

© 2018

Marilie Coetsee

ALL RIGHTS RESERVED

THE ETHICS OF DISAGREEMENT

by

MARILIE COETSEE

A dissertation submitted to the
School of Graduate Studies
Rutgers, The State University of New Jersey
In partial fulfillment of the requirements

For the degree of
Doctor of Philosophy
Graduate Program in Philosophy

Written under the direction of
Ruth Chang

and approved by

New Brunswick, New Jersey

October 2018

ABSTRACT OF THE DISSERTATION

The Ethics of Disagreement

By MARILIE COETSEE

Dissertation Director:
Ruth Chang

Deep moral disagreements must sometimes be worked *through*, rather than around. In *The Ethics of Disagreement*, I first examine a particular case in which a deep disagreement must be worked through, and then draw on work in philosophy of mind and epistemology to set the philosophical foundations for two approaches to deep disagreements that promise to help us make our way more virtuously through them.

Illiberal persons resist liberal values, and so also liberal justice, but liberals cannot in good conscience simply overlook their commitment to liberal justice to accommodate disagreement with the illiberal. Insofar as we do insist on liberal justice, however, illiberal persons become subject to coercive legislation whose justificatory merits they are not positioned to appreciate. Liberal political theorists often dismiss concerns about this justificatory alienation of illiberal persons as irrelevant to the legitimacy of liberal legislation. However, in “The Liberal Duty to Deliberate with the Illiberal,” I argue that our own liberal commitment to political autonomy militates against such a dismissal. I contend that liberal theorists have only been able to coherently dismiss worries about illiberal

citizens' lack of autonomy under liberal democratic law because those theorists harbor false stereotypes about what the political character of illiberal persons is like. I challenge these stereotypes by discussing the figure of Seyyid Qutb (1906–1966), an illiberal Islamic political thinker, and argue that concerns about the political autonomy of persons like Qutb generate a duty to engage them in *Deep Deliberation* that is aimed at reconciling them to the liberal justificatory bases of their coercion.

It is difficult to maintain mutual goodwill in conditions of deep moral disagreement, but preservation of such goodwill is critical to maintaining the resolve to work through it—whether via Deep Deliberation or in any other way. To maintain goodwill, parties to a moral disagreement must be able to (at least) understand their opponents as motivated by considerations they sincerely apprehend as moral reasons. But how do we come to understand our opponents as sincerely morally motivated in this way, when their “reasons” seem outlandish and alien to us? To answer this question, we need to know what it is for someone to appreciate a consideration as a moral reason, and that is the question I address in my second chapter, “The Phenomenal Appreciation of Reasons.” In this chapter, I draw on the resources of philosophy of mind to argue that to appreciate a consideration as a moral reason to ϕ is to present it under the light of a particular phenomenologically-mediated mode of presentation: one that presents the relevant consideration via the light of a felt directive force “pointing” towards ϕ -ing—lending weight to it, or soliciting it—in a particular authoritative way. If I am right, then to be able to understand another person as motivated by a consideration she sincerely takes to be a moral reason, you must be able to simulate what it is like to have that consideration “call out” to you (as it does to her) with a solicitive force directed at ϕ -ing.

In the course of working through deep moral disagreements, parties must consider and be ready to revise their initially opposing answers to the question of what ought to be done. Importantly, their disagreement arises not in the context of solitary contemplation about abstract truths, but rather in the context of shared social space where opposing judgments often give rise to practical social costs for disagreeing parties. In my final chapter, “An Introduction to Socially Problematic Disagreements,” I argue that the social costs of disagreement generates special norms of belief revision that epistemologists studying disagreement have problematically failed to acknowledge. In particular, I argue that the social practical costs of a disagreement can sometimes give us compelling reasons to hold open questions about the truth of opposing parties’ views—even when the relevant questions could, judging purely from an *epistemic* perspective, reasonably remain closed.

Deep moral disagreements may never be fully resolved, but their costs are consequential, and so there is nevertheless merit in endeavoring to work through them. By approaching our deliberations with appropriately open minds and charitably understanding hearts, we may manage to forge a flourishing society— *even without* resolving all our differences.

ACKNOWLEDGMENTS

First, thanks to my dissertation direction, Ruth Chang. My time before and my time after becoming her student at Rutgers has been like night and day. I'm exceedingly thankful for her support, and for her incisive comments on each of the chapters included here.

Second, thank you to my dissertation committee: David Estlund, Alex Guerrero, and Ernest Sosa. Each has provided me with helpful comments and conversation on papers that have now found their way into this dissertation. I am particularly grateful for David Estlund, who stepped in at a crucial moment in my graduate school career and provided invaluable feedback on my work in political philosophy.

Third, I am grateful for excellent discussions had through the years with Chris Hauser, Daniel Rubio, Sara Aronowitz, Amir Saemi, and Rutgers's Philosophy of Religion Reading Group. Thanks in addition especially to Bainbridge Cowell, who has more than anyone else talked through with me the core motivating ideas behind each of the chapters included here.

Last, thanks to Dean Zimmerman, Frankie Egan, and my undergraduate advisors Rob Reich and Joshua Cohen. Dean and Frankie have been incredibly supportive through my graduate school years. Frankie's encouragement and interest in my work was especially significant to me as I started at Rutgers and found my philosophical footing in graduate school. Rob and Josh showed faith in me at an early point in my philosophical career, and I may very well not have made it to graduate school at all without their guidance and without the model of philosophy they gave me. I was, and continue to be, inspired by the prospect of engaging in philosophy as a collaborative social project aimed at helping to create a just society.

DEDICATION

To Grace Monsma, Mavis Moon, Sandro Yee, Judy Vander Veen, and the congregation of San Jose Christian Reformed Church, for teaching me the value of truth ...*and love*.

To Nima Rahimi, Preetam Sarma, Mark Linsey, and Brian Ferrara for making the vibrant beauty of pluralism palpably present to me.

To Bainbridge Cowell, whose love and support has made this dissertation possible. You have continually inspired me with your sense of wonder, and enlivened my life with hope.

And to my parents and my family, who have borne with me through thick and thin. Thanks to my parents for their model of humility and integrity. Thanks to my Dad for my philosopher's heart, and to my mom for keeping it from going bad.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	v
Dedication	vi
Introduction	1
Chapter 1: On the Liberal Duty to Deliberate with the Illiberal	10
I. The Case of Seyyid Qutb	11
II. The <i>Prima Facie</i> Case Against Shallow Liberal Public Justification.	16
A. The Apparent Inadequacy of Shallow Public Justification.	17
B. The Apparent Need for Deep Deliberation.	22
III. The Deeper Case against Shallow Public Justification.	29
A. The Presumptive Claim to Autonomy.	30
B. The Need for Justificatory Engagement with the Illiberal and the Inadequacy of Shallow Public Justification.	34
C. The Duty for Deep Deliberation.	37
IV. Defenses of Shallow Public Justification.	41
A. Internalism.	42
B. A Failure to Distinguish between the Demands of Abstract and Public Justification.	42
C. An (Unduly) Unflattering Image of the Illiberal.	48
(i) Does Unfair-Mindedness Undermine the Presumptive Claim to Autonomy?. 52	
(ii) Does Unfair-Mindedness Override the Presumptive Claim to Autonomy? ..	52
(iii) Does Unfair-Mindedness Alter the Presumptive Claim to Autonomy?	55

V. Closing Concerns.	58
A. Too Demanding as a Duty for Ordinary Citizens?	58
B. The Short Term Failure of Deep Deliberation.	59
C. The Long Term Failure of Deep Deliberation.	60
Conclusion.	62
 Chapter 2: The Phenomenal Appreciation of Reasons (Or How Not to be a Psychopath)	 64
Introduction.	64
I. <i>De Re</i> Accounts of Implicit Appreciation.	68
II. <i>De Dicto</i> Accounts of Implicit Appreciation.	72
A. A Discursive Account of Implicit Appreciation.	74
B. A Causal Tracking Account of Implicit Appreciation.	76
C. A Functional Role Account of Implicit Appreciation.	79
III. A Phenomenological Account of Implicit Appreciation.	81
A. A Cognitive Phenomenology Account of Implicit Appreciation.	83
B. A Non-Discursive Experiential Representation?	86
C. A Distinctive Moral Phenomenology?	91
IV. Is the Phenomenal Representation of Moral Reasons a General Necessary Condition of Moral Reasons Appreciation?	94
A. The Final Elimination.	95
(i) An Affective Account of Appreciation.	95
(ii) A Discursive Account of Appreciation.	100
B. Making Sense of Moral Motivation.	103
Conclusion.	108

Chapter 3: The Ethics of Socially Problematic Disagreements.....	111
Introduction.....	111
I. Practical Problems for the Epistemology of Disagreement.....	114
A. Practical Considerations, Peer Disagreement, and Conciliationism.....	114
B. From Conciliationism to Doxastic Adjustment.....	118
C. The Plan.....	122
II. Socially Problematic Disagreements and Doxastic Adjustment.....	123
A. Socially Problematic Disagreements: The Cost of Error.....	124
B. Socially Problematic Disagreements: Separation Costs.....	127
C. Disagreement as a Practical and Epistemic Reason for Doxastic Adjustment.....	131
D. Upshots of Practical Considerations for Non-Peer Disagreements.....	134
III. The Purely Practical Approach.....	136
A. The Critic: A Purely Practical Approach to Socially Problematic Disagreements	138
B. The Need for Doxastic Adjustment in the Domain of Reasonable Compromise..	143
(i) “Left over” Separation Costs in Reasonable Compromises.....	144
(ii) The Costs of Doxastic Adjustment.....	146
C. The Need for Doxastic Adjustment in the Domain of Justified Coercion.....	149
(i) The Domain of Justified Coercion.....	149
(ii) Doxastic Adjustment and Disagreements about Justice.....	151
(iii) Doxastic Adjustment and Disagreements about Morality.....	153
IV. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment (Or, Why the Truth Matters to Socially Problematic Disagreements).....	156
A. Can Practical Considerations Give One Reasons to Withhold?.....	157

B. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment, Part 1: Separation Considerations as Truth-Indifferent Reasons for Doxastic Adjustment. .	159
C. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment, Part 2: Separation Considerations as Truth-Motivated Reasons for Doxastic Adjustment. .	164
V. Conclusion.	168
Bibliography.	169

Introduction

We want to be people of principle—but standing on principle in a pluralistic society can be problematic. Disagreements arise, and when they do, the dictates of conscience don't always permit us to put our differences aside. Moreover, ensuing conflicts are apt to be especially costly compared to conflicts stemming merely from competing interests.¹

Insofar as we can't work around moral disagreement, then, we must consider the question of how to work *through* it—and to do so better than we are naturally disposed to. This question is partially strategic, but it is also ethical. And it is this ethical question that drives the dissertation: how shall we approach those who disagree with us on moral matters, when the disagreement is so deep that we cannot in good conscience agree to disagree? In this dissertation, I first examine a particular case in which such a deep disagreement must be worked through, and then draw on work in philosophy of mind and epistemology to set the philosophical foundations for two approaches to deep disagreements that promise to help us make our way more virtuously through them.

The ethics of disagreement has not traditionally been treated as a distinctive area of inquiry in philosophy—as, say, health care ethics, immigration ethics, or environmental ethics has traditionally been treated. Nevertheless, the topic of disagreement has received a great deal of interest from political philosophers, who have long been keen to understand what sort of coercive force a state can permissibly exercise in light of deep differences between subjects' comprehensive visions of the good. This concern—which I'll sometimes

¹ See e.g. Atran et al (2007), Tetlock (2000), Tetlock et al (2000), Timothy (2013).

refer to simply as the concern with ‘deep moral difference’—is not new to post-Enlightenment political philosophy. Medieval authors, for instance, also struggled with understanding how to navigate political differences between religious majorities and opposing religious minorities and outsiders.² Still, in the wake of the European wars of religion of the sixteenth and seventeenth centuries, a particular way of approaching deep moral differences gained increasing popularity among European Enlightenment political philosophers. Philosophers in the “social contract” tradition, like Thomas Hobbes, John Locke, and Jean-Jacques Rousseau, suggested what we might call a “bracketing” strategy as a response to deep moral difference. On this approach, parties with deep moral differences are asked to set aside or “bracket” certain contested elements of their visions of the good, so as to assent to the terms of a core set of governing claims whose merits they can mutually and reasonably agree upon. Today, “public reason” liberals hold up the mantle of this social contract tradition. They argue that parties with deep moral differences can reasonably be expected to bracket certain contentious elements of their diverse comprehensive visions of the good, and mutually assent to a set of core governing claims that are more or less coterminous with the essentials of a liberal democratic constitution. In particular (public reason liberals say), such parties can reasonably be expected to accept the core liberal political values of freedom and equality, and the familiar rights associated with these values (including, for instance, freedom of thought and religion, freedom of association, freedom of speech, and freedom of movement and occupation).

Of course, not all those who are coerced in the name of liberal democratic values have a comprehensive vision of the good that in fact disposes them to assent to those values. In

² See e.g. Levy (2017) and Forst (2013).

“The Liberal Duty to Deliberate with the Illiberal,” I begin by discussing the figure of Seyyid Qutb (1906-1966), an illiberal Islamic thinker of the twentieth century who spent time in the United States and who has become regarded as one of the primary theoreticians of Islamism in the post-colonial Sunni world. Qutb’s comprehensive vision of the good led him to resist liberal democratic values, and I offer him as an example of someone who might live in a liberal democratic regime and be coerced in the name of its values, despite deep objections to those values. On the surface, it might appear that the objections of someone like Qutb to liberal democratic values would create a problem for public reason liberals. After all, don’t they say that we should bracket contested elements of our moral vision, to accommodate disagreement? And so shouldn’t they now bracket their own commitment to liberal democratic values, to accommodate Qutb? However, the fact that someone like Qutb would object to those values is not by itself decisively problematic for public reason liberals. They never said parties to a disagreement should always bracket any contentious element of their moral view, but only that they should bracket contentious elements of certain sorts, having to do with particular *comprehensive* aspects of their vision of the good. The bracketing strategy does not go so far as to demand that we compromise with someone like Qutb—and it was never intended to.

In a sense, then, the bracketing strategy runs its course as a response to disagreement once we confront someone like Qutb. If bracketing is our only strategy for responding to disagreement, then when we confront Qutb we will not have much to say to him. And indeed, public reason liberals have not historically had much to say to the illiberal. Rawls (1993) suggests that we “simply” say to someone like Qutb that his views are politically

unreasonable; “within political liberalism,” Rawls writes, “nothing more need be said.”³ Indeed, on Rawls’s view, we need not even venture to say that Qutb’s political views are *false*, and that our views are *true*—it is enough just to say that his political views are unreasonable, and that ours are reasonable.⁴ Since then, public reason liberals have widely come to agree that we can and should tell Qutb that his political views are *false*, and insist on our own political views as superior in part by virtue of the way they get at the *truth* about the values that should underwrite political justification.⁵ Nevertheless, that is about as far as things have come. Indeed, it has become popular to insist on an “internalist” view of liberalism, on which we liberals simply stipulate core liberal principles as starting points, and then focus our efforts on finding a program of political justification that can be coherently worked up from those principles. In being internally oriented in this way, it is thought, liberals escape the need to take on any “externalist” task of justifying their core principles to illiberal citizens.⁶

In my chapter, however, I argue that this is a mistake. For, I argue, liberal values *themselves* require us to accord fundamental importance to citizens’ realization of political autonomy. By virtue of being free and equal, I argue, citizens have a presumptive claim to a feasible opportunity to realize political autonomy. To appropriately honor that claim,

³ Rawls (1993), p. 178. See also p. 246–247.

⁴ Rawls (1993) famously suggests that in the context of discussions of liberal political legitimacy, questions of *truth* should be bracketed. “Once we accept the fact of reasonable pluralism,” he writes, “the idea of the reasonable is a more suitable basis of public justification... than the idea of moral truth” (p. 129.) On Rawls’s view, even the political conception of justice he advocates is right for us to accept, not because it’s true, but because it’s reasonable. On his view, “holding a political conception is true, and for that reason alone the one suitable basis of public reason, is exclusive, even sectarian, and so likely to foster political division” (p. 129). As a result, his political liberalism “does without the concept of truth” (p. 94).

⁵ See e.g. Estlund (2009), Quong (2011), Raz (1990).

moreover, I argue that we must engage illiberal persons like Qutb in *Deep Deliberation*. We must offer them Deep Reasons—reasons that don’t already presuppose agreement on liberal values and so can speak to them in the dialectical position they start from—as to why they should find liberal values and justifications persuasive, and reflectively consider what they have to say in return. One upshot of my argument is that we cannot on the one hand claim that it is *reasonable* to expect citizens to bracket relevant aspects of their comprehensive vision of the good, and on the other hand refuse to give them a *reason* for why they should do so.⁷ A second upshot is this: even after we have taken the bracketing strategy as far as we can in response to a disagreement, we still have more to do.

At the end of his *Liberalism and Perfection*, prominent public reason liberal Jonathon Quong claims that illiberal persons are “*beyond the reach*” of liberal public reason, since “liberalism’s core values do not ‘speak’ to them, or they are incapable of ‘seeing’ their moral force.”⁸ Setting aside the claim about “incapability”—which is by my lights far too strong of a claim to be plausible—Quong’s contention here gets at something important that ultimately serves as motivation for the concerns my second chapter is designed to address. Whether we are thinking of illiberal citizens’ relation to liberal values, liberal citizens’ relation to illiberal values, or (more generally) *any* citizen’s relation to the values of another citizen who holds a deeply different moral view, there is a worrisome consequence to a state of affairs where one party does not “see” the moral force of the other party’s values. This worrisome consequence is *not* necessarily the consequence that might

⁶ Quong (2011) is the most recent and most explicit advocate of this approach, though it also shows up in, for instance, Rawls (1993) and Gaus (2010).

⁷ Rawls himself says that “justice as fairness is not reasonable in the first place unless it in a suitable way can win its support by addressing each citizen’s reason, as explained within its own framework” (1993, p. 143).

⁸ Quong (2011), p. 314.

be thought to be implied if we use the word “see” in a factive sense. That is, the worrisome consequence is not necessarily the consequence that citizens may fail to *recognize* the moral force of values, held by other citizens, that actually pertain—for it’s surely the case that *some* citizens’ values are mistaken. Rather, what is worrisome, that is implied by a more subjective reading of the word “see,” is that citizens may not even be able to understand what it *would be* to see the world as instantiating the moral values that those they disagree with take it to instantiate. This is worrisome because it is crucial to maintaining citizens’ resolve to work through a deep disagreement that they preserve a basic level of goodwill towards each other. But to maintain goodwill towards one’s interlocutor in a moral disagreement, you must plausibly be able to (at least) understand her as motivated by considerations that she sincerely apprehends as moral reasons to ϕ . And it is difficult to genuinely understand her in this way when you can’t understand *how* she could see *those* considerations as moral reasons to ϕ . In these circumstances, we want to know: how do I come to understand my interlocutor as responding to considerations she genuinely apprehends *as moral reasons*?

This question provides the motivation for the second chapter of the dissertation, “On The Phenomenal Appreciation of Reasons (Or How Not to be a Psychopath).” To answer the question, we need to know what it is for someone to appreciate a consideration as a moral reason, and that is the issue I address in my last chapter. Drawing from resources in the philosophy of mind, I examine and reject four proposals for how to account for moral reasons-appreciation, three of which are derived from popular accounts of mental representation (inferentialist, causal tracking, and functionalist accounts of mental representation). I then argue that to appreciate a consideration as a moral reason to ϕ is to

present it under the light of a particular phenomenologically-mediated mode of presentation: one that presents the relevant consideration via the light of a kind of felt directive force “pointing” towards ϕ -ing—lending weight to it, or soliciting it—in a particular authoritative way. Though I don’t go on in the chapter to give an explanation of how this account of moral reasons-appreciation bears back on the initial concerns that motivated the chapter, the basic upshot of the account seems to be this: to understand someone as responding to a consideration she sincerely apprehends as a moral reason requires imaginative simulation of what it is like to “feel the force” of the relevant consideration as a reason. You must be able to simulate what it is like to have the consideration the other regards as a reason to ϕ “call out” to you (as it does to her) with a solicitive force directed at ϕ -ing.

In the final chapter, “The Ethics of Socially Problematic Disagreements,” I circle back to address the questions about disagreement that motivate the dissertation in a more general way. Rather than focus on a particular instance of a disagreement (as I do in the first chapter), I begin the chapter by defining a general class of disagreements that I call “socially problematic” disagreements. As I define them, these are disagreements where, first, there are “costs of error”—if parties to the disagreement act on an *erroneous* judgment about the relevant contested claim p , some non-trivial harms will come to them or to others in virtue of that action. Second, these disagreements are disagreements where there are “separation costs”—social costs that will be sustained in virtue of the persistence of the disagreement, like, e.g., the loss of mutual goodwill, or, as I focus on, the loss of full mutual self-determination. I consider a variation of the “bracketing” strategy advocated by public reason liberals in the political realm—that is, in this case, a strategy whereby parties

agree to set aside certain differences in order to mutually agree to abide by the terms of a reasonable compromise. As in my first chapter, I don't deny the important virtues of this approach. However—and also as in that first chapter—I focus on the limits of the approach. I argue that there are a range of cases in which, even *after* we have done all we reasonably can do to accommodate those we disagree with, there persist significant social costs to disagreement. Foremost among these cases are cases that fall in (what I call) the “domain of justified coercion,” where the disagreements bear on matters of justice and morality where the truth about what is right cannot easily be set aside. I argue that the social costs of disagreement in these cases give rise to reasons for disagreeing parties to make adjustments to certain of their *doxastic* attitudes. In particular, taking as a starting point certain discussions in the epistemology of disagreement, I argue that in the range of disagreements I demarcate, relevant social costs give parties to the disagreement compelling reasons to “doxastically adjust” by reserving judgment about or reconsidering their contested views, and inquiring further into the relevant contested questions. It might appear that any reasons for doxastic adjustment that could be generated by such *social* costs of disagreement would be reasons that would have to be in some way insensitive to the value of the truth. As a result, it might appear that subjects who make the relevant sorts of doxastic adjustments might be prone to be problematically insensitive to the costs of error that also pertain to these relevant socially problematic disagreements. However, I argue that that is not the case. To the contrary, I argue that the social costs of disagreement can sometimes give disagreeing parties reason to be *more*, rather than less, concerned about the value of truth, and to doxastically adjust on *that* basis.

Much more work remains to be done on the ethics of disagreement generally, and on the ethics of deep disagreements in particular. I hope, however, that in the course of my three chapters, I at least convince the reader that is *possible* to engage with those with whom we cannot agree to disagree, in an ethical way, and (moreover) that there is important ethical *value* to endeavoring to make good on this possibility. Since the bracketing approach to disagreement only gives us resources to respond to those with whom we *can* agree to disagree, it leaves us without the resources we need to carry out this further task of ethically engaging with those with whom we cannot agree to disagree. I hope that in addition to drawing attention to the possibility of such ethical engagement, I also begin to provide some of the much-needed resources for undertaking that task.

Chapter 1

On the Liberal Duty to Deliberate with the Illiberal

In this paper, I argue that there are *fair-minded* illiberal citizens—persons who are, despite their dissent from liberal values, actively concerned to cooperate with others on fair and mutually endorsable terms—and that the policy of public justification that public reason liberals currently favor is not adequate to show regard for such citizens' claim to a feasible opportunity to realize political autonomy.⁹ I'll argue that to show regard for this claim, we should engage fair-minded illiberal citizens in *Deep Deliberation*. We should offer them *Deep Reasons*—reasons that don't already presuppose agreement on liberal values, and so can speak to them in the dialectical position they start from—as to why they should find liberal values plausible, and reflectively consider what they have to say in return.

My argument will proceed as follows. First, I introduce an example of what a fair-minded illiberal citizen might look like. In particular, I examine the views of Seyyid Qutb (1906-1966), an illiberal Islamic thinker who has become regarded as one of the primary theoreticians of Islamism in the post-colonial Sunni world.¹⁰ Second, I show how the policy of *shallow liberal public justification* that public reason liberals advocate fails to facilitate someone like Qutb's realization of autonomy, and I offer a preliminary view of

⁹ I have in mind liberals like Estlund (2007), Gaus (2010), Gutmann and Thompson (1996), Rawls (1993), Quong (2011), Vallier (2014), and Weithman (2010). In this paper I engage primarily with Gaus (2010), Rawls (1993), Quong (2011) and (to a lesser extent) Weithman (2010). Public reason liberalism requires that the rules that regulate our common political life be, in one way or another, justifiable or acceptable to all those persons who are subject to their authority. For helpful overviews see Quong (2018) and Gaus (2003).

why that failure is problematic, in light of our own liberal democratic concern with political autonomy. Third, I give a more detailed account of how citizens' presumptive claim to autonomy is grounded in their status as free and equal, and offer an argument for why we should think that presumptive claim gives rise to a defeasible duty for us to engage illiberal citizens in Deep Deliberation. In Section Four, I address three objections that might be raised to my view. My response to the last objection leads to a discussion of the distinction between fair-minded and *unfair*-minded illiberal citizens. I grant that unfair-minded illiberal citizens' presumptive claim to a feasible opportunity to realize autonomy may be defeated, but argue that it is not similarly defeated with respect to *fair-minded* illiberal citizens, and that as a result the duty for Deep Deliberation with them remains intact. In Section Five, I close by briefly addressing concerns about whether and how Deep Deliberation can realistically be carried out.

I. The Case of Seyyid Qutb

Qutb agrees that democratic ideals contain some praiseworthy promises. On his view, though, these promises prove empty when we rely on secular democratic legislation to fulfill them. He writes:

When Europe rebelled against a Church that tyrannized under the false guise of religion, it tried to run away from God. People in Europe thought that they could best preserve their freedom, dignity and humanity under democratic government. They pinned their hopes on the guarantees provided by democratic constitutions, parliamentary systems, a free press, judicial and legal checks, majority rule, and similar ideals. But what happened in practice? Capitalism managed to exercise tyrannical power, reducing all checks and institutions into little more than slogans or myths. The great majority of the people became subservient to the powerful minority that owned the capital which enabled it to control the parliamentary

¹⁰ Qutb was also an early leading member of the Muslim Brotherhood. See Calvert (2010), Irwin (2001), and Musallam (2005).

majority, the constitution, the press and all other checks and balances that people imagined would guarantee their freedom and other rights.¹¹

In a secular democracy, Qutb claims, capitalism breeds materialism, and materialism breeds disregard for others' shared human dignity.¹² As those in power become blinded to the dignity of their disadvantaged neighbors, they use their power to circumvent secular safeguards of the marginalized and eventually drive the latter into practical enslavement to their own high and mighty material ends. Citizens are deprived of their effective political freedom, then, and the affairs of a secular democratic society come to be decided by "fear and threats."¹³

Qutb advocates for the institution of Sharia-based legislation as an antidote to the above ills of secular democracy.¹⁴ On his view, Sharia-based regulations nudge people to see each other in less materialistic ways, and—in this regard and also more generally—better sustain citizens' equal realization of human flourishing. Sharia-based legislation thus better protects persons' equal human dignity than secular democratic laws can.¹⁵ In a regime where Sharia law is established, Qutb says, political arrangements will

¹¹ See *In the Shades of the Quran* (from here on out, "SQ"), p. 15 and p. 267. See also *Islam: The Religion of the Future* (from here on out, "RF"), p. 62–63 and p. 71–74, and *Social Justice in Islam* (from here on out SJ), p. 315–318. I use abbreviation for names of texts rather than years because many of these texts are not associated with any definite year of publication.

¹² See especially RF, p. 71–74 and SJ, p. 315–318.

¹³ RF, p. 68 and *Milestones* (from here on out *M*), p. 27.

¹⁴ Qutb's advocacy of Sharia-based legislation is based on much more than just a (purported) advantage over secular democracy, of course. In addition to the (purported) advantages of such legislation discussed here, Qutb also had a great deal to say about the advantages of Sharia-based governance over Communism. See e.g. SJ p. 77–78, 315–316.

