Conscience of Huckleberry Finn.” *Philosophy* 49 (188):123-134.

Block, Ned. (1995) “On a Confusion about a Function of Consciousness.” *Brain and Behavioral Sciences* 18 (2):227-247.

BonJour, Laurence and Sosa, Ernest. (2003) *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues.* Oxford: Blackwell Pub.

Carruthers, Peter. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.

Craid, Edward. (1990) *Knowledge and the State of Nature: An Essay in Conceptua Synthesis*. Oxford: Clarendon Press.

Darwall, Stephen. (1977). “Two Kinds of Respect.” *Ethics* 88 (1):36-49.

Davis, Emmaon. (2016) “Typecasts, Tokens, and Spokespersons: A Case for Credibility Excess as Testimonial Injustice.” *Hypatia* 31 (3):485-501.

Dillon, Robin S. (1992) “Respect and Care: Toward Moral Integration.” *Canadian Journal of Philosophy* 22 (1):105-132.

Doris, John M. (2015) *Taking to Our Selves: Reflection, Ignorance, and Agency*. Oxford University Press.

Eagleman, David. (2011) *Incognito*. Pantheon Books.

Feldman, Richard and Conee, Earl. (1985) “Evidentialism.” *Philosophical Studies* 48 (1):15-34.

Fischer, J.M. and Ravizza M. (1998) *Responsibility and Control*. Cambridge University Press.

Franklin, Christopher Evan. (2013) “Valuing Blame.” In D. Justin Coates & Neal A. Tognazzini (eds.), *Blame: Its Nature and Norms*. Oxford University Press.

Fricker, Elizabeth. (2006) “Testimony and Epistemic Autonomy.” In Jennifer ackey & Ernest Sosa (eds.), *The Epistemology of Testimony*. Oxford University Press.

Fricker, Miranda. (2007) *Epistemic Injustice*. Oxford University Press.

Harman, Elizabeth. (2011) “Does Moral Ignorance Exculpate?” *Ratio* 24 (4):443-468.

Hurey, Paul. (2016) “Review of Phillip Pettit’s *The Robust Demands of the Good*.” *Notre Dame Philosophical Reviews* 201601.

Levy, Neil. (2009) “Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick.” *Ethics* 119 (4):729-741.

—. (2011a) “Exressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action.” *Analytic Philosophy* 52 (4):243-261.

—. (2011b) *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press.

—. (2014) *Consciousness and Moral Responsibility*. Oxford University Press.

Marušić, Berislav. (2015) *Evidence and Agency: Norms of Belief for Promising and Resolving*. Oxford University Press.

Medina, José. (2011) “The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary.” *Social Epistemology* 25 (1):15-35.

Moody-Adams, Michele M. (1993) "On the Old Saw that Character is Destiny." In Owen Flanagan & Amélie Oksenberg Rorty (eds.), *Identity, Character, and Morality: Essays in Moral Psychology*. MIT Press.

—. (1994) "Culture, Responsibility, and Affected Ignorance." *Ethics* 104 (2):291-309.

Moran, Richard. (2005) "Getting Told and Being Believed." *Philosophers' Imprint* (5):1-29.

Pettit, Phillip. (2015) *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect*. Oxford University Press.

Plato. (1997) *Complete Works*, John M. Cooper (ed.). Hackett Publishing Co.

Rosen, Gideon. (2002) "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103 (1):61-84.

—. (2004) "Skepticism About Moral Responsibility." *Philosophical Perspectives* 18 (1):295-313.

—. (2008) "Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility." *Journal of Philosophy* 105 (10):591-610.

Scanlon, T.M. (1998) *What We Owe to Each Other*. Belknap Press of Harvard University Press.

—. (2008) *Moral Dimensions: Permissibility, Meaning, and Blame*. Oxford University Press.

Setiya, Kieran. (2012) *Knowing Right from Wrong*. Oxford University Press.

Sher, George. (2009) *Who Knew? Responsibility Without Awareness*. Oxford University Press.

Sliwa, Paulina. (2016) "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 93 (2):393-418.

Slote, Michael. (1982) "Is Virtue Possible?" *Analysis* 42 (2):70-76.

Smith, Angela M. (2005) "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115 (2):236-271.

—. (2007) "On Being Responsible and Holding Responsible." *Journal of Ethics* 11 (4):465-484.

—. (2008) "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138 (3):367-392.

—. (2015) “Attitudes, Tracing, and Control.” *Journal of Applied Philosophy* 32 (2):115-132.

Smith, Holly. (2011) “Non-tracing Cases of Culpable Ignorance.” *Criminal Law and Philosophy* 5 (2):115-146.

Sosa, Ernest. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I.* Oxford University Press.

—. (2015). *Judgment and Agency*. Oxford University Press.

Stroud, Sarah. (2006) “Epistemic Partiality in Friendship.” *Ethics* 116 (3):498-524.

Sturgeon, Nicholas. (1992) “Nonmoral Explanations.” *Philosophical Perspectives* (6):97-117.

Wolf, Susan. (1980) “Asymmetrical Freedom.” *Journal of Philosophy* 77 (3):151-166.

—. (1987) “Sanity and the Metaphysics of Responsibility.” In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.

—. (2015) “Responsibility, Moral and Otherwise.” *Inquiry: An Interdisciplinary Journal of Philosophy* 58 (2):127-142.

Zimmerman, Michael J. (1997) “Moral Responsibility and Ignorance.” *Ethics* 107 (3):410-426.

—. (2015) “Varieties of Moral Responsibility.” In R. clarke, M. McKenna, & A.M. Smith (eds.), *The Nature of Moral Responsibility*. Oxford University Press.