¹⁵ SQ, p. 18; RF, p. 71–74; SJ, p. 68, p. 315–318. I use the term "Sharia-based legislation" rather than "Sharia law" because I take Sharia to be better construed as a system of moral norms that inform law, than as itself a system of laws in our normal legal sense of it. (For a discussion of the latter, we would need to consider the specific jurisprudence of one of the Islamic *madhhabs*, which Qutb does not do).

work towards “the benefit of all” and there will be a “unity of law to which both ruler and ruled [submit].”¹⁶ He writes:

When the human conscience has come to know [the Islamic faith], it will have no need of anyone to preach equality to it in words... it will not endure the distinctions that arise from worldly values at all.... Even the rich and the powerful will support it, because their conscience acknowledges those values that Islam is intent on establishing and confirming... This is what actually happened in Islamic society fourteen centuries ago.¹⁷

As the above begins to allude to, Qutb thinks that Sharia-based legislation is legislation that people, considered generally, can be persuaded to endorse. On Qutb’s view, once someone gets a proper understanding of Sharia and so sees the good it can accomplish, she won’t be able to “help [but be] affected by it”—she will feel how profoundly it corresponds to her human nature and, as a result, eventually accept it as authoritative.¹⁸ In contrast to secular capitalist democratic regimes that ultimately survive on threat of force, then, Qutb thinks that more or less all people could ultimately be persuaded to affirm the Sharia-based legislation he advocates to impose.¹⁹

Qutb echoes some of the same kinds of fundamental moral concerns that motivate our allegiance to liberal values—care for human dignity and the common welfare, equity, mutual responsibility, and persons’ ability to affirm the political structures they are subject to. Most importantly for our purposes, Qutb has a concern for the ideal of cooperating with

¹⁶ *This Religion of Islam* (from here on out RI), p. 71.

¹⁷ SJ, p. 68.

¹⁸ SJ, p. 169. More generally, see March (2010). Notably, Qutb’s approach here has something in common with how Rawls suggests citizens’ may come around to accept his liberal political conception of justice. Both seem to depend to some degree on the idea that citizens’ moral experience in the state will lead them to just “see” the moral benefits of the principles of justice each advocates for, and to accept those principles on that basis. See e.g. Rawls (1993), p. 160, Section Three, Part C and Footnote 59.

others on fair terms. Given the familiar moral values he echoes, and the way he weaves them together into a broader political view, we can recognize that he is genuinely invested in the value of a way of living together that we also aspire to—at least in broad terms—when we aspire to create a fair and just society. We do not need to work out a general account of the boundaries of the concept of fairness, then, to be able to say that we can recognize Qutb as having—and valuing—a *concept* of fair cooperation.²⁰ He is not (say) a Nazi or a ruthless and corrupt mob boss who might also espouse a concern with “fairness,” but whose sense of what that means is not—or is only in the barest way—recognizable to us as having any bearing on *fairness*.

Though Qutb’s concern for fair cooperation is recognizable to us *as* being a concern for fair cooperation, Qutb nevertheless has a very different *conception* of what fair cooperation amounts to than we do:²¹ his way of working out concerns relevant to fairness ultimately move him to embrace some decidedly illiberal approaches to politics. Qutb regards people as sharing uniformly in fundamental human dignity and wants cooperation between them to be fair, but he doesn’t measure that fairness against the kind of background presumption of equal liberty that (as we’ll later see) is fundamental to liberal

¹⁹ He writes (for instance): “Islam is concerned to persuade the conscience in the case of every duty which it prescribes. It imposes no more duties that are demanded by the welfare of society and no more than can be accepted by the ability of the general mass of mankind” (SJ, p. 108–109).

²⁰ The word “can” in “can recognize” is important here. That someone shares (enough of) our concept of fairness isn’t something we always *do* recognize right off the bat. We should be careful not to say that we *cannot* recognize a person’s use of “fairness” as a reference to *fairness* just because we *don’t* yet recognize it as such. Indeed, this is true even when we are talking about cases like the Nazi or the mob boss. There are probably at least some Nazis and mob bosses whose use of “fairness” we could recognize as a (badly misinformed) reference to *fairness* if we talked to them enough. Empirical misinformation can do a great deal in disguising a recognizable reference to *fairness*.

²¹ For discussion of the distinction between a concept and a conception see e.g. Rawls (1971/1999), p. 5. There are of course many discussions of this distinction; I have found Solum (2004)’s discussion especially helpful. Thanks to David Estlund for reminding me of the distinction in this context.

approaches to justice. Such liberal approaches to justice measure fairness against the baseline assumption that persons have an equal political claim to pursue their own chosen conception of the good, while Qutb instead measures fairness against the baseline assumption that persons have an equal political claim to be able to enjoy comprehensive human flourishing—and, of course, he understands such flourishing in Islamic terms.²² As a result, Qutb is prepared to impose a version of Sharia-based law that gives systematic preference to Islamic practices over the practices that might be favored by those of other religious and philosophical worldviews. More generally, he advocates for far-reaching social regulations that would restrict various freedoms we liberals take for granted. He claims, for instance, that voting rights should be restricted to Muslims,²³ that “free mixing” between men and women should be legally limited, that extra-marital sexual relations should be criminalized, and that evolution (and other views he takes to be anti-Islamic) should be not be permitted to be taught in public schools.²⁴

Standing at the distance at which liberal political theory typically encounters illiberal citizens like Qutb, we might easily be inclined to see him as intent merely on imposing his own religious views on others,²⁵ without any genuine concern for cooperating with those

²² Importantly, in his early work (as in *Social Justice in Islam*) Qutb talks a great deal about freedom. The kind of freedom he discusses, though, is explicitly one that involves a kind of human flourishing that he takes only to be available to Muslims.

²³ SJ, 120.

²⁴ UP p. 31, 58, SJ p. 78, 288–90.

²⁵ In fact Qutb did not see himself as imposing Islamic belief *per se*, but only creating just conditions that would also eventually motivate persons to willingly opt in to such belief. Qutb was aware of the picture that Westerners had painted of Islam as “a violent movement which imposes its belief upon people by the sword” (M, p. 76)—and explicitly and flatly rejected that picture not just as false, but as maliciously deceptive. Sharia-based law, he thought, clearly demanded that all citizens be given “complete freedom to accept or not accept” the “spiritual message” of Islam (M, p. 61).

others on fair and mutually endorsable terms.²⁶ That we might initially have failed to appreciate that Qutb is fair-minded in this way is not surprising, since much of his positive policy program manifests what is (from our standpoint) a systematic misassessment of what the requirements of fair cooperation actually are. Still, with even the marginally closer view of Qutb's convictions that the above discussion affords, we can see that Qutb's resistance to secular liberal democracy is not rooted in disregard for the ideal of fair cooperation, but rather in his own positive pursuit of it. As far as Qutb can tell, after all, it is secular liberal political arrangements that, via the workings of pernicious materialist and capitalist influences, undermines our ability to cooperate with each other on fair terms, and Sharia-based law that restores the potential for such cooperation.

II. The *Prima Facie* Case Against Shallow Liberal Public Justification

Despite Qutb's general concern for fair cooperation, his conception of what fair cooperation amounts to is different enough from our liberal conception of the fair cooperation to cause significant difficulty for the prospect of carrying out political life together. Imagine Qutb as a representative of a kind of illiberal citizen that very well might reside in our liberal democratic state today. In so far as we don't use Sharia as a guide to law, and instead impose legislation that fits with secular liberal democratic norms of fairness, the kind of legislation we will tend to impose will tend to be rejected by him as unjustified—and in relatively fundamental kind of way. It is important that we impose fair, liberally-justified legislation, but the dissent of someone like Qutb from that legislation would nevertheless be troubling. For, while we care about establishing liberally just laws,

²⁶ Indeed, as we'll see later, the idea that illiberal citizens are characteristically unconcerned about cooperating with others on fair terms in this way is often used as grounds for dismissing the moral

we *also* care about facilitating citizens' realization of political autonomy—and now it looks like there is a conflict between these two values. In this section, I'll outline that conflict and show how, on face, public reason liberals' policy of "shallow" public justification fails to adequately resolve it. I'll also give a preliminary suggestion as to how the alternative policy of Deep Deliberation I advocate for offers a better way of responding to the conflict.

A. The Apparent Inadequacy of Shallow Public Justification

It is a core requirement of liberal legitimacy that we should show fair regard for others as free and equal partners in a project of collective self-governance. As public reason liberals correctly note, this requirement gives rise to certain substantive requirements on the kind of legislation we should establish. Most broadly, we must establish legislation that stands in alignment with the liberal democratic values that support free and equal citizenship. More specifically, we must establish legislation that is *abstractly justified*—that is sufficiently supported by liberal justifying reasons that reflect an acknowledgement of liberal political values (*viz.*, chiefly, the values of political freedom and equality that underwrite free and equal citizenship).²⁷ So, for instance, legislation that restricts voting rights to adults might be appropriately imposed on the basis of the abstract justification that *children have not developed general capacities of moral reasoning*. On the other hand, such legislation could not be appropriately imposed on the basis of the abstract justification that *children have not yet understood the truth of Islam*. The latter consideration would do nothing to support that legislation in the way relevant to abstract justification, since

significance of their dissent from liberal legislation.

treating this feature of children as salient to whether they should vote does not treat them as free and equal. Different public reason liberals work out which reasons count as substantively liberal in the needed respect in different ways,²⁸ but I'll refer to whatever the relevant class of reasons is as *qualified liberal reasons*.²⁹ Following other public reason liberals, I will say legislation is abstractly justified if there are sufficient qualified liberal reasons speaking in favor of it, and assume that we have a duty to ensure that legislation is thus abstractly justified.

Though ensuring that legislation is abstractly justified is a necessary part of showing regard for other citizens as free and equal, it is not sufficient. This is because showing regard for others as free and equal puts constraints not just on the substance of the justifications for legislation we depend on, but also on *the manner in which we use* those justifications in the process of the enactment of the law. For consider: a philosopher king might secretly divine that a given piece of legislation is abstractly justified by qualified liberal reasons, and impose legislation on those—entirely privately held—grounds. That legislation will be abstractly justified. Nevertheless, citizens would not have been treated as fully free and equal in her regime. This is because, as free and equal, citizens should not simply be pushed around by those in power; instead, they should be able to act as moral agents to freely comport their conduct with the law's demands on the basis of their own recognition of relevant reasons as justificatory.³⁰ As a result (and as we'll say much more

²⁷ Thanks to David Estlund for putting me on to using the category of abstract justification in the way that I have used it here. The term is used in slightly different ways in different contexts, but my usage falls broadly in line with the standard usage. See e.g. Rawls (1993), p. 386–388 and Gaus (2010), p. 41, 42.

²⁸ For more on this see Section IV, B.

²⁹ Thanks to David Estlund, who I believe suggested this particular terminology to me.

³⁰ See e.g. Gaus (2010), p. 16, Quong (2011), p. 2 and Rawls (1993). Rawls (1993) writes (for instance) “It is fitting, then, that the fair terms of social cooperation between citizens as free and equal should meet the

about later) they have at least some presumptive claim to a feasible opportunity to come to understand, examine, and endorse the justificatory reasons for the coercive legislation they are subject to *as justificatory reasons* for themselves. They have, as I'll put it, a presumptive claim to a feasible opportunity to realize political autonomy—or, more simply, *a presumptive claim to autonomy*. Though the legislation imposed by the philosopher king is abstractly justified, citizens in her regime are not genuinely treated as free and equal because this presumptive claim to autonomy has apparently been ignored: legislation has been one-sidedly foisted on them for reasons they have not had any feasible opportunity to come to understand, examine, and endorse as justificatory for themselves. We who hope to secure the liberal legitimacy of law must improve on the philosopher king's approach.

Public reason liberals typically argue that we should improve on the philosopher king's approach by ensuring that legislation is not just abstractly justified, but also presently *publicly justified*.³¹ We should show regard for citizens' presumptive claim to autonomy, that is (they suggest), by subjecting legislation to a process of public deliberation in which citizens have an opportunity to come to understand, examine, and endorse for themselves

requirements of full publicity. For if the basic structure relies on coercive sanctions... the grounds of its institutions should stand up to public scrutiny... Publicity ensures... that citizens are in a position to know and to accept the pervasive influences of the basic structure... that citizens should be in this position is a condition of their realizing their freedom as fully autonomous, politically speaking" (p. 68).

³¹ Liberal theorists have sometimes used the term "public justification" to refer to a variety of the kind of abstract justification I have in mind. As discussed in Section IV, B, many understand abstract justification as itself a matter of what certain kinds of idealized, qualified liberal parties could accept. As a result (and somewhat confusingly) abstractly justified legislation is also sometimes described as "publicly justified" because it is justifiable to such parties. Here and throughout, I use the term public justification in the sense discussed in this paragraph, as an activity involving the mutual sharing and reflective consideration of reasons. This is a secondary use of the phrase that public reason liberals also recognize, and it is (by my lights) important to keep these two distinct uses distinct. See the next footnote for references to public reason liberals discussion of what I am here calling present public justification.

the reasons that will serve to justify their political coercion.³² More specifically, public reason liberals argue that legislation should be subject to a process of public deliberation in which citizens offer each other and deliberate about the qualified liberal reasons that serve in abstract justification.³³ For reasons that will become clear shortly, I'll call this kind of public justification *shallow liberal public justification*.³⁴

Unfortunately, it does not appear that shallow liberal public justification does enough to improve on the philosopher king's objectionable approach. For shallow liberal public justification does not require citizens to go beyond offering each other qualified liberal reasons for legislation, and these are reasons that only some (liberal) citizens will typically be in a position to understand and endorse as justificatory. Shallow liberal public justification, thus, apparently only shows regard for *those* liberal citizens' presumptive claim to autonomy. Given the congruence between liberal values and qualified liberal reasons, those citizens (who hold liberal values) are apt to be able to appreciate the justificatory merits of qualified liberal reasons with some not unreasonably demanding amount of reflection on their values. They will thus have a feasible opportunity to come to

³² See e.g. Rawls (1993), Lecture VI (especially on the duty of civility, see e.g. p. 216–227, 253), p. 77–79, p. 386, “Reply to Habermas” p. 430ff, “The Idea of Public Reason Revisited” (especially p. 448–462) and Quong (2011), Chapter 9. For more references on Rawls see Footnote 54. Importantly, Rawls is concerned (first and foremost) to advocate for shallow present public justification of the sort I discuss here (the sort that makes use only of qualified liberal reasons) as it bears on constitutional essentials and matters of basic justice (p. xlviii, p. 215). Quong argues that shallow public justification of this sort should extend more broadly—as he puts it, “whenever possible, to all decisions where citizens exercise political power over one another” (p. 274; see also p. 273–275). The importance of present public justification shows up in Gaus (2010) in his discussions of full public justification (e.g. p. 42–53, p. 455) and in his insistence that citizens should be able appreciate relevant reasons that figure in the justification of their coercion so long as they undertake a “respectable amount” of good reasoning (Chapter 13). See also Gaus (2015), where he replies to relevant criticisms by Enoch (2015).

³³ See previous footnote.

³⁴ The term “shallow” here is inspired by Raz (1990, p. 8–9) though I take him to be more concerned with abstract justification than with public justification.

endorse relevant liberal legislation as just and realize political autonomy.³⁵ Illiberal persons are free and equal citizens too, though, and so they too (it seems) have a presumptive claim to autonomy—but a policy of shallow liberal public justification will not similarly serve to honor this claim for them. The qualified liberal reasons we would offer someone like Qutb in support of legislation would not be apt to be understood and endorsed by him as justificatory, since those reasons would appeal to political values that he rejects; in particular, those reasons would often be understood by him to be insufficiently sensitive to the value of comprehensive human flourishing.³⁶ What's more, given the illiberal political framework Qutb starts with, even a reasonable amount of reasoning and reflection on his part would not be apt to change his mind about this. So, so long as we only made use of qualified liberal reasons in our practices of public justification, we would systematically preclude Qutb from having a feasible opportunity to realize political autonomy. Qutb would fare no better in our polity than citizens more generally would fare in the polity of the liberal philosopher king.

This is an in principle problem, but the practical force of the problem can be brought home by considering the concrete consequences of shallow public justification for our political relationship with illiberal citizens like Qutb. Without a feasible opportunity to come to see the reasons they are coerced to conform to as justificatory, illiberal citizens'

³⁵ It is not expected that even liberal citizens will endorse every piece of legislation as perfectly or maximally just; it is thought, though, that they will be able to endorse it as meeting at least some minimum threshold of justice. See Gaus (2011), p. 249ff for a related discussion of a "respectable" amount of reasoning.

³⁶ Thus, for instance, legislation that gives citizens of all religious orientations a right to vote is abstractly justified on the basis of (perhaps, and certainly among other things) the fact that all citizens have an equal claim to exercise political influence so as to protect the pursuit of their own comprehensive vision of the good. But Qutb would reject that legislation and those grounds for it. He would say that non-Muslim citizens should not vote (because they lack insight into how to steer society towards the proper Islamic comprehensive moral goods), and he would deny that citizens have an equal claim to exercise political influence so as to protect the pursuit of their own comprehensive vision of the good.

decisions to abide by the law will to a significant extent have to be motivated by the brute force of power we threaten to exert on them.³⁷ As a result—and as liberal theorists readily and explicitly acknowledge—a policy of shallow liberal justification has the effect of making our relation with the illiberal into a kind of a power-brokered *modus vivendi*.³⁸ We relate to illiberal persons at best (as Gaus puts it³⁹) “strategic partners” or “patients to be helped”—or, otherwise, as “dangers” on par with (as Rawls puts it in a related context) “war and disease” that simply need to be “contained.”⁴⁰ Whatever the precise form of these power-brokered relations, they are not intuitively fitting to someone like Qutb’s status as a free and equal citizen. To the contrary, these ways of relating to those who ideologically dissent from the principles of a political regime are intuitively more suitable to a regime (perhaps like that of the philosopher king’s) that regards such dissenters as brute forces to be subdued or as intractable children, than to a regime that regards all persons as free and equal. In practice as well as in principle, then, it appears that the failure of shallow public justification to facilitate feasible opportunities for illiberal citizens like Qutb to realize political autonomy, manifests a failure to show full regard for them as free and equal citizens.

B. The Apparent Need for Deep Deliberation

Since it is a core requirement of liberal legitimacy that we should show fair regard for citizens as free and equal, and since a policy of shallow liberal public justification is (on

³⁷ Thanks to David Estlund for pointing out that there are a variety of reasons to act as laws require even apart from those laws having proper political authority. See Simmons (1981).

³⁸ Rawls (1993), p. 489. Interestingly, Weithman (2010) floats the idea that stability in actual liberal democracies may generally be based more on a *modus vivendi* than on the kind of overlapping consensus that would obtain in a well-ordered liberal society (p. 321).

³⁹ Gaus (2010), p. 210, 282–283.

face) not fully adequate for fulfilling that task, a policy of shallow liberal public justification is (on face) not fully adequate to support the legitimacy of liberal law. But how should we alter our practices of public justification to make them more sufficient to support the liberal legitimacy of legislation? There are two further kinds of reasons that we could offer illiberal citizens like Qutb in the course of public deliberation besides qualified liberal reasons, so as to try to secure their understanding endorsement of the liberal legislation we impose on them. First, we could offer them substitute, *unqualified illiberal reasons* that appeal to and so presuppose the soundness of illiberal political values. We could, for instance, tell Qutb that he should accept universal voting rights because, if he does so, more non-Muslims will be eventually be attracted to his overall Sharia-based policy program. Public reason liberals have rejected this strategy of engaging with illiberal citizens,⁴¹ and I'll assume for the purposes of this paper that they have been right to do so. First and most obviously, such a strategy would be almost guaranteed to fail at reconciling Qutb to a great deal of liberal legislation, since there is plausibly only so far that illiberal reasons can be stretched to support legislation that is actually selected for its fit with qualified *liberal* reasons. Second (as public reason liberals have pointed out) to offer Qutb illiberal reasons as means to try to secure his assent to liberal legislation would (as far as we can tell) be to offer him no *genuine* reasons for assenting to legislation at all. To offer him such reasons would thus, from our perspective, be to engage in a manipulative pretense with him⁴²—to purchase his cooperation while handing him play money. Third (and even if one rejects the worry about manipulation), it's not clear that offering Qutb substitute illiberal reasons does

⁴⁰ Rawls (1993), p. 64 n. 19.

⁴¹ See next footnote.

much to improve on showing regard for him *as a free and equal citizen*. A sportswoman engaged in a game of basketball is not shown respect *qua* basketball player when the referee offers her justifications for his call that accord with the rules of netball, just because those are the sorts of justifications that—given her current predilection for netball—she is now apt to accept. Similarly (it might reasonably be objected), we do not honor Qutb as a free and equal citizen by offering him illiberal justifications for legislation that are at odds with that status of being a free and equal citizen of a liberal democracy, just because those are the justifications that he is now apt to accept.

Another category of reasons we might offer Qutb to secure his assent to liberal legislation are neutral reasons that do not presuppose his acceptance of liberal political values (like qualified liberal reasons do), but that also do not presuppose the soundness of the illiberal political values we reject (like unqualified illiberal reasons do). Such reasons may involve reference an array of empirical consideration, and also a wide variety of moral values that liberal democrats have no monopoly on. Thus, for instance, think of the familiar moral values we said earlier are also echoed by Qutb—e.g. care for human dignity and the common welfare, equity, mutual responsibility, and persons’ ability to affirm the political structures they are subject.

This class of neutral political reasons can be offered to someone like Qutb in one of two ways—only one of which I will ultimately be concerned with. First, they can be offered in the way I previously mentioned that illiberal reasons could be offered: that is, as *substitutes* for qualified liberal reasons that are meant to *directly* garner Qutb’s endorsement of the relevant liberal legislation we want him to impose on him. I see no obvious reason why this

⁴² See e.g. Quong (2011), p. 265–273. Though Quong advances his worry by countering Gaus, Gaus also has a similar worry, couched in terms of concerns about “unintelligible” reasons (see e.g. 2010, p. 292).

use of these reasons would be intrinsically objectionable, but—as was the case with illiberal reasons—they are almost guaranteed to be insufficient to earn Qutb’s support for a wide swath of liberal legislation. After all, the political values of freedom and equality (and associated qualified liberal reasons) are not playing an epiphenomenal role in determining what legislation is abstractly justified and on that basis suitable for coercive imposition. Neutral reasons will as a result only be able to be stretched so far to serve as bases for acceptance of liberal legislation.

Fortunately, there is a second kind of use to which neutral reasons can be put that is somewhat more promising: they can be used to deliberate with Qutb about the relative merits of liberal political values, and to pursue a more fundamental kind of justificatory rapprochement with him in this way. Because neutral reasons are being used to get “underneath” the cause of our core political differences with Qutb, I’ll from here on out refer to these reasons, used in this capacity, as *Deep Reasons*. I’ll refer to the kind of deliberation in which we use these reasons to pursue rapprochement on core political values *Deep Deliberation*.

Of course it isn’t guaranteed that Deep Deliberation will ultimately work to persuade people like Qutb of the merits of liberal values, or even to bring about a full justificatory rapprochement in some other way (I will have more to say about this later).⁴³ Nevertheless, Deep Deliberation provides what seems to be our only sensible hope of a long term, general solution to the problem at hand. Offering Qutb Deep Reasons in pursuit of justificatory rapprochement is not—like the offering of illiberal or neutral substitute reasons would be—almost guaranteed to fail to facilitate his realization of autonomy in a

⁴³ See Section V and the Conclusion.

general way. More generally, it is not (like the offering of those other sorts of reasons would be) a mere temporary patch for the symptoms of the fundamental justificatory differences that divide us. Rather, offering Qutb Deep Reasons and engaging in Deep Deliberation with him strikes at the root cause of what is working to prevent him from realizing political autonomy in our polity: that is, the divisions between us and him on core political values.

Since Deep Deliberation is apparently critical to creating and sustaining feasible opportunities for Qutb to realize political autonomy, it would seem essential to include it as part our liberal practice of public justification. Nevertheless, public reason liberals have typically denied that this “deeper” element of deliberation is in needed—they have insisted that it is enough that we use only qualified liberal reasons in public justification, and in this sense advocated for a “shallow” form of public justification. Some explicitly reject the need for Deep Deliberation in favor of shallow public justification. Thus, for instance—understanding the politically liberal in terms of the politically reasonable—Rawls (1993) writes:

Political liberalism itself cannot argue that... comprehensive doctrines should find the conclusions of [liberal] public reason... within its leeway. To argue that transcends public reason.⁴⁴

Along similar lines, Quong (2011) writes that illiberal persons are “*beyond the reach of liberal public reason*,” and that “obtaining their agreement to a liberal theory of justice” is

⁴⁴ Rawls (1993), p. 246–247. See also (1993), p. 178: “The idea of the politically reasonable is sufficient unto itself... Of course... [some] assert that the religiously true, or the philosophically true, overrides the politically reasonable. We simply say that such a doctrine is politically unreasonable. Within political liberalism nothing more need be said.”

“irrelevant” to the liberal political project.⁴⁵ Other public reason liberals imply that Deep Deliberation is not necessary more obliquely, by defending shallow public justification and condoning the consequent loss of autonomy for the illiberal.⁴⁶ Thus—though they sometimes express regret about our power-brokered relations with the illiberal,⁴⁷ or hope that in better, more well-ordered liberal democracies, we would be able to avoid the relevant problems altogether by ensuring that citizens don’t grow up to develop illiberal beliefs in the first place⁴⁸—the reports they make about our living with the illiberal on the terms of a *modus vivendi* (about our treating them as “strategic partners” or “dangers”) are ultimately delivered as just that: reports of matters of sad but given facts about the bounds of our justificatory community, and not as grounds for critical concern about the adequacy of shallow liberal public justification. There is thus virtually no positive entertainment of the idea that there is a political *duty* to seek justificatory rapprochement with illiberal citizens. Citizens are of course not forbidden to engage in Deep Deliberation with illiberal citizens, but that is, from the point of view of political legitimacy as they public reason liberals have typically seen it, *optional*. Engaging with Qutb might be a nice and benevolent thing for us to do for him, but in the end it is supererogatory—not something we *owe* him as a free and equal citizen

⁴⁵ Quong (2011), p. 314 and p. 181. See also Quong (2011), p. 5–6, p. 28, 152–153, p. 313–314, and especially p. 234 and p. 242. Quong’s rationale for this stance is discussed in more detail at the start of the next section and in Section Four, Part A.

⁴⁶ Gaus (2010) says that his project is focused on the problem of social morality under conditions of pluralism among “good-will intelligible” moral persons, and suggests that this project excludes from consideration those who do not treat others as free and equal (p. 281–282).

⁴⁷ Gaus (2010), p. 282.

⁴⁸ Perhaps via proper childhood education.

For the reasons I've started to discuss in this section, this seems to be a mistake.⁴⁹ Since illiberal persons are free and equal citizens, they are (apparently) owed a feasible opportunity to realize political autonomy, just like anyone else. As we'll see later in more detail, there may sometimes be excuses that permit us to impose legislation in certain circumstances despite having *successfully* provided such an opportunity. But it is not in keeping with showing regard for someone like Qutb as free and equal, to begin by adopting a policy of shallow public justification that treats it as entirely optional, across the board, whether or not the work that would need to be done to facilitate his realization of political autonomy, is ever even attempted—for that is to suggest that, contrary to fact, the void of a feasible opportunity for him to realize political autonomy would not confront us with any normative “push back” that we are compelled to address.⁵⁰ So, it appears, to show the regard for someone like Qutb that he is due as a free and equal citizen, we must go beyond a practice of shallow public justification. More specifically—and as I'll argue now argue in more detail—we should acknowledge a duty to engage illiberal citizens like Qutb in Deep Deliberation.

⁴⁹ As I'll discuss in more detail later, I grant public reason liberals that there is this grain of truth to their claim: there may often be such overriding importance to ensuring that liberal justice is accomplished, that we can legitimately impose coercive legislation on someone like Qutb even though Deep Deliberation hasn't *yet* succeeded at forging a feasible opportunity for him to realize political autonomy.

⁵⁰ In failing to take the non-facilitation of illiberal citizens' like Qutb's autonomy as grounds to query the adequacy of shallow liberal justification, public reason liberals go beyond just permitting a *modus vivendi* with these citizens: they also express their lack of active normative concern for that state of affairs. They imply, not only that the failure to facilitate someone like Qutb's autonomy doesn't decisively illegitimize our coercion of him, but also (and more significantly) that that failure does not even stack against the legitimacy of that coercion—that it does not confront us with any normative “push back” that we are compelled to address.

III. The Deeper Case against Shallow Public Justification

In this section I'll give a more precise account of how illiberal persons' status as free and equal citizens gives rise to a presumptive claim to political autonomy, and then show why that claim plausibly gives rise to a (defeasible) duty for Deep Deliberation. (For the purposes of argument, I'll assume in this section that the presumptive claim to autonomy is not defeated, but I consider that matter further in the next section.)

Before getting to the primary business of this section, though, it's worth pausing to frame the arguments in it as a response to one of the increasingly popular strategies used by public reason liberals to ward off criticisms of shallow public justification. In a move mimicked by others, Jonathon Quong has recently responded to criticisms about the failure of shallow public justification to justify liberal institutions to the illiberal by claiming that liberals should take an "internalist" approach to political justification.⁵¹ According to Quong, liberals should simply stipulate core liberal principles as starting points, and then focus their efforts on finding a program of political justification that can be coherently worked up from those principles. In being internally oriented in this way, Quong thinks, liberals escape the need to take on any "externalist" task of justifying their core principles to illiberal citizens.⁵²

Despite Quong's claims to the contrary, however, the choice to take an internalist approach to liberal justification does nothing to establish that shallow liberal public

⁵¹ Quong (2011), p. 137–145 (especially p. 140), p. 158. Quong reasonably reads Rawls as being an internalist and this sort. Rawls says, for instance, that a political conception of justice will have "served its practical purpose" if it manages to "[narrow] the gap between the conscientious convictions of those who accept the basic ideas of a constitutional regime," to "[clear] [their] view," and to "[make] [their] considered convictions more coherent" (1993, p. 156).) See also Weithman (2011), Gaus (2010), p. 281–282, and Dreben (2003).

⁵² See previous footnote and Footnote 37.

justification is admissible. After all, even if it is appropriate for liberals to simply begin theorizing with liberal values, and then to focus their energies on constructing an account of political justification that is coherent with it, it may well turn out that a coherent account of justification based on those liberal values will *itself* require that those values—and the justificatory reasons tied to them—should be publicly justified to the illiberal citizens who are coerced in their name. Public reason liberals can’t defend shallow liberal public justification, then, just by stipulating that they are internalists: rather, they actually need to take on the assignment of showing that and how it is compatible with our commitment to liberal values, not to undertake the relevant “externalist” task. We have already seen *prima facie* reasons to think that they will not be able to execute this assignment. In the rest of this and the following sections, I aim to show that these *prima facie* reasons also accord with the deeper facts of the case.

A. The Presumptive Claim to Autonomy

Public reason liberals typically hold that citizens are free in virtue of their possession of two moral capacities, and equal in virtue of having these capacities to the degree required to be a cooperative member of society.⁵³ These two capacities are: (i) a capacity for a conception of the good—that is, a capacity to form, revise, and direct her life according to a conception of the good, and (ii) a capacity for a sense of justice—that is (roughly), a capacity to understand, be motivated by, and willingly govern her political life in

⁵³ See Rawls (1993), p. 18–20, p. 29–35 and Quong (2011), p. 14, p. 38–39. Though Gaus doesn’t make use of the same terminology, the categories of moral power he discusses are similar. See e.g. Gaus (2010), p. 19. Especially importantly for my purposes, though the terminology of a “capacity for sense of justice” doesn’t figure prominently in Gaus’s work, he also makes use of something like a capacity for a sense of justice. See Gaus (2010), p. 254, p. 255, p. 258, p. 279, p. 283, p. 360–2, p. 480, and his claim that “the normative justification for the rules we live by” should be understood “not [to be] beyond anyone’s ken” (2015, p. 1093).

accordance with moral reasons, in particular as they pertain to fair and mutually endorsable terms of political cooperation between herself and others.⁵⁴ According to public reason liberals, to show regard for persons as free and equal citizens is to show regard for the fundamental importance of their development and exercise of these two capacities.

The fundamental importance of citizens' development and exercise of these two moral capacities (the capacity for a conception of the good and for a sense of justice) plausibly gives rise to certain presumptive claims (the presumptive claim to liberty and the presumptive claim to autonomy, respectively), which in turn, plausibly then give rise to two further kinds of justificatory entitlements (concerning abstract and public justification, respectively). A full story of how the moral powers give rise to these claims and entitlements is beyond the scope of the paper,⁵⁵ but briefly, we can understand things as follows.

Though it is less immediately relevant to us, I'll begin with a discussion of the capacity for a conception of the good, so as to better contextualize the overall argument. In so far as we regard it as of fundamental importance that a citizen should be able to develop and exercise her capacity to form, revise, and direct her life according to a conception of the good, it stands to reason that we should (as Rawls says) regard her claims on her political institutions so as to advance her conception of the good as "self-authenticating"—as not needing prior justification in terms of larger duties she owes to society, or some

⁵⁴ See previous footnote for references.

⁵⁵ One important addition to the account I present here that would need to be made is a further discussion of the ways in which the capacity for a conception of the good influences the entitlement to be publicly justified to, and the ways in which the operation of citizens' capacity for a sense of justice influences the kind of abstract justification of legislation that citizens are owed.

pre-ascribed role she has in an alleged imposed natural order.⁵⁶ This entitlement of a citizen to make self-authenticating claims on her institutions, in turn, plausibly gives rise to a *presumptive claim to liberty*—a claim that liberals characteristically accord to citizens of a liberal democracy.⁵⁷ Given that a citizen's claims to pursue her conception of the good are self-authenticating, the state should by default not interfere with her pursuit of the conception of her good. When the state does use its coercive power over a person in a way that restricts, shapes, or otherwise substantially affects her pursuit of her conception of the good, moreover, that will always require that positive justification be made *with respect to her* as to why the relevant coercive power should be used in the way it is. Importantly, of course, not just any kind of justification will do: given that all citizens share equally in their ability to make these self-authenticating claims, justifications of legislation will always need to show that relevant proposed restrictions are fair, *as measured against a background of assumed equal liberty*. As a result, the presumptive claim to liberty (and, more deeply, citizens' capacity for a conception of the good) plausibly generates substantive constraints on the reasons for legislation that may be used to justify their coercion. It is plausibly because of citizens' capacity for a conception of the good, that is, that we are required to ensure that legislation is abstractly justified by reference to fair, qualified liberal justifications.⁵⁸

Given the *second* moral power in virtue of which we are free, however, it isn't enough just that such fair justifications exist in the abstract, or even that some institutional authority has managed to discover them and privately makes reference to them, to defend

⁵⁶ Rawls (1993), p. 32.

⁵⁷ See e.g. Gaus (2010), p. 341ff for more on this.

its political decisions. No, rather: given my capacity for a sense of justice, it becomes important that legislation imposed on me should not only be justified *with respect to* me, but also (in some important sense) *to* me.⁵⁹ Given that I have a capacity to understand, be motivated by, and willingly govern my socio-political life in accordance with moral reasons pertaining to fair cooperation between myself and others, I am not—and shouldn't be treated as though I were—a kind of social loose cannon who must have political directives (even fair and just ones) simply imposed on me. Others, thus, should not defend the coercive legislation they propose to impose on me by relying on justificatory reasons as on esoteric truths that only *they* have special access to.⁶⁰ Instead, given that I have a capacity for a sense of justice, it should be assumed that whatever reasons are fit to justify the imposition of rules that regulate our common political life together, are also reasons I have a capacity to understand and be motivated by. Moreover, and more positively, in so far as it is fundamentally important that I should be able to develop and exercise this capacity to participate in the collective self-governance of my political society, it is fundamentally important that I should be able to come to understand and endorse the reasons that serve to justify my coercion.⁶¹

As a citizen's capacity for a conception of the good plausibly grounds her presumptive claim to liberty, it is plausibly a citizen's capacity for a sense of justice that grounds her

⁵⁸ A fuller account of how the moral powers give rise to relevant claims and entitlements would have a more complicated story to offer here. (I also mention this issue in Footnote 47.)

⁵⁹ For more on the idea of justification *to* see Quong (2011), p. 141–142.

⁶⁰ A regime that treats citizens as free and equal, then, cannot act as certain religious regimes might when they write some off as “sinners” who are simply beyond the reach of the truth.

⁶¹ Rawls (1993), p. 77–78. See also Quong (2011), p. 265–266.

presumptive claim to autonomy.⁶² It is plausibly because it is fundamentally important that citizens should be able to develop and exercise this capacity, that it is also fundamentally important they are afforded feasible opportunities to come to understand and endorse the fairness-pertaining reasons that serve to justify their coercion—and thereby come to realize full political autonomy. Just as the presumptive claim to liberty sets a fairness constraint on the substance of the justifications we employ to justify coercion, then, so also the presumptive claim to autonomy plausibly sets constraints on the *manner* in which we make political use of those justifications in the process of the enactment of the law. It is plausibly because it is fundamentally important that citizens should be able to develop and exercise their capacity for a sense of justice, that we are compelled to adopt a practice of public justification that provides citizens with feasible opportunities to realize political autonomy.

B. The Need for Justificatory Engagement with the Illiberal and the Inadequacy of Shallow Public Justification

We are now in a position to see with greater clarity why our adoption of a practice of shallow justification would fail to manifest regard for illiberal citizens as free and equal. In short, shallow public justification fails to treat illiberal persons as free and equal because it does not manifest regard for the fundamental importance of their development and exercise of their capacity for a sense of justice. But let's lay out this reasoning more carefully.

⁶² In saying this, I elaborate on what it seems suggested by Rawls's account but is not explicitly argued by him—he does not, for instance, at any time use the term “presumptive claim to autonomy.” Rawls writes, for instance, that “It is fitting... that the fair terms of social cooperation between citizens as free and equal should meet the requirements of full publicity. For if the basic structure relies on coercive sanctions... the grounds of its institutions should stand up to public scrutiny... Publicity ensures... that citizens are in a position to know and to accept the pervasive influences of the basic structure... that citizens should be in this position is a condition of their realizing their freedom as fully autonomous, politically speaking” (1993, p. 68). See also Rawls (1993), p. 77, 80 and Rawls (1971), p. 115, where Rawls first introduces the seeds of the idea of public justification in relation to Kantian contractarian theory.

First, showing regard for an illiberal person like Qutb as a moral agent who possesses a capacity for a sense of justice (to the degree required to be a cooperative member of society), means treating him as someone who is (at least) *capable* of being a partner in collective self-governance—and so as someone who is not doomed to be merely a “strategic partner,” “patient,” or “danger to be contained.”⁶³ Perhaps even more to the point, for us: in so far as we hold that the correct (“real”) fairness-pertaining reasons are qualified liberal ones, treating someone like Qutb as having a capacity for a sense of justice means treating him as though he has a capacity to understand and be motivated by those reasons. We may not, thus, treat these reasons or the liberal political values that underly them as esoteric truths that he cannot equally gain access to.

Second, since we not only must recognize that Qutb *has* a capacity for a sense of justice, but also treat its development and exercise as of fundamental importance, it becomes incumbent on us to adopt a practice of public justification that facilitates that development and exercise in a way that is commensurate with its great importance. We must—to put the idea in the less cumbersome term I’ll now switch to using—adopt a practice of public justification that facilitates Qutb’s realization of political autonomy in a way that is commensurate with its great importance. What does such a practice of public justification need to look like?

The first thing to point out is that such a practice of public justification need not generate a feasible opportunity for all illiberal citizens to fully realize political autonomy before any relevant legislation is imposed on them. This is because it’s possible to manifest a level of concern for illiberal citizens’ realization of autonomy that is commensurate with

⁶³ See Footnotes 30–32.

its great importance, *and also* to weigh in the balance other considerations that bear on showing appropriate regard for *other* citizens' status as free and equal. So, for instance, suppose the timely ending of a substantive injustice is at stake, but that relevant illiberal citizens do not yet recognize the injustice as an injustice, and that this is at least in part because they have not had feasible opportunities to facilitate their doing so. In this kind of circumstance, we can take the lack of those opportunities and the resulting non-autonomy of the illiberal as *some* reason to wait until we can better persuade the illiberal of the law's merits, but also weigh even more heavily as a reason to impose the legislation immediately, the more serious threat to freedom and inequality that comes from the threat of injustice. So, we could legitimately impose coercive legislation because *all things considered* that was the best way to show regard for citizens as free and equal, and we gave due consideration to the importance of illiberal persons' realization of autonomy.⁶⁴

We can imagine various extenuating circumstances along the lines of the above that would permit various degrees of excuse for various kinds of failures to successfully provide illiberal citizens' with a feasible opportunity to realize political autonomy (and so develop and exercise their capacity for a sense of justice). But there would need to be very extenuating circumstances indeed to excuse a policy of shallow public justification—and such circumstances are not the ordinary circumstances of political life. For a policy of shallow public justification puts no *requirements* at all on even *engaging* illiberal citizens in an effort to overcome their alienation from the law, and thus has as a predictable result the systematic failure to provide illiberal citizens with feasible opportunities to realize

⁶⁴ In my view, the kind of legitimacy that is possible here is a compromised one since we have in fact failed to treat illiberal persons as *fully* free and equal—in this sense (in my view) we have a “mark” on our record that needs to be made up for. But I also take it that in the real world, comprised legitimacy is often the best we can do.

autonomy. It's hard to imagine circumstances where attempts at such engagement would come at a significant cost of respect for others' freedom and equality. Perhaps in war to save our liberal polity's existence, any such efforts would come at too high a cost of political resources needed to defend freedom and equality elsewhere, and so perhaps in such circumstances, shallow justification would be compatible with a due—albeit outweighed—concern for the exercise of illiberal persons' capacity for a sense of justice. But other than that, it's hard to see how a blanket denial of a need for justificatory engagement with illiberal citizens, aimed to facilitate feasible opportunities for them to come to understand and endorse the justifications for their coercion, could be compatible with genuine regard for the fundamental importance of their development and exercise of their capacity for a sense of justice. It's hard to see, in other words, how shallow public justification could be compatible with genuine regard for illiberal persons as free and equal citizens.

C. The Duty for Deep Deliberation

So suppose that there is a (defeasible) requirement, and in this sense a (defeasible) *duty*, for us to engage illiberal citizens like Qutb as part of our practice of public justification, so as to help afford them with a feasible opportunity to realize political autonomy. What should this engagement look like?

Public reason liberals have commonly thought that public justification should involve an exchange of reasons in some sort of forum of public deliberation. And this stands to reason, since some form of public deliberation is what seems to be most in keeping with giving proper support to citizens' exercise of their capacity for a sense of justice. If that capacity consisted in the ability to understand and be motivated by fair reasons when fed a

special pill, then public justification would properly consist in the distribution of such pills; if on the other hand the capacity for a sense of justice consisted in the ability to understand and be motivated by fair reasons after undergoing a kind of transformative spiritual experience, public justification would need to consist in creating the conditions that are needed to stimulate such experiences. But significantly, these are not the ways that our capacity for a sense of justice is normally understood to operate, and so public justification isn't thought to properly take those forms. Rather, it is normally taken for granted that the capacity for a sense of justice involves the operation of reason—of citizens' use of inference, judgment, and moral imagination, and (often) the synthesis of these powers in the formation of a reflective equilibrium.⁶⁵ Given this, it is fitting that public justification should take the form of *deliberation* rather than something else, for it is via the exchange of reasons that we engage other citizens' powers of reason, and form and re-form our respective reflective equilibriums so as to move towards accord on contested issues.⁶⁶

⁶⁵ See e.g. Rawls (2001), p. 24, 29–32 (especially p. 29), (1993), p. 19, p. 81 and Quong (2011), p. 14, p. 42–43, p. 265–266. For Gaus, see Footnote 45 and (2010), Chapter 13. Also of note here are references to justifying legislation by reference to principles and ideals that are acceptable to citizens' "common human reason" (Rawls (1993), p. 137, p. 143; Quong (2011), p. 131). What "reason" itself amounts to is often left unclarified, but one way of model of how reason operates that is popular among public reason liberals—and that will serve our purposes perfectly well—is that of a reflective equilibrium. Roughly, this method consists of working back and forth between considered convictions about particular cases and general rules or principles that account for them until we can achieve an adequate level of coherence and unity between both of these two elements. See e.g. Daniels (2018), Rawls (1985), p. 29–32; (1971), p. 18; Quong (2011), p. 155. See also Gutmann and Thompson (1996), p. 5, n. 1 p. 323; Benhabib (1992), p. 30 and (2002), p. 55 and Stout (2004), p. 82.

⁶⁶ If we understand the capacity for a sense of justice as involving an operation of reason in this way, then public justification should plausibly involve an exchange of reasons in which citizens attempt to unsettle and then together reestablish their reflective equilibriums', with an eye to ascertaining how they might move to or adequately towards agreement on a contested issue. Citizens will of course never all reach perfect agreement on which pieces of legislation are maximally just, but if all goes well, such deliberation tends to move citizens towards commonly recognizing a pool of proposals that they can all recognize as meeting at least some minimum threshold of justice.

Given what we've just said, it would be natural to expect that the way we should engage illiberal citizens like Qutb is also via *deliberation*: this is what would seem to be required by way of regard for them for them as moral agents with a *reason-governed* capacity for a sense of justice. In fact, though, at the times that public reason liberals have gone so far as to discuss what engagement with the illiberal could look like, this is not what they have suggested. Though never in the context of discussing *requirements* of public justification, liberal theorists like Rawls (1993) and Weithman (2010) have suggested that the best we as liberal citizens can do to engage illiberal persons (if we want to do so) is to just exhibit to them the great goods that are to be had from cooperating with others in a political community committed to liberal norms. We are then supposed to hope that when they become experientially acquainted with those goods, they choose to simply disregard any old conflicting illiberal commitments.⁶⁷

⁶⁷ See e.g. Weithman (2010), p. 254, p. 318 (Weithman also expressed a similar view to me in conversation). Rawls (1993) writes that:

Many if not most citizens come to affirm the principles of justice incorporated into their constitution and political practice without seeing any particular connection, one way or the other, between those principles and their other views. It is possible for citizens first to appreciate the good those principles accomplish... and then... should an incompatibility later be recognized between the principles of justice and their wider doctrines, then they might very well adjust or revise these doctrines rather than reject those principles. (p. 160)

and

Much [depends] on the fact that most people's political conception are normally only partially comprehensive... [That way] when conflicts do arise, the political conception has a better chance of sustaining itself and shaping those views to fit within its limits. (p. 209)

Here, Rawls seems to want to win people over to liberalism by *distracting* them from bringing their own experiences, beliefs, arguments, and intuitions to bear on that experience: he suggests that they don't consider the connections "one way or the other" between experienced liberal goods and their own prior views, and that "much depends" on the liberal conception of justice taking advantage of the incompletenesses in others' theorizing so as to "shape" those views in favor of liberalism—rather than being made to interface with other such views on a dialectically responsive way. This kind of one-sided use of moral experience seems to be aimed, in short, to distract someone like Qutb from fully bringing his *reason* to bear on the liberal moral experiences he undergoes. But that fails to show regard for Qutb as a reasoning moral agent with a reason-governed capacity for a sense of justice.

This kind of one-sided “engagement,” however, does not seem to show respect for citizens like Qutb as serious moral reasoners, whose capacity for a sense of justice does not arrive on our doorstep as a blank slate. I will say even more about this below when I discuss differences between how we might approach a fair-minded illiberal citizen like Qutb and other, less fair-minded illiberal citizens, but suffice to say for now that—like any ordinary liberal citizen—Qutb too plausibly has a reflective equilibrium that involves balanced conscientious considered judgments and general principles. Comprehensive respect for him as a moral reasoner thus plausibly requires a more comprehensive kind of engagement with those moral reasons that he brings to the table, than just an experiential exposure to the goods of liberal cooperation.⁶⁸ In so far as we wish to persuade someone like Qutb of the merits of liberal political values, then, we should plausibly go beyond just offering him that kind of experiential “evidence” of the merits of liberal values;⁶⁹ we should plausibly also offer him a variety of other Deep Reasons—that don’t already presuppose agreement

⁶⁸ Indeed, Qutb may have something to teach us!

⁶⁹ Discussions of the moral experience approach are sometimes ambiguous between one of two tasks the approach could be meant to accomplish, and only one of these tasks is relevant to the case of Qutb. One of the tasks that this kind of approach is sometimes suggested to accomplish, is that of transforming people who are “unjust” into committed members of liberal society (I take it that this is what is at issue in Weithman (2010), p. 129, p. 216, Chapter VI). This is an important task, and (as I will say more about later) it points the way to how we might better show regard for the capacity of a sense of justice of those illiberal persons who are not fair minded. But it is important to distinguish Qutb from such persons. An alternative kind of use of moral experience that might well be an important part of public justification to Qutb is as a kind of evidence that is meant to convince rather than convert. Insofar as we offer illiberal persons opportunities to partake in moral experiences of the this sort we will go beyond just offering them qualified liberal reasons; in one sense, then, we will already be engaged in an enterprise of offering them “Deep” Reasons (that are not part of the set of qualified liberal reasons, and don’t already presuppose agreement on qualified liberal values). That said, it’s not plausible to think that these kinds of (experiential) Deep Reasons will work entirely on their own to enable illiberal persons’ to understand the justificatory merits of liberal values and qualified liberal reasons. To think that they would succeed on their own is, effectively, to think that the precedence of liberal values is so obvious to anyone who bothers to honestly consider what liberal social life is like, as to shut down any further doubts about those liberal values a person could rationally find herself with. But it’s unreasonable to think this. Citizens like Qutb, after all, may be conscientious and careful about how they form their reasoned

on liberal values, and so can speak to him in the dialectical moral position he starts from—so as to engage with his concerns and suggest in more complete, reasonable way why he should find liberal values appealing. In other words, in short, we should engage Qutb in Deep Deliberation, so as to earn his cooperation with liberally-justified legislation, not by the threat of political power or via reason-subverting maneuvers of conversion, but the by the force of the better argument.⁷⁰

IV. Defenses of Shallow Public Justification

If, as I have argued, Deep Deliberation is required to show such regard for illiberal persons like Qutb as free and equal citizens with a capacity for a sense of justice, then only legislation that is passed on the back of a practice of public justification that involves such Deep Deliberation will enjoy full and robust liberal legitimacy. As we said before, though, public reason liberals have typically denied this, and insisted that shallow public justification is fully sufficient to support liberal legitimacy. There have not (to my knowledge) been sustained attempts to defend shallow public justification from the kind of arguments I have raised, but in this section I'll discuss three general kinds of responses that have been put forward in reply to criticisms of how shallow public justification fails to address illiberal persons' dissent liberal law.

objections, and if they are, then those objections values aren't all apt to all fall away in the face of new experiential evidence.

⁷⁰ Rawls himself says that "justice as fairness is not reasonable in the first place unless it in a suitable way can win its support by addressing each citizen's reason, as explained within its own framework" (1993, p. 143). It's not clear to me what positive work is done by the last clause, but this gets at the spirit of my concern.

A. Internalism

We already saw that *internalist* liberals respond to the idea that liberals should pursue the “externalist” task of justifying liberal principles to illiberal citizens by arguing that liberals may simply stipulate their liberal principles as starting points and then focus on building a coherent program of political justification from those starting points. Deep Deliberation is a kind of externalist task, and so if internalists are right, then liberals have no duty, *qua* liberals, to engage in Deep Deliberation. But as we’ve now seen in great detail, even if one is an internalist, and so starts political theorizing by simply stipulating core liberal principles as premises, *those liberal principles themselves* seem to compel one to go beyond shallow public justification and engage in the (“externalist”) task of Deep Deliberation.

B. A Failure to Distinguish between the Demands of Abstract and Public Justification

A second and more general kind of explanation for why some public reason liberals have resisted the idea that we should go beyond shallow public justification to better accommodate the illiberal is that they have been primarily focused on working out the conditions of the proper *abstract* justification of legislation, and only addressed the matter of present public justification as somewhat of an afterthought. It may sometimes simply have been taken for granted that the contents of public justification will be filled out by whatever reasons are needed to abstractly justify legislation. But there is also an argument, hinted at briefly by (for instance) Quong (2011),⁷¹ for why one might think that this *should* be so.

⁷¹ Quong writes, “Political liberalism does not claim that some actual citizens, ‘the reasonable’, are owed justifications for the way political power is exercised, while other citizens, ‘the unreasonable’ can be coerced without justification. *All* persons are offered a justification for the exercise of political power that they could

Public reason liberals typically understand the abstract justification of legislation as itself a matter of a kind of acceptability of that legislation to citizens: what makes legislation abstractly justified (on their view) is that it is supported by reasons that free and equal citizens could, in their capacity as free and equal, reasonably accept as sufficient grounds for imposing the legislation.⁷² Of course, to see what reasons citizens can, in their capacity as free and equal, reasonably accept in this way, it won't do to simply take a poll to see what actual citizens do accept. So, public reason liberals typically demarcate a group of idealized parties whose acceptances are supposed to stand properly in accord with what their own nature as free and equal demands. What range of parties are qualified in this way is understood in different ways by different theorists—some hold that qualified parties hold to a specified range of “reasonable” points of view (Rawls (1993), Quong (2011)), while Gaus (2010) understands the relevant parties in a more capacious way, as “Members of the Public.” In all cases, however, it is stipulated that qualified parties will be liberal in that they will acknowledge that persons are free and equal citizens and be committed to associated liberal political values. The reasons that are endorsed by these (let's call them) *qualified liberal acceptors* as sufficient grounds for imposing legislation are meant to represent what persons more generally could reasonably accept, considered in their

endorse *in their capacity* as free and equal citizens committed to the idea of society as a fair system of social cooperation. In that sense each actual citizens is always provided with a justification for the exercise of political power. Of course some people may not accept the basic liberal values of freedom, equality, and fairness... But the fact that unreasonable people reject those political values does not entail that a liberal state treats such people disrespectfully when it justifies its actions by appealing values of freedom, equality or fairness [*sic*]. By relying only on public justifications based on core political values the liberal state treats each citizen in the same way—it treats all citizens as free and equal” (2011, p. 133).

⁷² Abstract justification is not, thus, a matter of (for instance) congruence with some independent given moral order of reasons. Thus, on public reason liberals view, legislation that denies citizens voting rights on the basis of their religious affiliation fails to be abstractly justified, not because there are some independent given moral facts about what freedom and equality demand, but rather because there are not reasons for such a

capacity as free and equal citizens. It is these reasons (thus understood), that then constitute the qualified liberal reasons that serve in the abstract justification of legislation.

If (as public reason liberals claim) qualified liberal reasons just are the reasons that persons, considered purely in their capacity as free and equal citizens, would accept as sufficient grounds for imposing legislation, then it starts to look puzzling how any person could, *considered in his capacity as a free and equal citizen*, be owed anything more than the qualified liberal reasons that serve in abstract justification. *In his capacity as a free and equal citizen*, after all (the thought is), Qutb could and would accept qualified liberal reasons as fully sufficient grounds for the legislation that is imposed on him. So what else could he be *owed*, *qua* free and equal? Since (it seems) *qua* free and equal, he should be satisfied with qualified liberal reasons, it appears that no further, other kinds of reasons can really be *owed* to him, thus properly conceived as free and equal; in particular (it appears) he cannot be owed Deep Reasons regarding why he should accept qualified liberal reasons in the first place. My status as a basketball player only gives me a claim to be offered the kinds of justifications that someone who accepts the rules of basketball would accept as sufficient grounds for the calls made against her, and not also a claim to be given justifications for why I should sign on for those rules in the first place. So also, Qutb's status as a free and equal citizen of a liberal democracy only gives him a claim to be offered the kinds of justifications for coercion that someone who accepts the "rules" of liberal democracy would accept as sufficient grounds for coercion (that is, qualified liberal reasons), and not also a claim to be given justifications for why he should sign on for liberal "rules" in the first place.

denial that citizens could, in their capacity as free and equal, all reasonably accept. See e.g. Rawls (1993), xxvi–xxvii.

To begin to see the trouble with this kind of argument, consider first the critical disanalogy between the basketball player and someone like Qutb. The basketball player is never *coerced* to play basketball; she opts in (or not), depending on whether she judges there to be sufficient reasons to participate in a game with the rules basketball has. Someone like Qutb, on the other hand, is *coerced* to play our liberal “game”—he does not get a chance to opt in (or opt out) depending on whether he judges there to be sufficient reasons to participate in a political activity run according to liberal rules. (Indeed, liberal political theory is partially built on the *assumption* that he does not, since it sees itself as providing an account of how citizens who enter a political system at birth, and exit only at death, can be legitimately coerced.⁷³) Given this disparity between the position of Qutb and the position of the basketball player, it is intuitively plausible that we have different obligations towards Qutb than basketball players (or a basketball referee) would have towards other participants in the game of basketball. Since basketball players opt into the game of basketball, it is reasonable for players to expect each other to be satisfied with justifications that make reference just to the rules internal to the game of basketball; players can reasonably be left to themselves to find their own reasons for participating in the first place. On the other hand, since we *force* Qutb to “play” our liberal political “game,” we cannot reasonably expect him to be satisfied just with the kinds of qualified liberal reasons that are internal to that “game.” Since he never gets the chance to opt in on his own accord, he plausibly has some claim on us to be offered reasons for *why* he should endorse the “rules” of liberal democracy with which he is coerced to comply.

⁷³ Rawls (1993), p. 445.

Similar reasoning can also be used to suggest that we owe Qutb reasons that go beyond the qualified liberal reasons that qualified liberal acceptors would accept as sufficient grounds for legislative coercion. When public reason liberals point out that qualified liberal acceptors accept qualified liberal reasons as sufficient grounds for legislation, they assume that there exists in the background some account of how these acceptors have reasonably been led to accept liberal democratic “rules” in the first place. (Thus, for instance, Rawls and Quong assume that the “reasonable” parties who figure as qualified liberal acceptors with diverse religious and moral doctrines have found resources within their doctrines to support their endorsement of liberal “rules”; this is how an overlapping consensus on liberal political values is understood to be sustained.⁷⁴) Qualified liberal acceptors endorse qualified liberal reasons as sufficient grounds for legislation, then, only because they have already been assumed to “opt in” to liberal political values—like the basketball players “opt in” to the rules of basketball. But Qutb has not opted in to these values, and so it is not reasonable for us to expect him to be satisfied with the same qualified liberal reasons that such qualified liberal acceptors are satisfied by, as sufficient grounds for legislation. Just as the kinds of reason we owe other players in such a game of basketball plausibly differs from the kinds of reasons we owe Qutb, so also the kinds of reasons that qualified liberal acceptors owe each other plausibly differ from the kinds of reasons we owe Qutb. We owe him reasons for *why* he should endorse the “rules” of our liberal political game, even though qualified liberal acceptors do not need to offer each other such reasons, to persuade each other of the merits of those rules.

⁷⁴ Rawls (1993), Lecture IV; Quong (2011), Chapter 6.

This argument gains particular force *because* the political “game” at issue is a liberal one.⁷⁵ In political games not founded on liberal values, there need be no assumption that participants have a claim to political autonomy that must be honored. We would judge such a game unacceptably authoritarian, but it at least would be internally coherent, if it did not acknowledge any need for its participants to be given Deep Reasons for why they should accept the political values the regime coerced them on the basis of. The same cannot be said, however, for our own liberal political game: our core political values require us to regard citizens as having a capacity for a sense of justice, and so also as having some presumptive claim to be able to actually realize political autonomy. Unlike more authoritarian regimes, then, a liberal regime is at risk of internal incoherence if it disavows the need to offer Deep Reasons.⁷⁶

⁷⁵ A further reason to doubt that shallow public justification can underwrite liberal legitimacy, that is related to this point, is that such a practice of public justification appears to undercut a dialectical advantage over the political regimes of (say) certain kinds of religious authorities, that liberals typically depend on to establish the political superiority of liberal regimes. Typically, liberals hold that even (say) a religious prophet who relies on substantively sound reasons to justify the legislation she proposes to impose on others, rules illegitimately if those reasons are esoteric reasons whose justificatory force is only revealed to qualified religious citizens specially initiated into her “enlightened” political perspective. If we say that substantively liberal reasons serve to legitimate coercion just on account of their being reasons that a specially qualified set of (hypothetical, idealized, or constricted) liberal citizens can accept, though, then our practice of justification seems in one important and problematic aspect indistinguishable from that of the religious prophet. Both of us, after all, say that we may properly coerce citizens on the basis of a set of reasons in part by virtue of the fact that those reasons are reasons that the right, specially qualified and “enlightened” citizens can accept—and regardless of whether the “benighted” subjects actually coerced in the name of those reasons, are privy to their justificatory force. In so far as liberal legitimacy requires us to improve on the religious prophet’s use of esoteric reasons to justify coercion, then, it might also be thought to require us to improve on the kind of use of qualified liberal reasons that public liberals sanction.

⁷⁶ Rawls denies that we should we make use of comprehensive doctrine-based, Deep Reasons in political justification because he is concerned to ensure that citizens with diverse, reasonable comprehensive doctrines cannot mutually endorse the justifications for their coercion. There are two ways in which comprehensive doctrine-based, Deep Reasons could be used that would threaten our ability to procure the relevant needed kind of mutual endorsement, but the kind of use of Deep Reasons I proposed is distinct from both and does not present a similar threat. First, citizens with diverse reasonable comprehensive doctrines would not be able to suitably mutually endorse the justifications for their coercion if the justifications that were offered directly in support of legislation were drawn straight from some controversial comprehensive doctrine. But I have not

C. An (Unduly) Unflattering Image of the Illiberal

A third reason why public reason liberals have not been particularly concerned about the consequences of shallow public justification for illiberal citizens' loss of autonomy is that

argued that Deep Reasons from citizens' comprehensive doctrines, rather than shallow reasons based on liberal values, should be used directly as substitutes for liberal justifications for legislation (see Section One). This means that, at one level, political justification will proceed in much the way it would under a more shallow Rawlsian justificatory program. My proposal merely adds an additional dimension of political justification that operates on a deeper plane. A second kind of scenario where citizens with diverse, reasonable comprehensive doctrines might not be able to mutually endorse the justifications for their coercion is one where the liberal political conception of justice is presented as abstractly justified by, and so dependent on, a single line of Deep Reasons all drawn from one controversial comprehensive doctrine. When Rawls makes the case for his shallow political liberalism, this is the primary kind of problem he is apparently concerned to avoid; he detaches support for his conception of justice as fairness from its previous Kantian roots because he realizes that the burdens of judgment make it unreasonable to expect fair-minded citizens generally to be able to form a consensus around the relevant elements of a Kantian comprehensive doctrine. We can grant that Rawls is right not to present the liberal political conception of justice as abstractly justified by the reasons of a single comprehensive doctrine. But then two moves are still possible for us. One move is to say that abstract justification can remain shallow, even while public justification goes "deep." (I begin to imply this kind of move in this subsection.) Another move is to point out that the fact that a political conception of justice should not depend for abstract support on some single comprehensive doctrine does *not* entail that it cannot depend for abstract support on *any* kind of comprehensive reasons. The alternative to one, after all, is not only none, but also *many*; we could present the political conception of justice as one that may gain deeper abstract support from a variety of comprehensive doctrines. We can preserve the benefits of shallow abstract justification while also making use of Deep Reasons in abstract justification, then, if we adopt a kind of "octopus" structured system of abstract political justification that allows in many legs of deep defense, to support some single set of central principles.

Two final notes on this latter alternative are worth making. First, even if further investigation leads us to discover that there are problems with adopting an octopus style structure of abstract political justification, those problems should not automatically lead us to think that returning to a shallow system of either *abstract* or *public* justification is the right answer. With respect to the latter, we've seen that such a system of public justification does not honor citizens' freedom and equality-based claim to realize political autonomy. With respect to the former: Rawls' advocacy for a shallow conception of justice presupposes that there will be an overlapping consensus, in which citizens with a variety of reasonable comprehensive doctrines will find, within their doctrines, Deep Reasons to support that conception. So Rawls already depends on there being a kind of octopus structured system of justificatory support for his shallow conception of justice. Whatever problems there may be with such a system, then, they are not obviously problems that are uniquely introduced just by my view. A second final note is a word of caution. We don't yet know precisely what the discursive content of Deep Reasons will be. What reasons will be appropriately rationally persuasive to Qutb, and to others like him, is still an open question, and it can only be answered by actually going through the dialogical process of Deep Deliberation. Indeed, we don't even know for certain that there will be any such Deep Reasons that are appropriately rationally persuasive to people like Qutb. As a result, we may not yet know that liberal democracy is (even) abstractly justified. That is a result I'm willing to live with. If it turns out that there is no set of Deep Reasons that can be appropriately persuasive to people like Qutb, we may need

they have focused their attention on the sub-group of illiberal persons whose loss of autonomy is intuitively least problematic.⁷⁷ When liberal theorists confront the fact that illiberal persons cannot (given their core political values), accept qualified liberal reasons as justifications for coercion, it is suggested that we need not be concerned about their justificatory alienation from the law because they (or at least the ones who are allowed to serve as representative illustrations of them)⁷⁸ are fanatical,⁷⁹ fools,⁸⁰ childish,⁸¹ hostile to human dignity,⁸² “immoral, self-obsessed, or otherwise troublesome”⁸³—and most generally and tellingly for our purposes—political analogues of psychopaths,⁸⁴ persons who simply don’t care about cooperating on fair and mutually endorsable terms with others.⁸⁵

to rethink the coherence of liberal democratic system that is built on the idea that persons are free and equal citizens with a capacity for a sense of justice. (I say more about this in the conclusion.)

⁷⁷ For more of this kind of concern, see Enoch (2015) and Friedman (2000).

⁷⁸ Nazi’s show up with surprising frequency as representatives of the class of illiberal persons who need not be publicly justified to. See Quong (2011), p. 8, Gaus (2010), p. 281, and Dreben (2003), p. 329. Dreben’s text is particularly remarkable: “What Rawls is saying is that there is in a constitutional liberal democracy a tradition of thought which it is our job to explore and see whether it can be made coherent and consistent... We are not arguing *for* such a society. We take for granted that today only a fool would not want to live in such a society... If one cannot see the benefits of living in a liberal constitutional democracy... then I do not know how to convince him... Sometimes I am asked, when I go around speaking for Rawls, What do you say to an Adolf Hitler? The answer is nothing. You shoot him. You do not try to reason with him. Reason has no bearing on that question. So I do not want to discuss it. But what I am perfectly prepared to discuss is whether the idea of liberalism as an ideal... is contradictory” (2003, p. 329).

⁷⁹ Larmore (1987), p. 60.

⁸⁰ Dreben (2003), p. 329.

⁸¹ Macedo (2000), p. 35.

⁸² Gutmann and Thompson (1996), p. 65.

⁸³ Quong (2011), p. 37.

⁸⁴ *Ibid.*, p. 314, Gaus (2010), p. 281–282. Quong and Gaus suggest that illiberal persons must be treated as the political equivalent of psychopaths.

⁸⁵ In discussions where he touches on the exclusion of the unreasonable, Rawls often characterizes them as rejecting the principle of reciprocity, implying that it is the reason for their exclusion. See e.g. Rawls (1993), p. 483. Quong (2011) also often characterizes the unreasonable as rejecting the ideal of fair cooperation, see e.g. p. 290, p. 294, p. 313.

It is of course not intuitively very concerning if public justification leaves fanatics and political psychopaths without an appreciation of the justificatory merits of the law. In so far as public reason liberals have focused on such cases as exemplars of cases of illiberality, as a result, it's not surprising that they have not been more disturbed by the failure of shallow public justification to facilitate illiberal citizens' autonomy.⁸⁶ That said, we've seen reason to think that these characterizations of illiberal citizens are too one-sided to serve as any general guide for intuitions about how we should treat illiberal citizens.⁸⁷ The kinds of figures who are used to represent illiberal citizens above importantly lack the kind of fair-mindedness that we saw Qutb plausibly possesses (in Section One); accusations of a general disregard for the ideal of cooperating with others on fair terms—not to mention, of childishness, fanaticism, or self-obsession—are palpably unfitting with respect to Qutb. Further—and importantly—we have no reason to think that Qutb's concern for fair cooperation makes him an utter outlier among illiberal citizens. Since liberals fail to supplement the above epithets they apply to the illiberal with any real examination of cases of persons with illiberal views, they don't give us any empirical reason to think that Qutb is an exceptional case. It's also hard to see any formal reason to think that he must be exceptional. Qutb's case goes to show that which we must surely anyways have been compelled to acknowledge: namely, that it's possible for people to be genuinely concerned to cooperate with others on fair terms, even while being seriously mistaken about what

⁸⁶ Importantly, when liberal theorists offer the above characterizations of the illiberal, they are often not careful to distinguish between questions of abstract justification and public justification. Part of the point of the above characterizations, then, is often to convince the reader that it is not problematic to exclude the illiberal citizens who reject qualified liberal reasons, from the constituency of qualified acceptors. Thanks to David Estlund for pointing this out.

⁸⁷ It would be a political prejudice to assume that all illiberal persons who resist qualified liberal reasons must, by virtue of that, also reject the ideal of cooperating with others on fair terms wholesale.

such terms actually look like. (Indeed, but for our own more liberal upbringing, we may very well have found ourselves in just such a position.)

So suppose we grant that there is a morally important distinction to make, within the category of the illiberal, between those who fail to be concerned about the ideal of cooperating with others on fair terms—call them *unfair-minded* illiberals—and those *fair-minded* illiberal citizens who are so concerned, but understand what fair cooperation requires in a non-liberal way. How does this distinction make a difference to how we should approach public justification to these two constituencies? Given what we’ve said so far, the answer is not obvious. So far, after all, what we have said is that it is illiberal persons’ status as free and equal that grounds a presumptive claim to autonomy. But unfair-minded illiberal citizens are presumably free and equal citizens as well, and so they would appear to have just as much of a presumptive claim to realize political autonomy as fair-minded illiberal citizens.⁸⁸

I happily grant that unfair-minded illiberal citizens may have a *presumptive* claim to realize political autonomy—but what I’ll now try to show is that that presumptive claim may be defeated, or (more plausibly in my view) at least *altered*, when unfair-minded illiberal citizens are in question. Given that it can be so defeated (or altered), we can preserve the idea that there is a duty for Deep Deliberation with citizens like Qutb, without also becoming committed to the claim that there is an identical duty with respect to unfair-minded illiberal citizens. Public reason liberals have not been clear about how unfair-mindedness might interact with illiberal persons’ claim to autonomy, but below I

⁸⁸ Thanks to David Estlund for impressing this point on me.

canvass three ways in which such unfair-mindedness might defeat (or transform) that claim.

(i) Does Unfair-Mindedness Undermine the Presumptive Claim to Autonomy?

There is one proposal about how unfair-minded illiberal citizens' presumptive claim to autonomy may be defeated that is badly unsuccessful, but that it is important to explicitly reject. This is the proposal—suggested by the above idea that illiberal persons are the political analogous of psychopaths or fanatics—that illiberal persons simply lack a capacity for a sense of justice. If this is right, illiberal persons would not really be free and equal citizens after all, and so any presumptive claim to autonomy would be defeated by being undermined. It would clearly be ridiculous to suggest that all illiberal persons not already convinced of the superiority of liberal democracy can be treated as lacking a capacity for a sense of justice. The very thought manifests an insulting imperialistic attitude towards those who are politically unlike us, that we should be happy to leave behind to the colonialists of previous centuries. It is also implausible, though, to suggest that unfair-minded illiberal citizens lack a capacity for a sense of justice altogether. Someone who has grown up in the mob may not now be convinced of the superiority of fair cooperation, but that does not mean she does not have the *capacity* to come to see the virtues of such fair cooperation, and to be motivated (in the right conditions) to abide by them. Here and more generally, then, illiberal persons unfair-mindedness cannot be assumed to entail a wholesale lack of a capacity for a sense of justice.

(ii) Does Unfair-Mindedness Override the Presumptive Claim to Autonomy?

A second more plausible way to argue that unfair-minded illiberal citizens' presumptive claim to autonomy is defeated, is to argue that it is *overridden* by their unfair-mindedness. In

particular, one might argue as follows. Though the presumptive claim to autonomy is attached to the *mere possession* of a capacity for a sense of justice, it is thus attached only in light of the reasonable expectation that individuals who have the capacity will *use it*, at least in the minimal respect of taking a sincere, general interest in the ideal of fair cooperation. The *effective* claim to autonomy—the claim that actually generates obligations for others—only comes into play on the condition that this reasonable expectation is then fulfilled. It is reasonable to make such concern for fair cooperation a condition of possessing an effective claim to autonomy (one might continue), because it is reasonable to expect citizens who are granted such an effective claim to autonomy, to display a basic degree of reciprocity in this respect, in their cooperative efforts with others: it is reasonable to expect someone who is granted an effective claim to autonomy, to display a sincere, good faith commitment to finding terms of cooperation that also enable *others* to mutually realize political autonomy for themselves. But (one might insist) one cannot sincerely and in good faith expect others to mutually endorse, and so realize autonomy in the context of, terms of cooperation that are admitted off the bat to be unfair to them. So, if you have a sincere commitment to enabling others' mutual realization of autonomy, you must also display at least an in-principle concern to cooperate with them on fair terms. If this reasoning is right, then in so far as an illiberal person fails to take even an in-principle interest in cooperating with others on fair terms, her presumptive claim to autonomy would be defeated by being overridden. There would thus be no grounds for a duty of Deep Deliberation.

Importantly, the same kind of reasoning cannot be used to insist that it should be a condition of having an effective claim to (a feasible opportunity to) autonomy, that one

first express concern for a specifically liberal way of understanding fair cooperation. For it is not reasonable to expect citizens to accept liberal political values, right off the bat, in the way that it might be reasonable to expect citizens to take a general interest in fair cooperation, right off the bat. Ordinary conditions of human experience *may* perhaps afford persons, generally, with a basic appreciation of the importance of reciprocity and its relation to fair cooperation, and *if* it does, that is part of what would make it reasonable for us to expect such appreciation from persons when they join us at the political table.⁸⁹ But the same can certainly not be said with respect to liberal principles of justice. The burdens of judgment—or everyday obstacles to reasoning—that liberals acknowledge are responsible for ordinary, run-of-the-mill pluralism about religion and morality among fair-minded people, also plausibly affects citizens like Qutb’s development of a conception of fairness.⁹⁰ This makes it unfeasible for them to reason their own way to a specific set of (liberal) political values—even if they are conscientious and otherwise fair-minded. (The

⁸⁹ I’ll not pursue this matter further here, but only register an important point, which is that it is not obvious that it is always reasonable to expect citizens to come to the political table already having learned to exercise their capacity for a sense of justice in even the minimal way indicated above. You cannot reasonably be expected to take an interest in cooperating with others on fair terms if (for instance) you have not been raised in a social community that modeled and made clear the benefits of fair cooperation. A state that shows regard for unfair-minded illiberal persons’ capacity for a sense of justice would thus seem to have a responsibility to take steps to provide this needed kind of social support. Further, if or when that support is initially lacking, regard for relevant unfair-minded persons’ capacity for a sense of justice would seem to compel us to institute practices of public justification that help re-establish it. In particular—and as will come up again in the next subsection—we may be obliged to add to our practices of shallow liberal justification, measures that provide unfair-minded citizens with something akin to transformative moral therapy, which aims revamp their sense of intrinsic rewards of living with others on fair terms.

⁹⁰ Qutb’s political views plausibly largely stem, for him, as all of our views do, for us, both on his background in a particular epistemic environment—one that no doubt presented him with a range of evidential considerations that we are not familiar with—and on a kind of moral reasoning that is, though careful and morally serious, still vulnerable to the normal gamut of obstacles that attend all human moral reflection. It’s no doubt much harder for us to appreciate just how these burdens operate in the mostly unfamiliar domain of non-liberal reasoning, than in our own intramural debates, but that is not itself any reason to think that they don’t. Rawls (1993), p. 52ff. Gaus (2010) and Quong (2011) refer to the relevant phenomenon as the “fact of reasonable pluralism.”

only thing that may separate how those burdens operate in the realm of religion and comprehensive morality, and in the realm of political opinion, is that—we hope—they may eventually feasibly be able to be overcome by reasonable deliberation in the latter realm of political opinion.)⁹¹

(iii) Does Unfair-Mindedness Alter the Presumptive Claim to Autonomy?

The third (and by my view most plausible) way to argue for a distinction in our justificatory obligations to the unfair-minded and the fair-minded illiberal is to argue that, though the unfair-minded retain some presumptive claim to autonomy, the way in which that claim should be honored differs significantly from the way it should be honored with respect to the fair-minded. In particular (one might argue), honoring an unfair-minded person's claim to autonomy does not require us to engage her in reasoned *deliberation* about contested issues in the way that is appropriate to engage with someone like Qutb.

One might argue for this idea as follows. If we were to try to engage unfair-minded illiberal citizens in deliberation about why they should accept liberal values, the kinds of considerations they would be apt to advance as objections to our reasons would be ones that we could with excellent warrant deny bear any weight as moral reasons. An unfair-minded person might object to liberal legislation simply because it does not prioritize her own well-being, for instance—but we are well-warranted in denying that that consideration bears any weight as a moral reason in a discussion about what terms of

⁹¹ Here I touch upon the often discussed Asymmetry Problem for public reason liberalism (see e.g. Quong (2011), p. 192–220 and Leland and van Wietmarschen (2012), p. 738–741). As I say more about at the conclusion, if difference in political opinion *can't* be overcome this way, I expect public reason liberals will need to modify either their commitment to the capacity for a sense of justice, or to the idea that liberal political values can be appropriately imposed on those all those who now reject them.

cooperation would be fair for us to jointly abide by.⁹² In so far as this consideration doesn't bear weight as a moral reason (one might claim), we do not do violence to the unfair-minded illiberal citizen as a citizen with a reason-governed capacity for a sense of justice if we try to change her mind in some way other than by directly engaging that consideration with counterposing arguments. To put the same idea in a different way, we might suggest that in so far as the unfair-minded illiberal citizen has not carried out any moral reasoning relevant to matters of justice, there is no reason-governed operation of her capacity for a sense of justice that has been operationalized, and so that needs to be respected by being deliberately engaged with. As a result, it *may* after all be permissible to one-sidedly expose her to liberal moral experiences that we hope simply upends her previous ways of thinking, serving as moral therapy (or perhaps sparks for a political "conversion") to "start up" that operation of the capacity for a sense of justice for the first time—without destroying any prior fruits of its exercise in the process.⁹³

For reasons we have already started to discuss in the last section, the same kind of argument would not succeed in making it permissible not to deliberately engage with someone like Qutb. For Qutb's objections to liberalism—even if ultimately mistaken—appeal to considerations that we can recognize as morally weighty and relevant to discussions about fairness: worries about materialism, about giving citizens a fair chance to flourish, and even about liberalism's "fit" with his religious commitments are worries

⁹² Similar thinking might also be applied to those who claim to be fair-minded but whose objections to liberal legislation definitively lack moral weight. The objection of a neo-Nazi that makes reference to the claim that citizens of certain religious and racial groups lack moral status as persons, for instance, would appropriately be counted failing to carry any moral weight. To the extent that such a neo-Nazi employed non-morally-weight-bearing reasons of this sort, she would be "fair-minded" in name only; she would be better classified, and treated, as unfair-minded. See however Footnote 12.

⁹³ This was discussed in more detail in Section Three, Part C.

that we can recognize as having a kind of politically reasonable, moral weight to them. Indeed—and importantly—it’s not just that we *can* recognize these kinds of considerations as bearing moral weight; at least public reason liberals are also antecedently *committed* to treating them as such. This is because in “good” cases (where citizens turn out to be politically liberal), public reason liberals are committed to regarding the various considerations embedded in persons’ comprehensive doctrines as morally weighty enough to provide reasonable support for citizens’ commitment to liberalism.⁹⁴ If a liberal citizen adverts to considerations regarding human flourishing and congruence with Islam as reasons that motivate her endorsement of liberal political values, public reason liberals typically will regard these considerations as bearing enough moral weight to support her *reasonable*, autonomous endorsement of liberal norms of fairness.⁹⁵ But any consideration that is granted to bear moral weight in *support* of a citizen’s reasonable commitment to liberal justice, should also be taken seriously as a morally load-bearing reason when a citizen who is not yet liberal, like Qutb, takes it to provide reasons *against* endorsing liberalism. Suppose then that we can—and that (at least) public reason liberals *should*—recognize the kinds of considerations that a fair-minded illiberal citizen like Qutb advances as objections to liberalism, as bearing moral weight. If that is the case, then we cannot simply set aside his objections and depend on moral therapy, or the effects of some other kind of liberal conversion experience, to bamboozle him into changing his mind. To do that would be to subvert the operation of a kind of reasoning that we recognize as bearing moral weight and (more specifically) that we recognize as relevant to supporting

⁹⁴ See Rawls (1993), Lecture IV; Quong (2011), Chapter 6; Gaus (2011), Section 3.2.

the operation of his capacity for a sense of justice. It would not, as in the case of the unfair-minded, be simply making that operation possible for the first time.

V. Closing Concerns

A policy of public justification that involves a duty for Deep Deliberation has a critical advantage over a policy of shallow public justification: the former makes it a requirement that we show illiberal persons like Qutb the respect they are due as free and equal citizens, while the latter fails to do so. In particular, by requiring us to engage in Deep Deliberation with fair-minded illiberal citizens, the former ensures that we manifest regard for (first) the fundamental importance of those citizens' exercise of a capacity for a sense of justice⁹⁶ and (second) the operation of their powers of moral reasoning that support that capacity. That all granted, however, the idea that we have a duty for Deep Deliberation is itself bound to raise some questions. Not all of those questions can be addressed here, but I briefly respond to three kinds of concerns that may be raised for it, below.

A. Too Demanding as a Duty for Ordinary Citizens?

It might be worried that a duty to engage in Deep Deliberation is beyond what we can reasonably require of ordinary citizens. Surely, one might say, ordinary citizens cannot be required to take on the giant task of reasoning with others about the philosophical foundations for liberalism. I agree, but have two responses.

The first thing to say is that though the duty for Deep Deliberation is naturally understood in terms of a duty to offer and discuss Deep Reasons to be liberal, that doesn't

⁹⁵ They do not, thus, want to rely on the kinds of diversely-motivated endorsements of liberalism that liberal citizens in a pluralistic society will have, as on successful manipulative ploys to buy their cooperation with liberal justice. See previous footnote for references.

require offering theoretical proofs of liberalism. Rather, the moral kernel of the duty lies in the critical importance of engaging with the illiberal in morally reflective and deeply dialogic ways. Deep Deliberation will thus not always—or perhaps, even normally—proceed by way of fully worked out, abstract philosophical persuasion. Much of it will occur instead in more piecemeal and practically manageable ways—listening to another’s views on a specific legislative matter, querying particular intuitions, sharing moral experiences, and so on.

The second thing to note about the duty to engage in Deep Deliberation is that even if it is in some regards applicable to all of us, there are other regards in which the duty is importantly collective, rather than individualistic. Certainly, much of the theoretical heavy-lifting will properly fall to professional thinkers—to philosophers, clergy, and theologians, for instance, who have the relevant needed skills and professional time and resources at their disposal. Last, in so far as prior familiarity with a tradition makes one better able to engage with others who are also in it, there may be some division of labor: the duty to deeply deliberate with (say) Qutb may fall on different people than the duty to deliberate with those coming from other religious and philosophical backgrounds does.

B. The Short Term Failure of Deep Deliberation

The considerations adduced in this paper suggest that, in an ideal world, we would persuade illiberal citizens of liberal values (or come to some kind of an agreement with them about fundamental political values) before coercively imposing legislation on them. This, I take it, is what fully responsive regard for their capacity for a sense of justice speaks in favor of when circumstances allow. Be that as it may, however, we have already seen

⁹⁶ It ensures that we do this by requiring us to make reasonable efforts to provide them with a feasible

that recognition of the value of such pre-coercion reconciliation does not keep us from also recognizing the value of other important countervailing considerations—for instance, the timely ending of a substantive injustice that is not yet recognizable as such to the illiberal. The existence of these kinds of countervailing considerations will often mean that, in the short term, the best way to show regard for citizens as free and equal (*all things considered*) is to go ahead with the legislation despite illiberal citizens' current inability to endorse the justifications for it. When these important countervailing considerations exist, then we will plausibly have moral excuse to legitimately coerce illiberal persons despite their dissent from qualified liberal reasons. Still, though we might in these cases legitimately *background* concerns about illiberal persons' realization of autonomy, we cannot permissibly—as public reason liberals seem to—totally discount the value of their realization of autonomy. We will thus always be required to make context-appropriate efforts to pursue Deep Deliberation.⁹⁷

C. The Long Term Failure of Deep Deliberation

A third kind of concern falls in line with the previous one, but brings matters to head in a more definite way. Jonathon Quong makes clear that one reason he supports an internalist conception of political liberalism that denies any duty for Deep Deliberation with the illiberal is that he thinks that illiberal persons are “incapable of ‘seeing’ [the] moral force of [liberalism’s core values],” and that liberal political theory is not “capable” of coherently

opportunity to realize political autonomy.

⁹⁷ Things will be different depending on, for instance, the urgency of the legislation, the importance of it, and the certainty with which we hold to the relevant element of our liberal beliefs.

justifying liberal values to those who do not already accept them.⁹⁸ If Quong is right, then Deep Deliberation will of course be futile, and so the duty for Deep Deliberation senseless.

If Quong *is* right, however, then public reason liberalism is wrong—or at least, not anything like we imagined it to be. For if Quong is right, then there are some people from whom the core truths of political morality will in principle always be veiled. If they are to be part of a just society, as a result, we will always have to rule over them—as a philosopher king or authoritarian religious prophet would rule over benighted subjects. But that’s not what democracy looks like. To treat people as free and equal is to treat them as sharing in common with us a capacity to see and understand moral reasons, and so to have faith that they can be capable moral partners in the development of a just society—rather than passive subordinates who must take esoterically justified orders from the Enlightened.

⁹⁸ Quong (2011), p. 314. See also p. 140, 152–153. More fully, Quong writes that “Unreasonable persons are not simply excluded from the constituency of public justification: they are *beyond the reach* of liberal public reason. Liberalism’s core values do not ‘speak’ to them, or they are incapable of ‘seeing’ their moral force. This should not undermine our confidence in liberal justice anymore than the existence of psychopaths should undermine our confidence in moral rules” (p. 314). Quong does not give a rational for why we should think that illiberal people are incapable of seeing the moral force of liberal values, but he does explain some rationale for thinking that liberal political theory cannot coherently justify liberal values to those who do not already accept them. Quong writes: “Liberal reasoning is not possible without [liberal] values—to query their validity is to step outside liberal political theory altogether and engage in some other form of reasoning...that must necessarily seem unsound to anyone committed to liberal values” (p. 313–314). This picture that Quong paints of our dialectical situation, vis-à-vis the illiberal, however, is overly simplistic—even fundamentalist and Manichean. Perhaps not surprisingly, given the insularity of the internalist political liberalism he is in the business of defending, his claim betrays a lack—at least, of imagination—and, more realistically and importantly, of serious exposure to or engagement with the kind of thinking and reasoning that the fair-minded illiberal persons partake in. There may very well be some ways of stepping outside of liberal political morality that are simply beyond the pale of what liberal democrats can countenance, but a great many of the forms of non-liberal reasoning about political morality that we realistically need to contend with are forms of reasoning that—like those of Sayyid Qutb’s—interweave elements of liberal values and that, further, appeal to other familiar but not distinctively “liberal” values and intuitions that we can nevertheless very meaningfully engage with. (Qutb, for instance, talks a great deal about mutual responsibility and human dignity—values which lay beyond any all-encompassing “liberal” vs. “non-liberal” divide. I discuss this further in Section One.)

Conclusion

Forms of liberalism that deny that we have an obligation to engage illiberal persons in Deep Deliberation presume that liberally grounded political orders can rightfully claim authority over those to whom they nevertheless have no political obligation to defend themselves. It should come as no surprise to us, then, if such forms of liberalism are accused of a kind of domineering imperialism vis-à-vis the illiberal. Such imperialism is not only intrinsically problematic; it is also contrary to liberalism's own central values. In particular, it is antithetical to showing regard for illiberal persons' capacity for a sense of justice and, following on that, their status as free and equal citizens. To better live up to our own central values, then, we should take an approach to public justification like the one that I have proposed, that recognizes a duty to deliberatively engage fair-minded illiberal persons with Deep Reasons aimed at bridging the justificatory divide between us. Only by taking this kind of approach can we properly acknowledge the importance of treating people like Qutb—not as relata in a power-mediated *modus vivendi*—but as free and equal citizens who (first) have the capacity to autonomously govern their own political life in accordance with fairness-pertaining moral reasons, and following on that (second) are owed a chance to see the merits of the liberal reasons we coerce them in the name of.

My argument for this conclusion has proceeded as follows. First, I introduced Seyyid Qutb (1906-1966) as an example of what a fair-minded illiberal citizen might look like. Second, I gave a preliminary view of the inadequacy of shallow public justification, suggesting that its failure to facilitate someone like Qutb's realization of autonomy is problematic even in light of our own liberal democratic values. In Section Three, I then offered a more detailed account of how citizens' presumptive claim to autonomy is

grounded in their status as free and equal, and argued that this presumptive claim gives rise to a defeasible duty for us to engage illiberal citizens in Deep Deliberation. In Section Four, I addressed three objections that might be raised to my view. I gave special attention to making a distinction between fair-minded and *unfair*-minded illiberal citizens, and argued that while unfair-minded illiberal citizens' presumptive claim to autonomy may be defeated,⁹⁹ it is not similarly defeated with respect to *fair-minded* illiberal citizens. As a result, I concluded, the duty for Deep Deliberation with them remains intact. I closed in Section Five by briefly addressing concerns about whether and how Deep Deliberation can realistically be carried out. Though I did not deny that it may not in the long run do the work we want for it to do, I suggested that public reason liberals' acceptance of the idea of a capacity for a sense of justice commits them to have faith that it will.¹⁰⁰

⁹⁹ More specifically, I argued that it may be defeated *or altered*.

¹⁰⁰ Thanks to Ruth Chang, Alex Guerrero, and especially David Estlund for their extensive and helpful comments.

Chapter 2

The Phenomenal Appreciation of Reasons (Or How Not to be a Psychopath)

Introduction

Stealing is wrong. Huck knows it's wrong, and he believes that by helping Jim escape to freedom, he is stealing—and so doing something wrong. But Huck does it anyway—explicitly going against what he believes he has decisive moral reason to do, he decides to help Jim escape to the North.¹

¹ Huck's case has received a great deal of theoretical attention in the philosophical literature on moral psychology and metaethics. See e.g. Arpaly (2002), Bennett (1974), Markovits (2010), Manne (2013), Sliwa (2016), and Teichman (1975). Here is a synopsis of the case. Huck Finn has been traveling down the Mississippi river on a raft with Jim, who is running away from his "owner," Miss Watson. They have become friends, but when they come near to the place where Jim would legally be free and Jim begins to talk about his excitement about freedom, Huck begins to realize that he might be doing something wrong:

I begun to get it through my head that he was most freed—and who was to blame for it? Why, me. I couldn't get that out of my conscience, no how nor no way.... It hadn't ever come home to me, before, what this thing was that I was doing. But now it did; and it stayed with me, and scorched me more and more... Conscience says to me: 'What had poor Miss Watson done to you, that you could see her [slave] go off right under your eyes and never say one single word? What did that poor old woman do to you, that you could treat her so mean?' I got to feeling so mean and so miserable I most wished I was dead. (Twain, p. 109–110)

Huck eventually goes towards shore to turn Jim in to some men looking for run-away slaves. But, at the last second, Huck can't bring himself to go through with it. Instead of telling them about Jim, he tells them that he has someone who is sick on board. When Huck returns to the raft, he "feel[s] bad and low" because, he says, "I knowed very well I had done wrong." He then reports that:

Then I thought a minute, and says to myself, hold on—s'pose you'd a done right and give Jim up; would you feel better than what you do now? No, says I, I'd feel bad. I'd feel just the same way I do now. Well, then, says I, what's the use you learning to do right, when... the wages is just the same?... I couldn't answer that. So I reckoned I wouldn't bother no more about it, but after this always do whichever come handiest at the time. (p. 114)

If all we saw of Huck was him rehearsing his moral reasons for turning Jim in, and then opting to act against the judgment about his reasons he came to, we would rightly suspect that his decision to help Jim merely

Why do we credit Huck for his deed when it constitutes a failure to respond to the moral reasons he explicitly believes he has? Huck's choice to help Jim corresponds with what he has objective moral reason to do, but our esteem for Huck isn't warranted if his choice only happens to accord with the moral reasons that apply to him, and isn't also—at least in some capacity—made for the sake of those reasons. To earn our esteem, thus, Huck must appreciate and respond to considerations speaking in favor of helping Jim as moral reasons.² But then, if Huck is worthy of our esteem, there must be some implicit way of appreciating and responding to considerations as moral reasons that does not involve explicitly believing that those considerations are moral reasons.

happened to accord with the moral reasons that applied to him, and was not in any meaningful capacity made for the sake of those reasons. But we see something more: we see Huck have extended interactions with Jim that involve a particular kind of moral relational experience with him. Thus, for instance Huck has a kind of natural regard for Jim as a friend and peer, and there's no hint—as there would be with any more thorough racist—of condescension towards Jim or moral superiority on his own part. Further, it only occurs to Huck to think twice about what he is doing far along the way down the river, and when it does, the thought that what he's doing amounts to stealing—and so is “wrong”—comes to Huck, at least initially, only as from the “outside,” as an abstract realization of his actions' formally filling certain criteria of wrongdoing that have been laid out for him.

Given the kind of relationship Huck has with Jim and the kinds of attitudes he expresses, it seems that Huck's decision to help Jim is based not on considerations that have impressed him as neutral stimuli. Rather, Huck's decision seems to be based on a kind of visceral impression of (something like) Jim's equal moral personhood. Huck has an appreciation of his reasons, then, grounded in his experience, and Huck's failure to explicitly conceptualize relevant considerations *as moral reasons* seems to be an upshot of the fact that his experience of moral reasons has simply outstripped the concepts he has readily available to him to make sense of that experience. While Huck's appreciation of his reasons does, by virtue of his failure to conceptualize them in the right way, seem to involve an unfortunate defect in packaging, it does not plausibly involve any critical break in the basic integrity of that appreciation.

² I focus attention here on moral reasons appreciation as a condition of moral praiseworthiness, but a much more detailed argument would be required to show that moral reasons appreciation is either a necessary or sufficient condition of moral praiseworthiness. I focus attention on appreciation as a condition of praiseworthiness here only to make salient that specific kind of representation of reasons that I am concerned with. I take it that at least in paradigmatic cases in which an agent is morally praiseworthy for ϕ -ing, she appreciates her reasons in the way I discuss here. Thanks to David Enoch for pressing me on this point.

Huck's case is fictional, but there are a range of real cases of implicit moral agency³ that are like his, that make understanding his implicit responsiveness to reasons practically important. An important feature of Huck's case is that the concepts he possesses are inadequate to capture the contours of his moral experience: he has a distorted concept of *stealing* and apparently lacks the kinds of concepts of (for instance) *free and equal moral agency* that he would need to apply to Jim, to form the appropriate explicit beliefs about his moral reasons to help Jim.⁴ Whether due to hermeneutical injustices or just bad hermeneutical luck,⁵ real moral agents often find themselves in a position analogous to Huck's. Thus, for instance, a man who has been sexually harassed but lacks the concept of "sexual harassment" may respond to his moral reasons—perhaps he gets indignant at his harasser in a way that only makes sense for one to become indignant at moral

³ Here and throughout I use the term "implicit moral agent" to refer to an agent who implicitly appreciates moral reasons *as* moral reasons. For more on why I often focus on the kind of appreciation of moral reasons that is relevant to moral agency as opposed to (specifically) *creditable* moral agency, see Footnote 2.

⁴ For more on Huck's case, see Footnote 1. Concepts can fail to cut at the joints of moral experience in one of three ways. First, an agent's moral experience can be such as to suggest that the application conditions inferentialistically associated with a set of specified entailments extend in different ways than has been conventionally recognized by users of the concept. Thus, for instance, Huck's experience suggests that concepts of free and equal agency should, counter-conventionally, be extended to blacks. Second, an agent's moral experience can be such as to suggest that the entailments conventionally associated with a set of conceptually-specified application conditions are not appropriate. Thus, consider Sam:

Sam: Sam is in a same-sex relationship, living in a social context that takes all such relationships to be morally unnatural, "queer" relationships that people have moral reason to avoid. Sam has sense of the moral pull to remain faithful to her partner, but she lacks the conceptual resources to understand herself as driven by anything but sexual perversion. "Hang it!" she finally decides; "I'll be queer!"—and forthwith resolves to be steadfastly faithful to her partner.

Sam's experience suggests that the moral entailments typically associated with the application condition of *queer* are inappropriate to those application conditions, and so her attempt to apply the concept to her experience leaves her unable to make discursive moral sense of it. A third kind of case of "mismatch" between agents' moral concepts and moral experience occurs in cases where the relevant concepts are not just warped, but entirely lacking—the agent lacks any concept that picks out *either* application conditions *or* entailments relevant to her experience.

trespassers—and yet not be positioned to make reflective sense of the actions of his harasser as being the morally salient grounds for indignance that they are. Conceptually impeded implicit moral agents of this sort are not able form explicit beliefs about their moral reasons and are thus also frequently unable to articulately defend their moral choices to others. As a result, their moral agency is easily overlooked or discounted.⁶ Nevertheless, they display morally creditable agency, and so are worthy of our moral esteem. By better understanding how Huck implicitly appreciates his moral reasons, we can hope to help redeem the moral credit this whole class of implicit moral agents is due.

So, what might implicit appreciation of and responsiveness to moral reasons involve? In this paper I argue that an agent can implicitly appreciate a consideration as a moral reason to ϕ without explicitly believing it to be a moral reason to ϕ by presenting it under the light of a particular phenomenologically-mediated mode of presentation: one that presents the relevant consideration via the light of a kind of felt directive force “pointing” towards ϕ -ing—lending weight to it, or soliciting it—in a particular way. An agent can then respond to the consideration *as a moral reason* by responding to the consideration as presented under this light.

Although I will focus on moral reasons, my main aim isn’t to differentiate the appreciation of moral reasons from, say, the appreciation of prudential or aesthetic reasons. Instead, my focus is on explaining how we can appreciate moral reasons understood as one kind of practical reason. A larger claim I won’t have time to defend here is that the appreciation of practical reasons, more generally, involves a

⁵ For more on this, see Fricker (2007)’s important discussion of hermeneutical injustices. She characterizes the problem well: “None of [the concepts] fits,” but relevant “collective understandings are... powerful,” and “the experiential promise of an alternative understanding... lonely and inarticulate” (p. 164).

phenomenologically-mediated grasp of relevant considerations. I focus on the moral case here because I take moral reasons to be the most important kind of practical reason, and I think they constitute a good starting point for discussion of the more general claim about practical reasons.

My argument proceeds as follows. In the first two sections of the paper I consider four major approaches to explaining implicit appreciation of moral reasons and outline the difficulties faced by each. In the third section, I propose my own—more intuitively plausible, I argue—account of implicit appreciation of moral reasons. In the fourth section I suggest that my account of implicit moral reasons appreciation leads to an insight about moral reasons appreciation more generally, even beyond the implicit case. I end by highlighting some theoretical advantages of my approach.

I. *De Re* Accounts of Implicit Appreciation

One natural way to try to accommodate cases of implicit appreciation of moral reasons like Huck's is to lower the cognitive bar on moral reasons appreciation so as to allow mere *de re* contact with moral reasons to underwrite appreciation. Arpaly (2003), for instance, claims that an agent is creditably responsive to moral reasons if (roughly) she responds to the genuinely moral reason-giving features of the world out of an intrinsic *de re* desire that the course of action that has those features be taken.⁷ Along similar lines, Dreyfus (1991) claims that exemplar cases of moral agency are cases in which subjects comport themselves with the reasons their circumstances afford in a direct and spontaneous way,

⁶ See Fricker (2007) for an extensive study of this phenomenon.

⁷ See e.g. Arpaly (2003), p. 73–79, 84 and Smith (1994), p. 71–76.

without the baggage of intermediary *de dicto* representations about those reasons. Using a version of Arpaly's formulation of the view,⁸ we might say:

(1) DE RE: An agent implicitly appreciates a consideration *c* as a moral reason *r* to ϕ iff *c* is a moral reason to ϕ and she has an intrinsic, *de re* concern about *c* that leads her to desire to ϕ (or to desire that ϕ -ing be done).

Unfortunately, this proposal won't work. Since *de re* contact with moral reasons can be secured without the help of *any* particular mode of presentation of those reasons, a merely *de re* relation between an agent and a moral reason-constituting consideration is by nature insufficient to establish that her responsiveness to that consideration has anything in particular to do with its being *a moral reason*. The result is that a *de re* account of appreciation of moral reasons can't distinguish the perspective of an implicit moral agent on her reasons from that of, say, a lucky psychopath. Consider Juliet, and her psychopathic cousin Silas:⁹

Juliet: Juliet's family is caught in a bitter rivalry with another family, characterized by repeated cycles of violent revenge that are prodded by the premium that her society puts on machismo and social status. Juliet sees her cousin Tybalt punch Benvolio, a member of the other family on her family's "turf," right in the eye. She sees that Benvolio is badly injured and realizes he's in great pain. Uneasy and disturbed about Benvolio's being in pain, she tries to coax Tybalt to stop. "Why? What's wrong with you?" Tybalt asks, "Don't you care about our family's honor?" Juliet is not sure what to say. "No, I care," she sighs, "...I guess I am just a softie who can't bear seeing people in pain."

Like Huck and other implicit moral agents, Juliet lacks conceptual resources to make sense of the moral reasons she implicitly appreciates for Tybalt to stop. Still, she merits some

⁸ Arpaly does not formulate her view in terms of the appreciation of considerations as reasons *per se*, but rather in terms of the question of what kind of contact we need to have with the features of a situation that make ϕ -ing right, to count as being morally praiseworthy for ϕ -ing.

⁹ The conception of the psychopath I use here is the philosopher's conception. As I understand it psychopathy is no longer characterized as a medical psychological condition by the American Psychiatric Association in the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders-5).

moral credit for her response to these reasons. But consider, in contrast, her psychopathic cousin Silas:

Silas (*De Re*): Silas is Juliet's fussy psychopathic cousin. Silas sees Tybalt deal Benvolio the blow to the eye and sees he's badly injured. Silas realizes that Benvolio is in great pain, and he finds that distasteful and unseemly; he is 'weirded out' by people's experience of pain. Uneasy and disturbed about Benvolio's pain, he tries to coax Tybalt to stop. "Why? What's wrong with you?" Tybalt asks, "Don't you care about our family's honor?" "No, I care," he sighs, "I guess I just can't bear seeing people in pain."

Here, Silas has an intrinsic concern about Benvolio's pain, *qua* unseemly, but since this concern gets treated by (1) in the mode of the *de re*, rather than in the mode of the *de dicto*, it comes out simply as an intrinsic, *de re* concern about Benvolio's being in pain. Thus, we have a consideration *c* (*Benvolio's being in pain*), *c* is a moral reason to ϕ (to try to get Tybalt to stop), and both Silas and Juliet have an intrinsic *de re* concern about *c* (*Benvolio's being in pain*), that leads them to desire to ϕ (to try to get Tybalt to stop). If a *de re* account of implicit appreciation of moral reasons is right, then it seems that Silas and Juliet equally appreciate the relevant consideration *c* as a moral reason and act as creditable moral agents in response to it.¹⁰ However, this is clearly not the right result. A *de re* account of implicit moral reasons appreciation is inadequate.

The advocate of the *de re* approach may revise her strategy as follows.¹¹ It is important, she might say, that Silas's desire has the right *de dicto* content, in the following sense: the mode of presentation under which *c* is given should be a mode of presentation that picks

¹⁰ Arpaly sometimes suggests there is a distinct quality to Huck's moral vision (see e.g. 2003, p. 76). To the extent that she puts weight on this aspect of Huck's relation to his reasons, she introduces a *de dicto* representational element into her view. Also, as will become clearer later, this *de dicto* understanding of reasons appreciation seems to draw on the very kind of phenomenological account of reasons appreciation I advocate for in this paper.

¹¹ Thanks to David Black and Geoff Sayre-McCord for pressing me on this point.

out the right, morally salient features of the agent's situation.¹² So, it must be out of concern for (say) the *painfulness* of Benvolio's pain, and not the unseemliness of his being in pain, that Silas forms a desire to try to get the pain to stop. On this more advanced account:

(2) ADVANCED DE RE: An agent implicitly appreciates a consideration c as a moral reason r to ϕ iff c is a moral reason to ϕ on account of f , the agent (*de dicto*) represents c as having f , and her intrinsic concern about c *so construed* leads her to desire ϕ (or to desire that ϕ -ing be done).

This account is in keeping with the spirit of the *de re* approach in so far as it does not require the implicit agent to represent the morally salient f features of the cognized consideration *as morally salient*. This fact about the ADVANCED DE RE approach, however, turns out to undermine its plausibility. Consider:

Silas (Advanced De Re): Silas is Juliet's fussy psychopathic cousin. Silas sees Tybalt deal Benvolio the blow to the eye and sees that he is badly injured. Silas realizes he's in great pain and becomes very concerned about it, since Silas has a fetishistic preoccupation with pain. This preoccupation is of such a nature that it leads him to desire to stop the pain. What it is like for Silas here is what it would be like for one of us if we developed a fetishistic obsession with purple balloons, and this obsession was such as to lead us to desire to pop purple balloons whenever we saw them.

Here, there is a consideration c (Benvolio's being in pain), c is a moral reason to ϕ (to try to stop Tybalt), Silas *de dicto* represents *Benvolio's being in pain* as involving *painfulness*, and his concern about *Benvolio's being in pain*, so construed, leads him to desire to ϕ (to try to stop Tybalt). But Silas does still does not present *Benvolio's being in pain* as a moral reason to stop Tybalt.

The problem is that though Silas has a concern about the *painfulness* of Benvolio's pain, that concern is not of the right sort: it does not construe the *painfulness* as being

¹² See Arpaly and Schroeder (2014), Markovits (2010).

bad—perhaps at all—and definitely not in the way relevant to appreciating the *painfulness* of Benvolio’s pain *as a moral reason* to ϕ . Nothing pertaining to the moral weight of painfulness makes any difference to Silas’s perspective on painfulness, and so there’s nothing about the moral badness of painfulness that affects Silas’s motivation to ϕ . The result for our assessment of ADVANCED DE RE is the following: even if it is the case that an agent must have a representation of a moral reason-constituting consideration that targets the morally relevant features of that consideration, such a representation still won’t underwrite implicit moral reasons appreciation so long as those features are not themselves presented, at least in some way, *as morally relevant*.

II. *De Dicto* Accounts of Implicit Appreciation

We need to say something about the kind of perspective that Juliet—and implicit moral agents like her—have on the considerations that constitute their moral reasons, that a lucky psychopath like Silas does not. To distinguish an implicit moral agent’s perspective from that of the lucky psychopath’s, however, it seems we will need to advert to the different ways in which they *de dicto* represent the moral reason-constituting considerations they react to *as moral reasons*.

Moral theorists excessively focused on reflective deliberation have tended to treat the kind of *de dicto* representation of moral reasons that occurs in contexts of reflective deliberation as the only kind of *de dicto* representation of moral reasons that there is.¹³ If one does that, then it is of course natural to conclude that the only way to account for cases

¹³ Markovits (2010) and Arpaly (2003) both motivate their accounts by pointing to the implausibility of the idea that a moral agent must *believe* that ϕ -ing is right to be creditable for it, and the sort of belief they seem to have in mind is a kind of explicit judgmental belief involving a discursive representation of reasons—that

of implicit appreciation like that of Juliet and Huck is to do away with *de dicto* representation of moral reasons altogether, and opt for a DE RE or ADVANCED DE RE account of appreciation instead. But in fact there are a variety of ways in which a consideration can be represented as a moral reason that don't require the kind of *de dicto* representation that occurs in contexts of reflective deliberation.

To clarify what our options are, it is important to begin by distinguishing the *mode of affirmation* an agent takes towards her reasons from the *mode of presentation* under which she construes the reasons thus affirmed. In contexts of reflective deliberation, agents typically affirm their reasons by forming an *explicit judgmental belief* that *a consideration c is a reason*. But agents can also affirm reasons by way of having an *implicit belief* that *c is a reason*, by *seeing* or by having an *intuitive impression* that *c is a reason*, and so on.

Agents can also present their reasons under different kinds of modes of presentation: *c* may be represented as a reason by way of a representation whose contents are inferentialistically characterized, causally determined, functionally determined, phenomenologically determined, and so on.¹⁴ Particular modes of affirming our reasons may tend to go with particular modes of presenting those reasons, but there is no obvious, in principle reason to assume that particular modes of affirmation must of necessity be allied with particular modes of presentation. So, any lucid discussion of implicit moral reasons appreciation must treat these as separable dimensions of analysis. For the purposes of this paper, I will be focusing just on what kind of *mode of presentation* of their moral reasons implicit moral agents plausibly utilize. I leave it to later investigation to ascertain

is, the kind of belief that agents form in reflective deliberation. They do not consider, for instance, perceptual beliefs. They also do not distinguish modes of affirmation from modes of presentation of reasons.

¹⁴ These are the different kinds of representational contents I consider in the sections following.

whether there is some particular *mode of affirmation* of their reasons that such implicit moral agents must also employ.

A. A Discursive Account of Implicit Appreciation

To begin, let's look more closely at what distinguishes the kind of *de dicto* representation of moral reasons that occurs in cases of implicit moral agency from the kind of *de dicto* representation of moral reasons that occurs in contexts of reflective deliberation. Consider a case of reflective deliberation:

Dez's Deliberation: Dez receives a request from a student for extra credit. She's not inclined to grant it at first, but she remembers that she recently allowed another student to do extra credit—and she realizes that both students are in the same class, failing, and have now both asked for the same opportunity. Given the parity of the cases, she realizes that considerations of fairness apply. Given that they do, she infers that she has moral reason to grant the request. She grants the request.

Here, Dez presents the relevant considerations of parity as together constituting a *moral reason* to grant the extra credit request by making use of a fairly sophisticated *inferentialist* grasp she has of a moral reason-giving concept: *fairness*. Dez is able to pick out considerations of parity as morally salient because (i) she knows that such considerations constitute an application condition of the concept of fairness, and (ii) she knows that that concept of fairness, once applied, supports inferences about moral reasons.

Dez's inferentialist grasp of her reasons is critical to her being able to talk herself through her decision in the deliberative, step by step way she does—and it also makes it possible for her to articulate an account of her reasons to others. Because of these features of her inferentialist grasp of her reasons, I'll refer to her kind of appreciation of moral reasons as a *discursive* appreciation. An attempt at a discursive account of implicit moral reasons appreciation would go as follows:

(3) DISCURSIVE: An agent implicitly appreciates a consideration *c* as a moral reason *r* *iff* her representation of *c* plays the inferentialist role it is characteristic of moral reasons-type representations to play—underwriting inferences from salient application conditions to conclusions about (say) moral obligation—and she is able to articulate these inferences (at least with Socratic prodding).¹⁵

Now, this discursive account has more resources for explaining cases of implicit moral reasons appreciation than might initially appear to be the case. As Dez continues to teach, she'll become more adept in her assessments of fairness, and may ultimately be able to bypass deliberation; she may eventually simply be able to intuit or “see” what to do in new cases without first having to formulate any *conscious, occurrent* discursive thoughts about what her reasons are. If she does this, her old discursive, inferentialist grasp of her moral reasons will not thereby have been made obsolete. Her old inferentialist conceptual computers may be running in more sophisticated ways—even having some influence of cognitive penetration on her experience—and so still be fundamentally responsible for how she sorts through information and draws conclusions. If this is indeed the way that Dez's appreciation of her moral reasons develops, it will be evidenced by her continued and non-confabulatory ability to articulate an account of her moral reasons when called on to do so.

A good deal of those who have been interested in making sense of non-deliberative modes of reasons-responsiveness have put emphasis on cases that involve a particular kind of *articulate skilled agency*. These cases involve agents who, like the advanced teaching version of Dez, “see” what to do without first explicitly reasoning about it, but are nevertheless also poised to answer Anscombean “what” and “why” questions about what

¹⁵ This account follows roughly with Jackson and Pettit (1995)'s account of moral concepts. Brandom (1998) and Peacocke (1992) advocate for inferentialism as a general account of concepts.

they do.¹⁶ It's not difficult to see why we should want to understand such cases as cases of reasons-responsiveness, but what accounts for these sorts of cases will not also necessarily account for cases where agents like Huck and Juliet implicitly appreciate and respond to moral reasons. Unlike the skilled Dez, implicit moral agents like Huck and Juliet can't articulate a *post facto* account of their ϕ -ings that advert to a background, tacitly operative inferentialist grasp of their reasons for ϕ -ing. Huck (for instance) may have an inferentialist grasp of concepts like *stealing* and *property*, and he could potentially make use of these concepts in the intuitive, skillful ways that the advanced Dez does. But implicit moral agents like Huck have no similar background possession and use of inferentialistically-characterized concepts—like, in Huck's case, *free and equal moral agency*—that would speak in favor of their chosen course of action. Unlike articulate skilled agents, thus, implicit moral agents do not have a discursive appreciation of their reasons.

B. A Causal Tracking Account of Implicit Appreciation

Philosophers of mind have conventionally offered two ways of thinking about mental representation that might help account for implicit agents' *non-discursive*, but still *de dicto*, appreciation of their moral reasons as moral reasons. According to causal theories of representation, a mental representation *m* represents *p* in virtue of being a token of a type of representation that stands in suitably reliable kinds of external causal tracking relations with *p*-types of things.¹⁷ According to functionalist theories of mental representation, on

¹⁶ See, e.g., Annas (2011), p. 20. Horgan and Timmons (ms) and Railton (2009) also sometimes talk this way.

¹⁷ Here, the notion of representation is very roughly analogous to the notion of representation we use when we say that rings on a tree trunk represent the years a tree has lived: *m*-type representations represent *p*'s by being reliably "triggered" by *p*-type things and so "tracking" them. For classic statements of the position, see (e.g.) Dretske (1988), Fodor (1990), and Millikan (1984, 2004 (Part IV)).

the other hand, a mental representation m represents p in virtue of being a token of a type of representation that plays a certain distinctive kind of internal causal role in a subject's cognitive economy—that is, whatever causal role it is characteristic of p -related representations to play.¹⁸ Neither of these accounts of reasons-representation requires that the subject making use of the representation has much by way of discursive competencies. Indeed, bees are understood as employing mental representations of the relevant sorts despite their obviously limited discursive capacities. In this respect, these accounts of mental representation both hold promise to account for implicit moral agents' non-discursive appreciation of moral reasons. Unfortunately, as we'll see, this promise is not fulfilled.

Using a causal theory of mental representation to give a specification of what it is to implicitly (that is also, *non-discursively*) appreciate a reason, we might say:

(4) CAUSAL TRACKING: An agent implicitly appreciates a consideration c as a moral reason r to ϕ *iff* her representation of c is of a type of that causally tracks moral reasons for ϕ -ing.¹⁹

I'll allow that it might be a necessary condition of implicit moral reasons appreciation, that an agent employ a representation that is a token of a type that causally tracks moral reasons.²⁰ The question I want to investigate is whether the fact that a representation reliably tracks moral reasons in this way is also *sufficient* to underwrite such appreciation.

¹⁸ Here, the notion of representation is very roughly analogous to the notion of representation we use when we say that the mermaid Ariel represents a fork as a hairbrush: her contact with the fork is mediated by a representation that plays the kind of role that hairbrush-representations typically play, causing her to brush her hair with the fork, to think of the fork when she wishes to untangle her hair, put in in her armoire, and so on. For classic statements of the position, see (e.g.) Block (1986) and Harman (1973).

¹⁹ One might prefer to take out “for ϕ -ing” here: perhaps all that is needed is that c tracks moral reasons of any old sort. Even if this is so, though, it does not affect my critique of the CAUSAL TRACKING account.

²⁰ Typically, moral philosophers' interested in causal accounts of mental representation have used such accounts to help explain how moral terms refer. This question is distinct from the question of what it is to

It's not difficult to see the appeal of claiming that it is sufficient. An agent who represents her reasons by "tracking" them, after all, will display a reliable sensitivity to moral reason-constituting considerations.²¹ Nevertheless, a causal account of reasons appreciation is inadequate. This is because there will always be *deviant* ways of tracking considerations. In particular, so long as there is a supervenience of moral reasons on subvening natural properties (or for that matter, on *any* subvening base properties) which can be tracked independently of an agent's needing to take notice of the relevance of those subvening properties for the right kinds of questions of action guidance, it's going to be possible for an agent to track moral reasons in a deviant way unrelated to *appreciation* of those reasons.²² The possibility of these kinds of deviant tracking relations is again going to make it difficult to distinguish a lucky psychopath like Silas from an implicit agent like Juliet:

appreciate a moral reason, but I'll assume for the sake of argument that it is connected at least in this way: that it is a necessary condition of a representation's underwriting appreciation of a reason that the referential content of that representation corresponds in some appropriately intimate way to that of moral terms. (Since theorists making use of causal theories of representation typically argue that moral terms refer fundamentally to natural properties, I'll assume that, if they are right, then the referential content of appreciation-underwriting representations must be intimately related to the relevant natural properties.) Stepping away from the specific question of the reference of moral *terms*, I'll assume for the sake of argument, more generally, that it may be the case that it is a necessary condition of a representation being a reasons-appreciation underwriting representation, that it be of a type that tracks moral reasons. For more discussion of relevant matters see e.g. Wedgwood (2007).

²¹ Though no one has made an explicit claim of such sufficiency (that I know of), we see hints of sympathy in some discussions of implicit skilled agency: Railton (2014), for instance, talks about a sense skilled agents get of their reasons that is marked by a kind of "attunement" to their reasons, and one critical feature of this attunement is that it involves a "tacit susceptibility to information," selected for because it is a "generally reliable guide" to reasons in a "typical range of circumstances" (p. 816, p. 829, p. 832–3). Railton emphasizes the role of affect in attunement; I address the matter of affect in the final section.

²² This claim can be made to work even if Cornell Realism is true. If it is, then there won't be any orderly way to correlate moral properties with other natural properties, but we can still imagine that (say) God could tell Silas what natural properties contingently correlate with any given instantiation of a moral property in a particular case.

Silas (Causal Tracking): Silas sees Tybalt deal Benvolio a blow to the eye and sees he's badly injured. Silas realizes he's in great pain, and the pain is represented under the aspect of the same type of mental representation that, because of (say) the fortuitous forces of species evolution, is also deployed when Silas sees a murder, theft, assault, etc. Like in other cases where that representation is deployed, Silas is weirded out by the represented target. He tries to coax Tybalt to stop.²³

If a causal account of implicit moral reasons appreciation is right, then Silas will appreciate his moral reasons just as Juliet does, because he, like her, will represent considerations via a moral reasons-tracking representation. But clearly, Silas does not appreciate his moral reasons as Juliet does, and so a purely causal account of reasons appreciation won't work to explain implicit moral reasons appreciation.

C. A Functional Role Account of Implicit Appreciation

A further way we might try to distinguish Silas from Juliet is by considering how each represents moral reason-constituting considerations from the standpoint of a functional theory of mental representation. Employing this theory to give an account of implicit reasons appreciation, we might say:

(5) INTERNAL CAUSAL ROLE: An agent implicitly appreciates a consideration c as a moral reason r to ϕ iff her representation of c is of a type that plays the internal causal role that it is characteristic of moral reasons-type representations to play.²⁴

Unfortunately, the possibility of a *spectrum inversion* will always keep functionalist accounts of appreciation from distinguishing Silas from Juliet.²⁵ Take your favorite

²³ The conception of Silas's causal tracking-representation assumed here is Millikan's teleological conception (1984, 2004).

²⁴ The most popular versions of this kind of account are expressivist accounts. I discuss these in more detail in Section Four, Part A (i).

²⁵ Noting that there is a striking isomorphism between what reasons-representations a subject employs and the courses of action she opts to take, some suggest that we understand the internal causal role that it is characteristic for a moral reason-representation r to play as *disposing the subject to do the ϕ -ings that r*

causal-dispositional profile for moral reasons-representations and match it with the wrong phenomenal feel; I submit that you will get a state that intuitively fails to underwrite moral reasons appreciation.

Silas (Functional a): When Silas thinks about Benvolio's pain, he experiences (associated with that thought) a feeling of itchiness all over his body. In virtue of this, Silas has the urge to stop the pain, which he senses will relieve the itch, and so tells Tybalt to stop. Silas has the same experience with other representations of people being in pain: they feel itchy, and steps to relieve the represented pain relieve the itch.

Silas still fails to appreciate Benvolio's injury as a moral reason for Tybalt to stop, even though (we can imagine) he has all the same functional causal dispositions as Juliet: he rushes equally quickly to aid Benvolio, he attends to all the same features as she does, perhaps displays punishing behaviors towards Tybalt in the future, and so on.²⁶ And we can generalize the problem. We can "wire" into Silas, at birth, all the dispositions we need

speaks in favor of (see e.g. Blackburn (1998)). Of course, as stated, this proposal is too general: it is not just *moral* reasons-representations, but reasons-representations as a whole that play the role in question. And it is difficult to mark out what more specific set of dispositions D_M would be necessity be associated *specifically* with *moral* reasons-representations. A theorist might claim that D_M dispositions are especially likely to win out over other competing dispositions (Mackie 1977), that they have a longevity or sticking power that other dispositions don't (Blackburn 1998), that they tend to involve stronger tendencies towards self-reinforcement and self-correction than others, or that they involve distinctive patterns of attention—but there seems to be no in principle reason why someone who prioritized (say) prudential goods over moral ones, could not also display these same kinds of dispositions. These are further reasons to reject a functionalist account.

²⁶ It's tempting here to want to appeal to differences in emotions that Silas and Juliet are apt to feel. For instance, one might want to say that Juliet can be differentiated from Silas because she is disposed to feel guilty if she doesn't help, and Silas is not. If emotions are understood purely in terms of the non-phenomenological causal dispositions they are associated with, then this proposal won't work for the same reason that non-phenomenological causal functionalist accounts generally don't work. You could have, for instance, the feel of itchiness play the same role in bringing about behaviors, controlling attention, etc. as guilt does. On the other hand, if you characterize guilt in terms of its phenomenology, then you are already part of the way to my account—in fact, I argue in Section Four, Part A (i) that you are already implicitly committed to a version of my account.

him to develop to duplicate the causal-functional profile of a typical moral agent, but then invert the phenomenological feels normally associated with these dispositions:²⁷

Silas (Functional b): The dispositions associated with our representation of a moral reason not to ϕ , instead of being harnessed to feelings of indignation and the like, are harnessed to feelings of unseemliness and being weirded out, and the dispositions associated with our representation of a moral reason to ϕ are harnessed, instead of to feelings of moral admiration and the like, to feelings of arousal and titillation.

I'll grant that it may be the case that agents' implicit appreciation of their reasons may necessarily involve some distinctive internal functional dispositions. But, if a functional account of appreciation of reasons is right, and that is *all* that such implicit appreciation consists in, then Silas (Functional a and b) and Juliet will equally well appreciate their moral reasons and act as creditable moral agents in response to them. Clearly, however, this is not the right result, and so a functionalist account of implicit moral reasons appreciation is inadequate. It, like the *de re* account and the causal account of implicit appreciation, fails to differentiate lucky psychopaths from genuine moral agents.²⁸

III. A Phenomenological Account of Implicit Appreciation

We have seen that having the *wrong* kind of phenomenal feel associated with your representation of a consideration can disqualify you from implicitly appreciating that consideration as a moral reason. So, it looks to be the case that having the *right* kind of phenomenal feel associated with your representation of a consideration is a necessary

²⁷ Of course, some may argue that phenomenology itself is best understood functionally. My account permits that this may be the case. I am concerned only to show that that category of phenomena normally grouped together under the category of "phenomenology" are critical to moral reasons-appreciation.

²⁸ I discuss another version of the functionalist account in the Section Four, Part A (i).

component of implicitly representing that consideration as a moral reason.²⁹ If so, then the kind of mental representation of reasons that underwrites implicit moral reasons appreciation will be accounted for by a *cognitive phenomenology* theory of mental representation. Advocates of this theory hold, roughly, that the content of a given mental representation is determined by the distinctive phenomenal quality of the experience a subject undergoes when she deploys that representation in her cognitive life. Put more formally, they say that a mental representation *m* represents *p* by virtue of being a token of a type of representation *M* that is associated with a distinctive type of phenomenal feel *Q*.³⁰ There are weaker and stronger versions of the cognitive phenomenology theory of mental representation, but for the purposes of this paper I'll assume in the background a weaker version of the theory that holds (first) that it is enough for a type *M* to be "associated with" a feel *Q*, that a token *q* of *Q* is experienced in certain paradigmatic and non-derivative cases of *M*-tokenings,³¹ and (second) that *Q* may in principle be either a *sui generis* cognitive phenomenal feel or otherwise a kind of cognitive phenomenal feel that is reducible to other kinds of phenomenal feels (sensory, affective, etc.), perhaps "worked up" in particular ways, or integrated together, so as to serve in a representational capacity.³²

²⁹ At least in so far as appreciation is concerned. There may be ways of representing moral reasons as moral reasons that are not sufficient to underwrite appreciation.

³⁰ See e.g. Horgan and Tienson (2002), Horgan, Tienson, Graham (2004), Horgan and Graham (2013), Loar (2003), Pitt (2004, 2011), Siewert (1998, 2011), and Strawson (2011).

³¹ One can think of this roughly on the model of color representations. There is a kind of representation of *red* that regularly-sighted people have, that others don't. This doesn't mean that *every* time that a regularly-sighted person thinks or talks about red things, that she is overcome with the relevant *red* representing phenomenology. Rather, the phenomenology associated with that kind of representation of *red* is experienced only in paradigmatic and non-derivative cases of that kind of *red*-tokening.

³² See Footnote 30. I'll also assume that we do *not* need to accept that the cognitive phenomenology theory of representation explains why *all* kinds of mental representations have the contents they do, to show that it might explain why one kind of mental representation has the contents it does.

A. A Cognitive Phenomenology Account of Implicit Appreciation

Some cognitive phenomenologists take the phenomenal feel associated with a representation to be wholly responsible for determining the contents of that representation. For our purposes, though, we only want to know whether the phenomenal feel associated with moral reasons-representation is at least *partially* responsible for determining the contents of those representations—and that, only in so far as moral reasons-representation is relevant to implicit *appreciation* of moral reasons. To show that cognitive phenomenology can help account for implicit moral reasons appreciation in this way, we need to show that (1) there is a distinctive phenomenal feel associated with the appreciative representation of a moral reason (at least in paradigmatic cases)³³ and that (2) whether or not an agent has this distinctive feel is at least one critical factor determining whether or not she represents a consideration as a moral reason in the relevant appreciative way. Below, I begin by discussing *prima facie* reasons for thinking that phenomenology is partially responsible for moral reasons appreciation, *generally*, and then discuss the more specific case of *implicit* reasons appreciation. I return again to defend the general claim about moral reasons appreciation in more detail at the end of the paper.

First, there does seem to be a distinctive phenomenology associated with the appreciative representation of a moral reason.³⁴ Consider:

Fair Trade: You are an avid clothing shopper and a bargain hunter: you don't see any reason not to insist on getting the absolute best deal on every item of clothing you buy. Then you attend a lecture about the labor practices of major clothing retailers, and find out about the miserable conditions that foreign workers are

³³ It may be important to distinguish “appreciative” representation of a moral reason from representation of a moral reason, *per se*, because it could turn out that philosophers of mind become interested in a notion of representation of moral reasons that is perfectly respectable in its own right, but has little direct connection to the kind of representation of reasons that is relevant to evaluations of moral agency and creditability.

³⁴ See previous footnote for use of the terminology of “appreciative representation.”

forced to bear to allow retailers to offer bargains. You see pictures, hear about the hours the workers have to labor, and are told about the history of colonialism and undercurrent of racism that seems to fuel these practices. You're also told that things might be even worse for the workers if retail companies simply stop producing in their area, and that the only solution is to support fair trade retail companies instead. The clothing from these fair trade retailers is significantly more expensive and so you're not otherwise inclined to purchase them. Still, you now feel compelled to change your shopping habits, since you realize there are excellent moral reasons to purchase the more expensive fair trade items.

When you realize there is a reason to buy fair trade clothing, you feel a force directed at *buying fair trade clothing*, lending weight to it and soliciting you to do it, in a particular authoritative way—even despite your own prior standing desire to only get the cheapest items. This force is anchored in the considerations you appreciate as moral reasons: *the workers' miserable conditions* (for instance) presents itself with an authoritative directive force “pointing” you to purchase items from fair trade companies. Similar phenomenology is also present in other cases where a consideration is appreciated as a moral reason. You experience, say, the consideration *that your elderly neighbor is lonely and loves baseball* as lending weight to taking her to the game;³⁵ you think of the cruel treatment of animals and you experience that consideration as mandating you to take your hands off the veal and to reach for the tofu instead; perhaps you consider the categorical imperative, and feel solicited to be honest in your tax papers.³⁶

Whether or not the above distinctive solictive phenomenology is associated with a representation seems to make a critical difference to whether that representation underwrites moral appreciation. Thus, in the examples of Silas that we have considered,

³⁵ Granted, the moral reason here is not a requiring reason. As a result, the nature of the feel of the authoritative solicitation may take on a slightly different aspect. It may solicit without (for instance) *commanding*.

Silas's failure to appreciate considerations as moral reasons has consistently been correlated with a faulty phenomenology. We have considered Silas as entertaining *Benvolio's pain* under the aspect of the feeling of a fetishistic preoccupation, itchiness, and an unseemliness that weirds him out, and we have determined in each of these cases that he does not appreciate relevantly entertained considerations *as* moral reasons. Part of what is "off" about Silas's representation of *Benvolio's pain* in these cases seems to be the quality of Silas's experiential presentation of *Benvolio's pain*. Silas's experiential presentation of pain is quite different from the experience that the moral reasons-appreciating agent in FAIR TRADE has when she considers the *miserableness of workers' conditions*. Silas's experiential presentation of *Benvolio's pain* is also plausibly quite different from Juliet's experience of *Benvolio's pain*. In both instances, the relevant moral reasons-appreciator plausibly entertains the relevant considerations under the aspect of a particular kind of felt authoritative solicitation that is unlike (say) the obsessive feel of Silas's fetishistic preoccupation, or the irritation of his itch, or the kind of unpleasantness associated with his being weirded out.³⁷ Whether an agent does or does not experience a felt authoritative solicitation, then, seems to make some difference to whether she does, or does not, appreciate a consideration as a moral reason.

In light of the above, I propose (6) as an account of how agents like Huck and Juliet implicitly appreciate their moral reasons:

(6) PHENOMENAL MORAL REASONS: An agent implicitly appreciates a consideration *c* as a moral reason *r* to ϕ only if the representation of *c* is (at least in

³⁶ Consider also Mandelbaum (1955) "I sense the embarrassment of a person, and turn the conversation aside, I see a child in danger and catch hold of its hand; I hear a crash and become alert to help" (p. 48).

³⁷ Silas's unusual experiences of *Benvolio's pain* do give rise, perhaps, to some feeling of being *incited* to ϕ (to stop Tybalt), but—as we'll say more about soon—that feeling of being incited to ϕ seems to be quite different from the authoritative solicitation felt when we appreciate a consideration as a moral reason.

non-derivative cases) associated with phenomenal quality q_M , where q_M involves a felt authoritative soliciting force aimed at ϕ -ing.³⁸

To help us get a better grip of what q_M is, however, and to show the credibility of (6) as a proposal about *implicit* moral reasons appreciation, I'll now suggest that the feeling of felt solicitation in question stands as one among a variety of similarly implicit kinds of non-discursive, experiential representations that have modal contents. The feeling q_M , I'm going to argue, is particularly tethered to experiential representation of *deontic* modal contents.³⁹

B. A Non-Discursive Experiential Representation?

To begin, consider our representational capacities as they pertain to a broad and generic class of possibilities concerning how physical states of affairs in the world *are*, *might*, or *will be*. When I see a ball start rolling from the top of an incline with nothing in its way, I can see that it will roll to the bottom. When I see a log teetering just at the edge of a cliffside, I see at first that it is precariously perched, and so *might possibly* fall; then I see, as it teeters, that it is *almost certainly* going to fall, and then that *it is now about to and is definitely going to fall*. Here, I experience what we might roughly categorize as *nomic modalities*.⁴⁰ These are experiences as of realities that we don't need sophisticated

³⁸ This proposal of what it is to appreciate a reason is intuitive and has some historical roots. It draws on a basic idea already familiar from the Stoics' (and later Augustine's) moral psychology: that we experience "hormetic" impressions, which are imperatival practical perceptions that command us to ϕ . See e.g. Brennan (2003) and Inwood (1985). On Augustine see Byers (2013).

³⁹ In using the term "deontic" here I do not mean to connote a specifically Kantian notion of moral reasons, or even a notion of moral reasons more generally. Rather, I use deontic in the sense that it is used by logicians when they talk about "deontic logic" or metaphysicians when they talk about "deontic" as opposed to (say) "nomic" or "epistemic" modality.

⁴⁰ Siegel (2014) talks about these as proto-affordances. These experiential representations may or may not have any distinctive phenomenal quality associated with the different contents. They may not—but then again, they might, as evidenced by the differences we feel when these kinds of representation come in large doses: walking through giant slabs of solid granite cliff faces—where it seems nothing is going

discursively articulable concepts to represent. Thus, a small child who lacks an inferentialistically sophisticated grasp of the concept of *nomic possibility* can nevertheless see that the log *might fall* (*now will fall*, etc.) in the relevant way.⁴¹

A second kind of modality that it seems we can non-discursively, experientially represent is *dynamic modality*—that is, for our purposes, modalities involving possibilities and necessities that involve us as agents. Here, on the side of possibility, you can (for instance) look down a path, see obstructions, but also see that the path *can be traversed*. This kind of representation of a possibility for agency in an environment has (following the work of J.J. Gibson) been termed by some psychologists as an “affordance.” Thus, they say, a chair *affords* sitting, a door *affords* opening, and so on.

In addition to experientially presenting ϕ -ings as things that (dynamically) *can possibly be done*, we can also experientially present ϕ -ings as things that (dynamically) *must be done*.⁴² Thus, for a faint person, a chair might be experienced as something that *must be sat on*, or a door can be experientially represented by an obsessive-compulsive or claustrophobic person as something that *must be opened*. As before, presentations of dynamic possibilities and necessities for agents don’t seem to require much discursive competence on the part of the representing subject. Plausibly, a small obsessive-compulsive child could present the door as *having to be opened*, for instance, without having much of anything by way of inferentialist conceptual competences to talk

anywhere—plausibly has a different feel than would (say) walking through the world of Dali’s paintings—where anything could happen.

⁴¹ Similarly, a small child who lacks an inferentialistically sophisticated grasp of the concept of *fragile* can see at the moment that her brother lets go of her Lego-like structure from the top of the stairs, that it *will break*.

about the dynamic necessity of opening doors.⁴³ In these cases more so than in the nomic ones, the distinctive phenomenology involved in experiential presentations seems to play a role in determining the contents of what is experientially presented. It seems to be by virtue of the kind of phenomenal feel that compulsion has, for instance, that the targeted ϕ -ing is represented as something that *has to be done*:⁴⁴ it is plausibly on account of the feeling of being driven, pushed, or propelled to ϕ which is involved in a compulsion, that ϕ -ing is represented as something that (dynamically) *has to be done*.

In addition to having experiences that present it as being the case that something *might* or *will happen*, or that something *can* or (dynamically) *has to be done*, we also have experiences that present it to be the case that something normatively *may be done* or *is required to be done*. Here, on the one side, we have representations on the side of possibility in (say) cases of arbitrary choice. Consider a case where you look down two paths in a wood, and see what appears to be a nice place to rest under a tree down one side. The representation of this consideration may have a feeling of felt relevance associated with it as you consider whether you might want to sit. But then, when you decide you have no desire to sit, that felt relevance comes to naught: it provides no soliciting force pointing you towards or urging you to go down that path more than the other.

⁴² As I read Siegel (2014), her talk of affordances and experienced mandates fall primarily in the dynamic category. However, she is not clear to distinguish dynamic from deontic experiential representations, and some of what she says could be read to support a more deontic reading.

⁴³ Moreover, she could plausibly distinguish that kind of necessity from the pure causal necessity involved in apprehending that (e.g.) the giant boulder will knock down the door.

⁴⁴ Thus (for one), if you take the feeling of the compulsion away from an experience, it seems you will have affected our grounds for understanding the experience as presenting ϕ -ing as something that *has to be done*. Further, if you feel a compulsion to ϕ but judge ϕ -ing not to be necessary, you would plausibly experience some *cognitive* dissonance in a way that you would not with, say, an accompanying feeling of nausea. This seems to be evidence that the compulsion involves represented content in a way that the sensation of nausea does not.

Contrast this with cases, on the other side, where we have stronger feelings of deontic necessity. You see someone being harassed on the street and have an experience as of being mandated to step in; or an artist coming close to the end of her performance feels compelled by the music—aesthetically demanded—to perform a grand *finale* of *this* sort. These considerations—that *someone is being harassed*, or the flow of the music up to this point—are associated with an authoritative, mandating force aimed at ϕ -ing.⁴⁵ It's not just that the relevant ϕ -ing *could* be done; rather the consideration *calls out* for the ϕ -ing to be done in a solicitive and mandating way. Further, the relevant felt force doesn't just drive, push, or propel you to ϕ (though it might *also* have the effect of stimulating such a feeling): rather it asks or *bids* you to ϕ by virtue of a kind of governing authority. Thus it is that you “feel the force” of the relevant consideration *as* a reason.⁴⁶

As in the previous cases, an agent plausibly doesn't need discursive grasp of relevant deontic concepts to have experiences with these deontic contents. Consider Jasmine:

⁴⁵ These experiences of deontic force can come in a range of different strengths; I just highlight the extremes. So, for instance, ϕ -ing can be presented in experience as something that there would be some merit in doing, without its being experienced as *mandated*. Thus, the consideration that (say) my elderly neighbor is lonely and loves baseball may be experienced as having a felt relevance to the question of whether to bring her to a baseball game, and as having some directive force, “pointing” in a directive way to an answer of whether to bring her to a baseball, without also being experienced as exhorting me to bring her to a baseball game.

⁴⁶ A fourth kind of modality that it seems we can non-discursively and experientially represent—that may, taken in the right light, overlap with the deontic case—is epistemic modality. Here, on the far side of possibility, we have cases where a representation of a consideration *c* has no felt relevance to an entertained proposition *p*. The representation of *c* and *p* may each other in a stream of consciousness without any felt connection. On the side of necessity, we might employ language used by Locke and say that a representation of a consideration *c* imbues an “evident luster” on *p*—a luster that involves a pointing to *p* as *to be assented to*. To see the contrast between these, consider a case where you represent *That it is May 12* and then, in a stream of consciousness, you also happen to think *Today is my sister's birthday*. Contrast that case with a case where you represent *That it is May 12* and this directs you to the thought *Today is my sister's birthday*, imbuing the latter with an evident luster. Some propositions may come with their own evident luster without needing help from represented considerations; these propositions may be “clear and distinct” *à la* Descartes, and carry their *to be assented to-ness* on their sleeve. Other discussions of the phenomenal appreciation of epistemic reasons include Feldman and Conee (2004) and Plantinga (1993), p. 57–58 (1993), p. 105–6.

Jasmine: Jasmine is a young jazz piano virtuoso. Listening intently to the flow of the music, she hears how the particular timing and way in which a E-flat-diminished-seventh chord is played by the band. Jasmine doesn't know any formal jazz theory and doesn't have any discursive or inferentialist grasp of the concept of an "aesthetic" reason, but when she hears that chord, it is experienced by her as having a kind of directive force pointing to and demanding her to ease into a D-minor-seventh chord *just so*.

Jasmine plausibly appreciates the consideration that the E-flat-diminished-seventh chord was played *just so*, as an aesthetic reason to ease into the D-minor-seventh chord *just so*, even despite not having any discursive grasp of the relevant concepts in jazz theory that would allow her to articulate what considerations served as her reasons. Even more to the point, she plausibly appreciates the consideration that the E-flat-diminished-seventh chord was played *just so, as an aesthetic reason* without having the kind of discursive grasp of the concept of an *aesthetic reason* she would need, to understand herself as responding to an aesthetic reason.

Plausibly, Huck is like Jasmine. Just as the particular way in which an E-flat-diminished-seventh chord is played by the band can strike Jasmine as an aesthetic reason to ease into the D-minor-seventh chord *just so*, without her employing (or being able to employ) any discursive representation of the former as an *aesthetic reason* for the latter, so also, a consideration can strike someone like Huck as a *moral reason* to ϕ without his employing (or being able to employ) any discursive representation of the former as a moral reason for the latter. A consideration can strike someone like Huck as a moral reason to ϕ without his employing (or being able to employ) any discursive representation of the

former as a moral reason for the latter, in virtue of its being associated with a kind of directive or soliciting force “pointing” to ϕ -ing.⁴⁷

C. A Distinctive Moral Phenomenology?

Does Huck’s experience of his reasons *as moral reasons* have any distinctive quality to it that would set it apart from other kinds of experiences of reasons—say, e.g., Jasmine’s experience of her reasons *as aesthetic reasons*?⁴⁸

To begin, there’s at least one category of phenomenal feels associated with presentations of non-moral reasons that is pretty easy to distinguish from the phenomenal feels associated with presentations of moral reasons: that is the phenomenal feel associated with presentations of what we might think of as *enticing reasons* (following Dancy and others, we can think of these as reasons that make some ϕ -ing look appealing or attractive).⁴⁹ Plausibly, when Silas presents the *itchiness of Benvolio’s pain* as a reason to try to stop that pain, that consideration is presented as a kind of enticing reason to try to stop the pain in virtue of the fact that doing so will relieve the itch, and it is enticing and appealing to have the itch stop. Here, it seems, Silas’s reason can be presented as an enticing reason by virtue of being associated with a feeling that, unlike feelings of

⁴⁷ These phenomenal qualities aren’t just arbitrarily paired with contents; instead, they are plausibly part of what makes the relevant experience have the contents it does. Consider the difference between, on the one hand, (i) judging that it is not good that you feel itchy (since it is distracting that you feel that way), and yet feeling itchy, and, on the other hand, (ii) judging (perhaps by rationalizing) that you have no moral reason to give someone begging on the street charity, and yet experiencing a felt solicitation to give him charity. In (ii) there is a kind of cognitive dissonance that occurs that does not occur in (i). Like when (e.g.) you look at a Müller-Lyer illusion and it appears to you that the lines are different lengths, but you judge that they are not, in (ii) you feel the world to be presented to you as being a certain way, and this feels to conflict with your judgment. This is evidence that the feel of directive force has representational content. See also Footnote 44 and Footnote 61.

⁴⁸ For other discussions of moral phenomenology see Hampton (1998), Horgan and Timmons (ms), Mandelbaum (1955), and Kriegel (2008).

⁴⁹ Dancy (2004).

associated moral reasons, is at its core a feeling of being incited, propelled, or *driven* to ϕ . (These feelings will fall on the same spectrum as non-discursive presentations of dynamic facts, in other words.) Such feelings do not “pull” or solicit you to ϕ in the authoritative, mandating way that considerations presented as moral reasons do.⁵⁰

Beyond this, it gets more difficult to specify the precise sense in which a phenomenal feel might be associated with the appreciation of a specifically *moral* reason to ϕ , but there are plausibly some salient phenomenal contrasts in the vicinity. Consider the quality of the experience at stake in these variations on a scenario:

You are at a cafe relaxing, and there is a band playing some great early 1900s ragtime music in the background.

(i) Your devious friend gives you a dare: she’ll fix whatever is causing the check-engine light on your car to come on, if you go up to the singer and tell him to stop singing. You do need the check-engine light checked out and the mechanic is exorbitantly expensive. With that in mind, you consider whether to go ask the singer to stop singing.

(ii) You close your eyes and take in the music. You notice that the singer is off key and also not staying on beat with the rest of the band. You know the instrumental version of the song the band is playing well, and can feel how beautiful it would be if he was not singing. With that in mind, you consider whether to ask the singer to stop singing.

(iii) You start listening to the lyrics of the song and realize they are very racist and misogynistic. With that in mind, you consider whether to ask the singer to stop singing.

The feeling of solicitation you feel in case (iii) is the directedness involved with appreciating a moral reason (q_M) and it seems to be distinct from the feeling of directedness one would feel in (i), where you consider practical reasons—call that feeling q_P —and the feeling of directedness one would feel in (ii), where you consider aesthetic reasons—call that feeling q_A .

⁵⁰ Hampton (1998), Ch. 3 discusses this in great detail.

With respect to (i): q_M is distinctive in that the force soliciting you to *ask the singer to be quiet* is authoritative and governing, in a way that q_P is not. Part of what seems to be operative here is that there is a kind of *categorical* feel to q_M : the directedness feels to be sourced in a ground independent of your pre-existing desires.⁵¹ Further, there is plausibly some kind of *inescapability* in the feel of q_M , that there is not with q_P . I am bound by prudential reason to ask the singer to be quiet *if* I want a free engine repair, but, unlike in the moral case, I can untie this binding: I can let go of that desire and escape the authority of the reason. This plausibly comes through in the phenomenal feel associated with this representation of a practical reason, since there is a difference of feeling between being tied to a mast by a rope you can untie, and being tied to a mast that is entirely inescapable.⁵²

With respect to (ii): the feel of q_A does intuitively have some difference of quality from that of q_M , but it is more difficult to capture that difference in a description. The feel of the soliciting force associated with q_A may, after all, also involve at least some kind of categoricity and inescapability. Part of what may distinguish q_A is that the relevant feel of categoricity and inescapability is plausibly rooted in a sense of the fittingness of the unified coherence or completeness of the object of aesthetic appreciation that is to be had.⁵³ The feeling of being compelled to establish such coherence is distinct from what is felt with q_M .

⁵¹ Horgan and Timmons (ms) also discuss this.

⁵² Importantly, the relevant feeling of inescapability can pertain even in cases where you appreciate a moral reason that is not requiring. If you don't take your lonely elderly neighbor to the baseball game you perhaps don't accrue the dis-merit associated with failing to act on a requiring reason, but you will not escape what (non-requiring) moral pressure there is, to take her. As a result you may, for instance, "feel bad" that you didn't take her, without feeling "guilty" as such.

⁵³ See e.g. Beardsley (1958). According to Beardsley, "unity is a matter of coherence and of completeness... [and] coherence, in turn, is a matter of having elements that are properly connected one to another such that [o]ne thing leads to another; continuity of development, without gaps or dead spaces, a sense of overall providential pattern of guidance, an orderly cumulation of energy toward a climax, are present to an unusual degree" (p. 527–528).

A further difference between q_A and q_M may have to do with the (alleged) fact—often adverted to in philosophy of aesthetics—that appreciation of aesthetic reasons is necessarily allied to a kind of pleasure or anticipated pleasure, which the appreciation of moral reasons does not seem to be necessarily allied to.⁵⁴

Given the above, we can update (6) to (6'). (6'), I propose, is the best account of implicit moral reasons appreciation available to us:

(6') PHENOMENAL MORAL REASONS: An agent implicitly appreciates a consideration c as a moral reason r to ϕ only if the representation of c is (at least in non-derivative cases) associated with phenomenal quality q_M , where q_M does not by necessity involve pleasure, and does by necessity involve a felt authoritative directing or solicitating force aimed at ϕ -ing, that has a categorical, inescapable feel to it.

IV. Is the Phenomenal Representation of Moral Reasons a General Necessary Condition of Moral Reasons Appreciation?

So far I have argued that part of what it takes for an agent to implicitly appreciate a consideration as a moral reason to ϕ , is for her to present that consideration via the light of a felt directive force “pointing” towards ϕ -ing—lending weight to it, or soliciting it—in a particular way. Some of the discussion above, however, has intimated that this condition on implicit moral reasons appreciation may also be a condition of such appreciation more generally. In this section, I give two preliminary arguments for thinking that it is. The first argument proceeds by process of elimination of familiar views in the literature. The second points to problems about understanding moral motivation that suggest that there is a specific positive need for an account of moral reasons appreciation like mine.

⁵⁴ Philosophers of aesthetics drawing from Kant’s *Critique of Judgment* often advert to the role of pleasure in aesthetic appreciation.

A. The Final Elimination

We've already seen that causal and functionalist theories of mental representation are not adequate to account for implicit moral reasons appreciation without some reference to phenomenology. The arguments used to show that inadequacy did not hang on anything specific to do with the *implicitness* of the reasons appreciation at stake, and so those arguments also serve to show the inadequacy of the relevant theories of mental representation to serve as full accounts of ordinary moral reasons appreciation, more generally. There are still two final kinds of accounts of ordinary moral reasons appreciation, however, that we should discuss. If we can show that these, too, are inadequate to support ordinary moral reasons appreciation all on their own, we will have given strong presumptive reason to think that the phenomenal representation of moral reasons is a necessary condition of genuine moral reasons appreciation not just in cases of implicit agency, but also generally.

(i) *An Affective Account of Appreciation*

Let us return to the ADVANCED DE RE account, according to which:

(2) ADVANCED DE RE: An agent implicitly appreciates a consideration c as a moral reason r to ϕ iff c is a moral reason to ϕ on account of f , she (*de dicto*) represents c as having f , and her intrinsic concern about c *so construed* leads her to desire ϕ (or to desire that ϕ -ing be done).

We saw that (2) is inadequate because the quality of an agent's concern about c can be of the wrong sort: Silas's concern about *Benvolio's pain* can be constituted by a mere fetishistic preoccupation with that pain. The phenomenal account of reasons appreciation suggests that what is wrong with this fetishistic preoccupation with *pain* is that it fails to experientially present *pain* under the light of the right kind of categorical, moral solicitive force aimed at ϕ -ing. But an advocate of the ADVANCED DE RE approach might propose

yet another revision to her proposal: she might propose that we can set aside questions about how *Benvolio's pain* is phenomenally construed if we just ensure that the relevant *desire* to act on *Benvolio's pain* is of the right sort. More generally, one might argue that an agent's appreciation of a consideration as a moral reason to ϕ can be explained just by reference to the kinds of *pro-attitudes* that agent has towards acting on that consideration.⁵⁵

We already saw reason to think that this kind of proposal won't fully explain moral reasons appreciation if the relevant pro-attitudes are characterized entirely in terms of functional causal dispositions that do not account for phenomenology.⁵⁶ But it's worth considering whether such an account can be made to work if phenomenology is taken into account⁵⁷—if the pro-attitudes *are* characterized in part by the phenomenology of (say) desire, approval, or the like. To mark out the special phenomenal element I am adding, I make reference to *pro-attitude affects*:⁵⁸

⁵⁵ This is roughly the view advanced by Blackburn (1998) and Gibbard (1992).

⁵⁶ In Section Two, Part C. Blackburn (1998) especially focuses on the causal dispositional properties of the relevant pro-attitudes. To further see the inadequacy of such an account, consider also moods and obsessive desires. These can be made to play the same kinds of *causal* roles as stereotypical pro-attitudes, but yet not support moral reasons appreciation. The obsessive desire case is discussed in the text. To see the argument in the case of moods, consider:

Silas's cheer: Silas sees some children playing. Though he's a psychopath, this makes him happy and cheerful. When he's cheerful in this way, he's not inclined to curse.

Silas's cheer plays the causal role of connecting a represented consideration *c* (*the presence of children*) to his disposition to do the ϕ -ing that *c* gives him moral reason to do, but it doesn't manage to make out that represented consideration *as a moral reason*—or really, any kind of reason at all—to ϕ (*not to curse*). The causal properties of the affect, considered in and of themselves, are thus not enough to underwrite reasons appreciation.

⁵⁷ Of course, some may argue that phenomenology itself is best understood functionally. My account permits that this may be the case. I am concerned only to show that that category of phenomena normally grouped together under the category of “phenomenology” are critical to moral reasons-appreciation.

⁵⁸ Railton (2009, 2014, 2016) has recently written a lot about how *affects* underwriting agents' creditable responsiveness to reasons. See also Starkey (2008). Unlike Blackburn and Gibbard, Railton and Starkey are not particular intent on avoiding reference to phenomenal features of the relevant affects.

(7) **AFFECT:** An agent appreciates a consideration c as a moral reason r to ϕ iff her representation of c underwrites an appropriate pro-attitude affect a towards ϕ -ing.

What needs to be true of a for it to support appreciation of c as a moral reason? One feature that a must have can be brought out by contrast with a kind of pro-attitude affect that does *not* succeed at underwriting moral reasons appreciation. Consider:

Silas's Compulsion: When Silas thinks about Benvolio (who is by now in the hospital) as being in pain, it triggers in him a brute compulsive desire to *send a get well card*. The desire, he feels, could be resisted, but it feels totally inexplicable and out of the blue. What it is like for Silas is what it would be like for us if hearing a high-pitched sound suddenly triggered a brute compulsive desire to stand up and jump up and down. The drive to ϕ feels to be caused by the stimulus, but the ϕ -ing you thus feel driven to do doesn't *feel* to "make sense" in light of that represented consideration.

Here Silas's brute compulsive desire doesn't serve to underwrite appreciation of *Benvolio's pain* as a moral reason—or really, *any* kind of reason at all—to ϕ (to *send a get well card*).⁵⁹ The problem is that, though the feeling of brute compulsion involves a kind of "pro" presentation of the relevant ϕ -ing as *to be done*, that "pro" presentation of ϕ -ing presents ϕ -ing as *to be done* in the wrong manner: in a manner that fails to *connect* that *to be donedness* of ϕ -ing (*sending a get well card*) in an appropriate way to the relevant "triggering" consideration (*Benvolio's pain*). If *Benvolio's pain* is going to be construed as a reason to *send a get well card*, then the relevant pro-attitude affect a has to feel different than this compulsion: it must make the *to be donedness* of *sending a get well card* feel, not *merely* causally triggered by the representation of *Benvolio's pain*, but rather made

⁵⁹ This is not to say that appreciated moral reasons might also give rise to something like compulsive desires. There may be some kind of compulsion-like desire that is sometimes involved with appreciating a practically pressing moral reason to ϕ . My claim is just that such a compulsion will not be *all* there is to appreciating the relevant consideration as a reason. Also, I think it's plausible that a moral compulsion will have a different

worthwhile, warranted, or—most broadly—as something that “makes sense” to do, in light of that consideration. Silas must feel that ϕ -ing is to be done *on the basis* of that consideration, where this feel of ϕ -ing as to be done on the basis of *Benvolio’s pain* must be kept distinct from the feel of being “triggered” by *Benvolio’s pain*. Thus, if the desire to *send a get well card* is going to help elevate *Benvolio’s pain* to the level of a represented reason, that desire must be of a particular sort: it must construe *sending a get well card* as something that is worthwhile or “called out for” by *Benvolio’s pain*.⁶⁰

If a pro-attitude affect *a* must construe the ϕ -ing that is to be done, as something that is worthwhile to do on account of *c*, though, that means that *c* itself must also be construed in a particular way: *c* must be construed so as to make ϕ -ing feel warranted or worthwhile. *Benvolio’s pain* must thus itself in some way be construed so as to make it possible for *sending a get well card* to feel worthwhile, in light of it. But how is it that a consideration must be construed, so as to make that consideration, in turn, make some ϕ -ing feel worthwhile or sensible on account of it? Plausibly, a consideration will be construed in the needed way if it is construed via the light of some kind of felt directive force aimed at ϕ -ing. If I have desire to *send Benvolio a card* that manages to elevate *Benvolio’s pain* to the level of some kind of represented reason to *send Benvolio a card*, that will be on account of the desire building into itself some feeling of *Benvolio’s pain* “directing” me towards *sending Benvolio a card*.

Now, since there may be multiple ways in which *c* can be construed so as to have it direct me towards ϕ -ing, a desire to ϕ on account of *c* that supports appreciating *c* as a

phenomenal quality to it that the “brute” compulsion to jump up and down when you hear a high pitched sound. Thanks to Laura Callahan for pressing this point.

moral reason to ϕ has to construe c under the light of the *right* kind of feeling of directedness towards ϕ -ing. If *Benvolio's pain* is unseemly to Silas, *Benvolio's pain* will be associated with a feeling of directedness towards ϕ -ing. A desire to ϕ that incorporates *that* feeling of directedness may serve to construe *Benvolio's pain* as *some* kind of reason to ϕ , but that desire won't serve to construe *Benvolio's pain* as a *moral* reason to ϕ . If the desire to *send Benvolio a card* is going to serve to construe *Benvolio's pain* as a *moral* reason to *send Benvolio a card*, then, it will have to be on account of that desire building into itself a feeling of *Benvolio's pain* directing the agent towards *sending Benvolio a card* in a particular manner. Plausibly, this particular manner of directedness that is needed is the directedness that comes from feeling commanded or authoritatively solicited. But if that is right, then a desire to ϕ will only support moral reasons appreciation if it involves a moral solicitation of the sort I have discussed. In so far as some version of the AFFECT account of moral reasons appreciation can be made to work, then, it will only work if an account of the phenomenal appreciation of reasons like mine is also assumed in the background.⁶¹

⁶⁰ The considerations adduced in this paragraph might also be used to help give an account of the question of “basing relations” in epistemology. For an overview of the relevant issues is Korcz (2015).

⁶¹ One might in principle hold that the pro-attitudes are fundamental and that the phenomenal construal of reasons I discuss is grounded in the having of these pro-attitudes. Or, one might hold that the phenomenal construal of reasons I discuss is fundamental, and that the relevant pro-attitudes are grounded in some kind of responsiveness to reasons so construed. I favor the latter, but it is not integral to my account so far that one does so. I favor the latter because it seems to me that pro-attitudes presuppose, rather than give rise to, representations of reasons. Consider: you have great moral admiration for a moral hero of yours, but then find out that, behind the scenes, he did terrible things. You judge that he is morally bad and unworthy of admiration, but (suppose) your feelings of moral admiration sometimes creep back up on you and linger. Plausibly, when this occurs, you will experience a *cognitive* dissonance not unlike that of a person who sees a Müller-Lyer illusion: though you judge one thing, you will (in your feeling of admiration) also feel a different and opposing vision of the world being given to you. That your feeling of moral admiration creates a cognitive dissonance in this way with your moral judgment about your reasons is evidence that that moral admiration builds into itself a kind of representation of reasons. Gibbard (1992) considers the possibility that moral emotions contain representations of reasons but rejects it on faulty grounds. He assumes that if moral emotions have any representational content at all, that they must be a kind of content that is judgmentally affirmed. He rejects a “cognitive” account of (e.g.) moral anger just because one can make a *reflective*

(ii) A Discursive Account of Appreciation

We've shown the shortcomings of *de re*, causal, functionalist, and affective accounts of ordinary moral reasons appreciation. The only major account of ordinary appreciation that still needs to be considered is the discursive account. If this account also fails to fully explain ordinary moral reasons appreciation, then we will have established strong grounds for thinking that the phenomenal component of moral reasons appreciation is a necessary condition of such ordinary appreciation.

It seems that a purely discursive representation of a consideration as a moral reason is not fully adequate to support appreciation of that consideration as a moral reason. Consider this variation on Huck's story:

DISCURSIVE HUCK: Huck realizes that by helping Jim escape to the North, he is aiding in dispossessing Miss Watson from her property, and so stealing, and he knows there is strong moral reason not to steal. But somehow, in this case, these believed facts have no purchase on him: he doesn't feel the weight of them as considerations bearing on what he should do. He knows that technically he is supposed to report Jim, but at the present moment he just doesn't, in any way, experience that as actually demanded of him.

Despite having a discursive representation of the relevant considerations (regarding property, stealing, Jim, etc.) as moral reasons to report Jim in, Huck does not genuinely seem to appreciate those considerations as moral reasons to report Jim. Plausibly, this is because Huck doesn't have the right sort of phenomenology associated with his discursive representation of those considerations as moral reasons.

judgment about there not being reasons to feel angry, and feel angry anyways (p. 129ff). He does not take into the account that the feeling of anger may involve presentational, rather than, judged, contents. See also Footnote 44 and Footnote 47.

An advocate of a discursive account of appreciation might object by pointing out that on classic accounts of inferentialist representation,⁶² agents only count as having an inferentialist grasp of a concept if they find the inferences that that concept is supposed to underwrite *primitively compelling*.⁶³ If Huck has an inferentialist grasp of a moral reason-giving concept like stealing, then he will find the inferences from (e.g.) *Helping Jim is stealing* to *I have moral reason to report Jim* to *Morally speaking, the thing to do is to report Jim* primitively compelling. And, the advocate of discursive appreciation might claim, if Huck does find these inferences primitively compelling, it looks like he does after all appreciate *helping Jim is stealing* as though it is a moral reasons to *report Jim*.⁶⁴

This objection misses the mark. On the one hand, if Huck finds the conclusion *Morally speaking, the thing to do is to report Jim* “primitively compelling” because he feels the kind of moral soliciting force I have discussed, then maybe the inferentialist is right about Huck, but then of course this inferentialist account of reasons appreciation will not stand entirely by itself as an account of moral reasons appreciation: my phenomenal account will still be a necessary supplement.

On the other hand, and more plausibly, if the inferentialist wants to invoke a notion of primitive compellingness that does without my moral solicitive force, then he will need to offer some other account of primitive compellingness that does the job he wants done. But it looks like the best alternative account of primitive compellingness he could employ will not be up to the task.

⁶² Recall that inferentialist accounts of representation are the accounts of representation that a discursive account of appreciation makes use of.

⁶³ Peacocke (1992).

⁶⁴ Thanks to Laura Schroeter and Geoff Sayre-McCord for setting me on this path of inquiry.

Such an alternative account of primitive compellingness will plausibly characterize “compellingness” in epistemic terms: the relevant kind of primitive compellingness will plausibly be of a sort that makes it the case that when relevant premises are believed, supported inferential conclusions seem true or correct, and on that account also to be believed.⁶⁵ But this will not be enough to support moral reasons appreciation. Consider Silas:

Silas (Discursive): Silas receives and believes expert moral testimony regarding the whole network of inferences and application conditions caught up with moral reason-giving concepts. He has a complete moral theory, and acts as a pure ratiocinator with respect to it. If you present Silas with a consideration, he will crank it through his theoretical algorithms, and tell you whether that consideration constitutes a moral reason to ϕ . Silas even draws conclusions about action; he reports things like *B*: “Benvolio’s pain is a moral reason to stop Tybalt, and that means that, morally speaking, the thing to do is to stop Tybalt.” But the relevant believed facts have no practical purchase on him: he doesn’t feel the weight of them as bearing at all on the practical question of what to do. He feels compelled to believe *B* but doesn’t experience *stopping Tybalt* as in any way practically demanded of him.

Because Silas believes his moral theory to be true and finds the inferences that underwrite *B* primitively compelling, *B seems true* to him, and on that account *to be believed*. Even if this is so, though, Silas still fails to appreciate *Benvolio’s pain* as a moral reason in the way relevant to moral agency, because *Benvolio’s pain* is appreciated (at best) only in its capacity as an *epistemic* reason to believe something, and not in its capacity as a *practical* reason to help Benvolio.

My phenomenal account of reasons appreciation is well positioned to improve on a pure discursive account of appreciation on exactly this point in which it is lacking: if Silas comes to present *Benvolio’s pain* under the light of a felt moral solicitation demanding him

⁶⁵ See Footnote 46 for more on this.

to ϕ , he will feel the weight of that consideration as having practical relevance to the question of what he should do. Like the affective account of appreciation, a discursive account of appreciation also looks to require supplementation from a phenomenal account of appreciation.

We've now seen that each of the major theories of representation of moral reasons seems to require some supplementary reference to phenomenological representation to give a fully adequate account of moral reasons representation. (We saw that causal and functionalist theories of mental representation are insufficient to account for moral reasons representation on their own previously, and in this section have also shown how affective and discursive accounts of moral reasons representation are insufficient to account for moral reasons representation on their own.) Since all of the major theories of representation of moral reasons seem to require a phenomenological supplement, we have given strong presumptive reason to think that the phenomenal representation of moral reasons is a necessary condition of genuine moral reasons appreciation not just in cases of implicit agency, but also generally.

B. Making Sense of Moral Motivation

Judgments about moral reasons have a marked tendency to motivate action. If someone tells me that she judges that she has decisive moral reason to ϕ , I expect that (barring extraordinary circumstances) she will tend to be motivated to ϕ . If she gets new evidence and changes her mind, so as to judge that she does not have moral reason to ϕ , but rather to ψ , I expect that her motivation will track with her new judgment—that she will tend to be

motivated to ψ .⁶⁶ What is it about judgments about moral reasons that explains this tendency they have to motivate action?

On one account—call it the *standing moral desire* account—judgments about moral reasons motivate in fundamentally the same way as judgments about ordinary descriptive facts do: both kinds of judgments motivate if and in so far as they are appropriately paired with standing *de dicto* desires. Thus, my judgment that *My car won't drive* motivates me to bring it to the mechanic because (and only because) I have a standing *de dicto* desire to drive my car, and my judgment that *I have moral reason to ϕ* motivates me to ϕ because (and only because) I have a standing *de dicto* desire to do *the moral thing*.⁶⁷

Standing moral desire accounts of moral motivation have traditionally come under criticism from advocates of the DE RE accounts of appreciation considered earlier.⁶⁸ These critics argue that is not adequate to account for how (at least) ideal moral agents are motivated to ϕ , to say that they have a single generic *de dicto* desire to do what's right, that simply gets channeled through particular, concrete judgments so as to make agents desire to undertake particular ϕ -ings in a derivative way. If moral agents' motivation to ϕ is explained in this way (it is complained) then moral agents will all come out to be objectionably “fetishistic” about abstract morality. As Smith (1994) points out: “commonsense tells us that... good people care non-derivatively about honesty... the

⁶⁶ See e.g. Smith (1994), p. 71–76, for a sampling of the contemporary debate about judgment internalism and externalism. For a more general overview see Rosati (2016), Section 3.2.

⁶⁷ The standing desire account assumes Humeanism in the background. Though the element of the *de dicto* is not always made explicit by proponents, something like the standing moral desire account may potentially be gleaned from, e.g., Railton (1986).

⁶⁸ Arpaly (2002), Markovits (2010), Smith (1994).

well-being of their fellows, people getting what they deserve... and the like, not just one thing: doing what they believe is right, where this is read *de dicto* and not *de re*.”⁶⁹

Advocates of DE RE accounts of appreciation think their approach fixes this problem with the standing moral desire account of motivation: Juliet counts as morally praiseworthy for her efforts to help Benvolio, they say, because she is *directly* motivated by *Benvolio's pain*. We saw reasons in Section One, however, to think that this *de re* “fix” to the standing moral desire account does not work, since DE RE accounts of appreciation give rise to their own kind of problem of moral fetishism. There, we saw that Silas can also be directly motivated by his *de re* contact with *Benvolio's pain*, just by virtue of having a fetishistic preoccupation with *Benvolio's pain*—but Silas is intuitively not praiseworthy in that case. In fact, the problem for the DE RE account is even worse than this. It's not just that it's possible for agents who have a fetishistic preoccupation with morally-relevant descriptive properties to come out as praiseworthy, on the DE RE account of appreciation, it looks like this is how the ideal moral agents *must* be. For if the best kind of moral agent's motivation is supposed to be grounded in an *intrinsic de re* concern with descriptive properties that have moral salience, *qua* their being the kind of *descriptive* properties they are, and not *qua* their being morally salient properties as such, then it looks like that concern not only *can* be, but *must by nature* be grounded in a fetishistic preoccupation with those properties. For what the agent must be concerned about is *really just* the properties themselves—without any moral, aesthetic, practical or other sheen of evaluative significance cast over them. And if there is no such sheen of evaluative significance that

⁶⁹ Smith (1994), p. 75. He writes further: “just as it is constitutive of being a good lover that you have direct concern for the person you love, so it is constitutive of being a morally good person that you have direct concern for what you think is right, where this is read *de re* and not *de dicto*” (p. 76).

the agent is responding to, then it's hard to see how the quality of the concern can be described as anything but a compulsive fetish. Pure intrinsic *de re* concern about an ordinary descriptive property thus fails to make sense of our understanding of the nature of moral motivation.

A third and final kind of account of moral motivation is that offered by *expressivists* (who are broadly aligned with the kind of affective account of appreciation I discussed in Section Four, Part A). On their account of moral motivation, my judgment that *Consideration c gives me a moral reason to ϕ* motivates me to ϕ because (and only because) it is internal to what it is for that judgment to be a judgment about moral reasons, that the judgment presupposes an antecedent vested concern with or care about *c*, that makes me disposed to desire to ϕ on *c*'s account. There are several difficulties for this account of motivation,⁷⁰ but the one I want to focus on is this. On this account of motivation, it is impossible for a subject to genuinely judge herself to have moral reason to (e.g.) *help Benvolio* and yet not be at all disposed to desire to *help Benvolio*. But it seems that such cases *are* in fact possible, at least in principle: it seems that an evil demon, for instance, could genuinely judge herself to have moral reason to (e.g.) *help Benvolio* and yet not have

⁷⁰ One issue with this account that is worth mentioning but that I won't explore in detail is that it leaves us with the burden of having to make sense of how to understand the antecedent vested concern about *c*. Gibbard (2008), for instance talks about a kind of "*brute* care and concern" (p. 91, my emphasis)—but this seems to lend the account to the problems posed for the *de re* account of motivation discussed above. If the relevant concern about *c* is a *brute* concern in the sense that it is a concern that is responsive to *c* purely *qua* ordinary descriptive property, and not responsive to *c*, *qua* reason (represented as a reason), then it looks like the brute care about *c* will again come out as fetishistic preoccupation with *c*. But if it the concern *is* about *c*, *qua* reason, then we already need in hand a prior account of what it is to represent *c* as a reason, that can also account for how that care gets motivated. So grounding the motivational force of judgments about moral reasons in antecedent cares only works if you have some prior story about how those cares are grounded, and that prior story seems to require a prior story about motivation that the expressivist has yet to give us. This provides further motivation for my view. For related worries see Footnote 47 and Footnote 61.

any disposition at all to desire to *help Benvolio*.⁷¹ (That she can judge herself to have a moral reason to help and yet not have any disposition to help is precisely what makes her so evil.)⁷²

We need a way of explaining moral motivation that (i) (contra standing moral desire accounts) *doesn't* require that your desire to ϕ is just derivative of an abstract *de dicto* desire to *do what you have moral reason to do*, but also (ii) (contra DE RE accounts) makes it important that there is something specifically moral about how you construe the considerations you respond to, and last (iii) (contra expressivist accounts) doesn't require that your judgments about moral reasons internally involve a desire to ϕ . Though further work is required to discern the details, a phenomenal account of reasons appreciation appears to be just what we need to satisfy all of (i) through (iii). On the phenomenal account of reasons appreciation, an agent appreciates a consideration as a reason in part by entertaining it under the aspect of a certain kind of felt commanding solicitation to ϕ . This means, first, that (iii) judgments about reasons do not have to internally involve desires. On a phenomenal account of appreciation, an evil demon can consider *Benvolio's pain* under the aspect of a representation that “asks for” or solicits her to help, but just continually

⁷¹ Or so at least judgment externalists will say. I do not try to resolve the judgment internalism/externalism debate here but simply want to make room for an externalist position.

⁷² A further problem for this account is that it leaves us unable to distinguish this case of the evil demon from another case that we should intuitively be able to distinguish from it: that is, the case of the psychopath. Intuitively, the evil demon is *evil* and morally culpable for her lack of desire to *help Benvolio* precisely because she apprehends that (and in this sense judges that) that suffering gives her a moral reason to help, *and just doesn't care*; on the other hand, the *psychopath* is just *blind*—and so *not* morally culpable in the same way—because she, unlike the demon, never begins with that apprehension (or judgment) of having a moral reason to help. But if we assume the expressivist account of motivation there is no distinguishing between the evil demon and the psychopath in this way: since neither of these figures has any disposition at all to desire to *help Benvolio*, neither of them ever genuinely judges that she has moral reason to *help Benvolio*, and so both come out as equally “blind” and not culpable for their non-responsiveness to Benvolio's suffering.

dismiss or refuse to answer that call, and so remain unmoved by this solicitation.⁷³ That said, though feeling a moral solicitation to ϕ does not in itself require you to be motivated to ϕ , it does by its nature summon a kind of response that would constitute such a motivation. A solicitation does after all, *solicit* a response, and to “say yes to” the call of a solicitation to ϕ is naturally understood as constituting or directly giving rise to a motivating desire to ϕ :⁷⁴ becoming motivated to ϕ can be understood just as a matter of “answering” the call that demands you ϕ . This motivation to ϕ will (ii) be responsive to something specifically moral about how you construe the considerations you respond to, but also (i) not need to be derivative of a general *de dicto* desire to *do what you have moral reason to do*. *Benvolio’s* pain itself calls out to you, categorically demanding your concern, and it is this consideration, thus appreciated as a moral reason, that directly and underderivatively gives rise to your motivating desire to help.⁷⁵

Conclusion

What it’s like to be a moral agent in the world is more than what it’s like to feel pain or see red; there’s a way it feels to walk in the space of moral reasons, and if you lack this feeling, you lack something important about that walk. I have tried to show what this feeling is—what, in particular, it is for a moral agent to appreciate a consideration *as* a moral reason. I have argued that such appreciation occurs when an agent presents the relevant

⁷³ She differs from the psychopath on the phenomenal account of appreciation, because the psychopath is simply unable to feel the solicitation from the start. Thus it is that the evil demon is *evil*, and the psychopath merely blind.

⁷⁴ Indeed, the Stoics thought that motivation was just constituted by assenting to a hormetic impression. See Inwood (1985). See also Footnote 38.

⁷⁵ Though I have not had time to go into it here, my account may also explain why moral judgments can sometimes be motivating on their own, in a way that ordinary descriptive judgments are not. For more, see Shafer-Landau 2003, p. 122ff.

consideration via the light of a felt authoritative force “pointing” towards ϕ -ing—lending weight to it, or soliciting it—in a particular way.

I focused for the most part on implicit agents like Huck and Juliet. I suggested that such agents can implicitly appreciate a consideration as a moral reason without explicitly (discursively) representing it to be a moral reason by representing it under the light of the relevant phenomenological mode of presentation. To support that claim, I first showed the implausibility of four other proposals for how to account for implicit reasons appreciation. I showed that we cannot make adequate sense of implicit moral reasons appreciation by understanding it as involving either mere *de re* contact with moral reasons (Section One), or a mental representation of moral reasons whose reasons-content is determined discursively (Section Two, Part A), or by the representation’s external causal “tracking” of moral reasons (Section Two, Part B), or by the representation’s playing the particular kind of internal causal functional role that moral reasons are disposed to play (Section Two, Part C).

In Section Three, I showed how a cognitive phenomenology account of mental representation could plausibly be applied to understand appreciative moral reasons representation. I presented my account of phenomenal reasons appreciation as one that makes sense of *implicit* appreciation in a way that is more intuitive than the other major accounts of implicit appreciation on offer. Finally, in Section Four, I made a preliminary case for the idea that my account of phenomenal reasons appreciation not only captures a critical aspect

of implicit reasons appreciation for agents like Huck, but also gives insight into a more general necessary condition of such appreciation.⁷⁶

⁷⁶ Thanks go to the audience at the 2017 MadMeta Conference (in Madison, Wisconsin) and to the Rutgers Epistemology Dissertation Group (led by Ernie Sosa) for helpful discussion. Special thanks to Ruth Chang and Carolina Flores for their extensive written comments.

Chapter 3

The Ethics of Socially Problematic Disagreements

Introduction

In recent decades, epistemologists have become increasingly concerned about the epistemic significance of disagreement.¹ In particular, they have sought to determine whether and how one's justification for believing p changes when one finds out that an *epistemic peer*—that is, roughly, someone who is equally epistemically competent and who shares your evidence with respect to p ²—forms an opposing belief after assessing the evidence. Epistemologists investigating this question have for the most part attempted to assess the answer without attending to the practical settings that characterize so much of everyday human disagreement. Abstracting away from the human cares and concerns that motivate most of our ordinary anxieties about disagreement, their analysis of how we should respond to disagreement between peers could (for the most part) be applied equally well to understanding how we should respond to getting opposing readings on the temperature from two equally reliable thermometers.³

¹ See Christensen (2009) for an overview. See also Christensen (2007, 2013), Elga (2007), Kelly (2005, 2010), Lackey (2008, 2010), Feldman (2007).

² As I understand it, Gutting (1982) was the first to use the term “epistemic peer.” I have found Lackey (2010)'s characterization of peerhood especially helpful. See also Kelly (2005), p. 11, p.3 n. 3.

³ See e.g. White (2009) for the thermometer analogy. Christensen (2009), p. 196 also makes use of a similar analogy involving watches. Enoch (2010) shares in the spirit of my worry here.

In this paper, I explore how epistemic analyses of disagreement might be informed by closer attention to the practical considerations that bear on so much of ordinary human disagreement. More specifically, my aim is to survey whether, why, and how practical considerations can make a difference to the kinds of doxastic attitudes that parties should take up towards contested claims in the context of particular kinds of socially costly disagreements, which I call *socially problematic disagreements*. In the course of giving a general review of the kinds of worries and concerns that come up in answering these questions, I hope to do two things. First, I hope to show that there is strong reason to think that the added social costs of these disagreements can indeed give parties added reasons for “doxastic adjustment,” where doxastic adjustment with respect to a judgment *p* involves reservation of one’s judgment about *p*, or reconsideration of *p* and inquiry into the question of *whether p*. Second (toward the end), I hope to make it clear that at least some of the reasons parties have for such adjustment are “truth-motivated” reasons that compel them to seek out a greater level of epistemic support for their contested views. In this way—and contrary to what some might think⁴—I aim to show that the social costs of disagreement can sometimes give disagreeing parties’ reason to be *more*, rather than less, concerned about the value of truth.

The argument proceeds as follows. In Section 1, I consider some stock examples of disagreement in the literature on the epistemology of disagreement, and suggest that the

⁴ Rawls (1993) famously suggests that in the context of discussions of liberal political legitimacy, questions of *truth* should be bracketed. “Once we accept the fact of reasonable pluralism,” he writes, “the idea of the reasonable is a more suitable basis of public justification... than the idea of moral truth” (p. 129.) On Rawls’s view, even the political conception of justice he advocates is right for us to accept, not because it’s true, but because it’s reasonable. On his view, “holding a political conception is true, and for that reason alone the one suitable basis of public reason, is exclusive, even sectarian, and so likely to foster political division” (p. 129). As a result, his political liberalism “does without the concept of truth” (p. 94).

practical features of some of these disagreements contribute to the intuitive propriety of doxastic adjustment in response to them. In Section II, I home in on *which* practical considerations seem to make a difference to the enhanced propriety of doxastic adjustment, and make reference to these considerations to help characterize what is distinctive about the class of socially problematic disagreements that are the focus of my study. In Section III I consider a “purely practical” approach to socially problematic disagreements, which suggests that—contrary to the intuitions I have tried to conjure in Sections I and II—socially problematic disagreements can be fully adequately managed *just* by having parties make adjustments to their practical conduct, and so *without* giving parties any added, compelling reasons for doxastic adjustment. I indicate some of the limits of this approach, suggesting that even when we do all we can reasonably do to adjust our conduct towards each other in a socially problematic disagreement, there can still remain significant social costs to disagreement, and these persisting costs continue to give rise to reasons for disagreeing parties to make *doxastic*, and not *just practical*, adjustments. In Section IV, I then canvass two approaches to understanding how the social costs of a disagreement about *p* might generate reasons for doxastic adjustment with respect to *p*: one approach that treats these costs as providing (what I call) “truth-indifferent” reasons for doxastic adjustment, and another approach that treats these costs as providing (what I call) “truth-motivated” reasons for doxastic adjustment. While I won’t deny that relevant social costs can provide truth-indifferent reasons for doxastic adjustment, I nevertheless note the limits of that approach and focus on the way in which the social costs of disagreement can generate truth-motivated reasons for doxastic adjustment. The payoff of this is to show that

the social costs of disagreement can give disagreeing parties reason to be more, rather than less, concerned about the value of truth.

I. Practical Problems for the Epistemology of Disagreement

A. Practical Considerations, Peer Disagreement, and Conciliationism

Does your justification for believing p change when you find out that an epistemic peer disagrees with you? Epistemologists' varied answers to this question fall along a spectrum from *thoroughly conciliationist* to *thoroughly dogmatist*. On one end of the spectrum, *thorough conciliationists* argue that both parties to a peer disagreement should adjust their confidence in p so as to give equal weight to the peer's opposing view; in the typical kinds of cases we will focus on, this means that both parties are epistemically obliged to suspend judgment about whether p .⁵ On the far other end of the spectrum, *thorough dogmatists* hold that neither party to the disagreement need adjust her confidence in p at all.⁶ Though a variety of positions fall between these two ends of the spectrum, I'll simplify by calling those that fall closer to the former end of the spectrum *conciliationists*, and those that fall closer to the latter end of the spectrum *dogmatists*.

⁵ See e.g. Christensen (2007, 2009, 2013), Elga (2007). Kelly (2010) argues—and Christensen mostly seems to agree—that conciliationists should not frame their argument in terms of “on-off” belief states like belief, withholding, and disbelief, but rather in terms of credences. In this paper I work with the categories of belief, withholding, and disbelief for two reasons: first, to simplify discussion, and second, because there is reason to think that much of our practical life is guided by attitudes of belief, withholding, and disbelief about relevant propositions, rather than degrees of credence or degrees of confidence in propositions. Since much of the paper is focused on the practical upshots of our judgments, it seems most prudent to work with the “on-off” doxastic states of belief, withholding, and disbelief rather than with credences. See e.g. Buchak (2014), Ross and Schroeder (2014), and Jackson (forthcoming).

⁶ Kelly (2010) briefly defends this most extreme claim in reference to certain kinds of peer disagreements. Others who defend positions on the dogmatist side of the spectrum include Kelly (2005), van Inwagen (1996), and Plantinga (2000a, 200b).

As in other philosophical debates, conciliationists and dogmatists often leverage readers' intuitions to motivate their arguments. To start, consider these cases that figure prominently in widely-cited papers arguing for conciliationism and dogmatism, respectively (the cases have been renamed for the purposes of the paper):

Social Math: Your friend and you have been going out to dinner together regularly for many years. You always try exactly the same dishes together, tip 20% and split the check. You each do the requisite calculation in your head upon receiving the check. Most of the time you have agreed, but in the instances when you have not, you have taken out a calculator to check; over the years, you and your friend have been right in these situations equally often. This evening seems typical, in that you don't feel unusually tired or alert, and neither you nor your friend has had more wine or coffee than the usual. You calculate that you each owe \$43 and are confident in your result. But then your friend announces that he calculated that you each owe \$45 and that he, too, is confident in his result.⁷

Abstract Opinion: You and an epistemic peer disagree about

- (1) The extent to which a desire to intimidate the Soviet Union played a role in Harry Truman's decision to drop the atomic bomb on Japan in 1945.
- (2) Whether Truman's decision to do so was morally justified.
- (3) Whether there are in fact any truths of the kind that Immanuel Kant called "synthetic a priori."⁸

Christensen (2009) points readers to *Social Math* to motivate conciliationism: he expects that readers will have the intuition that once you find out that your peer disagrees with you about how much you each owe, it is appropriate to become much less confident in your own initial view of the matter. Indeed, he seems to expect that readers will agree that you cannot rationally favor your own initial assessment of the answer over that of your epistemic peer's, and that (as a result) you should suspend judgment on the question of how

⁷ This example is from Christensen (2007), p. 193. It has no title in the original paper but is titled "Mental Math" in Christensen's (2009) paper, where the example is repeated.

⁸ Kelly (2005), p. 167. These cases have no title in Kelly's original paper.

much you each owe.⁹ In his important paper arguing for dogmatism, on the other hand, Kelly (2005) begins by considering *Abstract Opinion*. Kelly seems to think that readers who consider these examples of peer disagreement will be inclined to agree that a peer's disagreement about the correct answer to one of these questions does not in itself undermine the rationality of maintaining your own distinct view about it. As a result (he thinks) you are epistemically authorized to stand steadfast in your belief even after finding out about your peer's opposing judgment.

Suppose for the moment that one has the intuitions Christensen and Kelly expect you to have in *Social Math* and *Abstract Opinion* (respectively). What would account for these intuitions? As with other epistemologists examining disagreement, Christensen and Kelly only make reference to epistemic considerations when they try to account for our differing intuitions about cases. However, the intuitions they appeal to in *Social Math* and *Abstract Opinion* appear to be informed not just by epistemic features of the relevant cases, but also practical ones. Consider the following variants of the cases, which keep the epistemic considerations bearing on disagreement more or less fixed while shifting the practical considerations at stake:

Solitary Math: You are playing an online math game called Mental Math Mania on a Sunday afternoon, just for fun. The game gives players a series of math problems, and tallies and reports their answers and confidence levels to each other. (It also uses reliable video technology to make sure players keep their eyes on the screen while they do mental math.) You are getting towards the end of the game when you are asked to add 20% to a given number and then divide that result in half. You calculate that the answer is 43 and are confident in your answer. You type it in and press enter. Then the game reports that someone else who played the game in the past, who had a similar score and confidence level, got 45.

⁹ See Footnote 5.

Collective Opinion: It is 1948 and Taylor and her partner Tom have been proud and consistent supporters of Harry Truman, who is now up for re-election. They have his political ads on their front lawn and in their car windows, and they frequently work together on and attend political functions for his campaign over the weekends. They are both politically savvy, conscientious reasoners, and they trust each other's judgment. They are listening to the news one night when a reporter raises the question of what Truman's motives were for dropping the atomic bomb in Japan. They agree that if he dropped the bomb largely to intimidate the Soviet Union, that that would be awful, and cause to withdraw political support for him and switch over support to Thomas Dewey. Over the course of the next couple of days they both listen to similar newscasts, and read similar papers and magazines in the chances they have between work and other duties. When conversation on the topic comes up again, Taylor says: "Well, it looks like we have to take the signs down and change our weekend plans. Looks to me like Truman's guilty after all." Her partner looks at her quizzically, "Really? Actually, I had the opposite view. I think he's innocent on this count. I don't know that I want to take down the signs and change our plans just yet!"

Though the epistemic considerations bearing on these disagreements have not been significantly changed from *Social Math* and *Abstract Opinion* (respectively), there is plausibly some shift in the intuitive appropriateness of dogmatism in these cases. To whatever extent one was inclined to think that you are compelled to reserve judgment on your view of the answer (43) in *Social Math*, there is plausibly comparatively *less* of a sense of being so compelled to reserve judgment in *Solitary Math*. After all, in *Solitary Math* it doesn't make any difference to anyone what you think about what the answer to the math problem is, whereas in *Social Math*, you and your friend would presumably like to agree on what price you should each pay. Along the same lines, to whatever extent one was inclined to think that it is suitable for someone to persist unmoved in her commitment about Truman's motives for dropping the bomb in *Abstract Opinion*, there is plausibly comparatively less of a sense of the suitability of such persistent steadfastness of

commitment in *Collective Opinion*.¹⁰ After all, in *Abstract Opinion* it's not likely to make any difference to the believer's life what she thinks about Truman's political motives half a century ago, whereas in *Collective Opinion*, it would be valuable for Taylor and her partner to reach some accord on the matter. Given that epistemic considerations are more or less the same across each pair of cases, it seems that it is these relevant *practical* considerations that relate to the social circumstances of each disagreement that make the difference to the enhanced appeal of conciliation in *Social Math* (as compared to *Solitary Math*) and *Collective Opinion* (as compared to *Abstract Opinion*).

B. From Conciliationism to Doxastic Adjustment

Importantly, not all will have had the intuitions about conciliationism and dogmatism that Christensen and Kelly seem to have expected in the original cases of disagreement offered (*Social Math* and *Abstract Opinion*). For those who didn't, shifts in intuition in these newly altered cases (*Solitary Math* and *Collective Opinion*) may not be obvious—and (in particular) it may not be obvious whether there is really an enhanced propriety to conciliation in the more socially embedded case in each pair (*Social Math* and *Collective Opinion*). This is not too problematic for my purposes, since I won't in the end be concerned to insist that there is an enhanced propriety to conciliation, *per se*—at least, not of the sort that some epistemologists of disagreement have been concerned with, that

¹⁰ Granted, readers with conciliationist sympathies may not find it *wholly* acceptable for the subject in *Solitary Math* to stick to her “4e” answer, and readers with dogmatist sympathies may not find it *wholly* necessary for Taylor in *Collective Opinion* to shift to withholding judgment about Truman's motives. Nevertheless, I expect that it can be widely agreed that holding fast to “43” in *Solitary Math* looks *more* permissible than doing the same in *Social Math*, and that conciliation looks *more* obligatory in *Collective Opinion*, than in *Abstract Opinion*.

involves full-out suspension of one's judgment that p .¹¹ I will however want to suggest that there is at least an enhanced propriety to a weaker kind of conciliatory move, which I will call *doxastic adjustment*. I will take this kind of conciliatory *doxastic adjustment* with respect to p to involve an adjustment to one's doxastic attitudes regarding p , *either* via (on the one hand) suspension of judgment about p , or (on the other hand) at least a judicious reconsideration one's judgment that p and pursuit of equitable inquiry into the question of *whether* p .¹² (I make a note of "equitable" inquiry here in order to bring to attention the idea that relevant inquiry should be unprejudiced in a certain way. While one's initial judgment might carry some weight, I assume that the relevant kind of inquiry isn't aimed simply at finding further confirmation for p , and (more generally) is not guided by a vested, one-sided interest in vindicating that initial judgment.) I take it that what unifies these two kinds of doxastic adjustment is that they both involve a kind of holding open of the question of *whether* p , that is not present in cases where subjects may come to a final and "once and for all" kind of settled judgment with respect to p . I'll sometimes contrast this relevant response of conciliatory doxastic adjustment with an opposing response I'll call *entrenched steadfastness*, where the latter will involve "sticking to one's guns," not just by persisting in belief p , but also by not reconsidering p or equitably inquiring about the question of *whether* p .

By framing discussion in the rest of the paper in terms of *doxastic adjustment* (and *entrenched steadfastness*) rather than in epistemologists of disagreement's terms of

¹¹ Many epistemologists of disagreement discuss norms of belief-revision in terms of norms of credence-revision, and when conciliationists in particular do this they don't necessarily argue that disagreeing peers should always suspend judgment in response to a disagreement. See Footnote 5 on why I focus on suspension of judgment rather than credence revision.

conciliation (and dogmatism) I hope to do three things. First, as I started to indicate above, I hope to win greater intuitive plausibility for the idea that practical considerations might make a normative difference to how we ought to regulate our doxastic attitudes in response to disagreement. I expect that the intuition that there is at least some enhanced propriety to *doxastic adjustment* in the more socially embedded cases of disagreement considered above will come easier to many than the correlate intuition about the enhanced propriety of full-out suspension of judgment in those cases. Even if you don't intuit that you should suspend judgment in *Social Math* (as say Christensen might expect readers to intuit), I expect that there is at least some intuitive force to the idea that there is added reason for you reconsider and inquire about your judgment of "\$43" in the case—an added reason, moreover, that does not pertain equally to *Solitary Math*, and so that does not compel you in a similarly strong way to reconsider and inquire about your judgment there. (After all, as we said before, in *Social Math*, it would be good if you and your friend who are splitting the check could come to the same judgment as to what amount you each owe, whereas it doesn't much matter much what you end up judging in *Solitary Math*. This seems to give you some added reason to at least reconsider your judgment in *Social Math*, that does not pertain in *Solitary Math*.) *Mutatis mutandis* for *Collective Opinion* and *Abstract Opinion*. (Since it would be valuable for you and your partner to reach some accord on the matter of Truman's motives in *Collective Opinion*, there is intuitively more reason for you to at least reconsider and inquire about your judgment about his motives, than there is in *Abstract Opinion*.)

¹² For reasons given later in the paper, I change this slightly. Doxastic adjustment is later characterized as involving *reserving* or *reconsidering* judgment about *p*, and inquiring about *p*.

A second reason for framing discussion in terms of doxastic adjustment rather than conciliation is more theoretical. Some doubt that practical considerations can provide normative reasons for believing, suspending judgment, and disbelieving; since (they take it) truth is the constitutive standard of correctness for belief, practical considerations can at best provide reasons for certain kinds of belief-*management*.¹³ Thus, practical considerations can't serve as normative reasons for me to believe or withhold judgment about *p*, but they can serve as normative reasons for me to expose myself to arguments for or against *p*, which might end up having the effect that I believe or withhold judgment about *p*. Framing discussion in terms of doxastic adjustment allows those with this theoretical concern to get on board with the discussion, since they can agree that practical considerations might provide reasons for subjects to (at least) reconsider and inquire further about a contested belief *p*.

Last and finally, framing discussion in terms of doxastic adjustment allows me to link up to some degree to a broader idea that has important social currency, but has received relatively limited philosophical attention—that is, the idea of *open-mindedness*.¹⁴ Suspending judgment about *p*—or as I'll later put it (for reasons that will come out) *reserving* judgment about *p*—or otherwise, reconsidering one's judgment about *p* and inquiring further about it, is one way of being open-minded about *p*. So, the question of whether and how the practical considerations at stake in a disagreement about *p* can give us reasons for doxastic adjustment with respect to *p* is in some ways the question of whether

¹³ I have been particularly influenced by Hieronymi (2005, 2006, 2009b), but see also Shah (2003, 2006), Thomson (2008), and Whiting (2014). For arguments against this view, see Leary (2017), McCormick (2014), Rinard (forthcoming), and Schroeder (2012).

¹⁴ Baehr (2011) is a recent exception. The topic of open-mindedness has also received some attention in the philosophy of education literature. See e.g. Adler (2004), Gardner (1993), Riggs (2010).

and how those practical considerations can give us reasons to be open-minded with respect to p . I take it that this latter question is important and worth further investigation, and I hope that my discussion throws some light on it. I stick to the language of doxastic adjustment, however, because “open-mindedness” has certain connotations I don’t wish to take on. In particular, I take it that becoming “open-minded” to someone’s view that $\neg p$ can sometimes connote taking a kind of positively sympathetic stance towards $\neg p$, where you come to grant $\neg p$ some considerable degree of credibility. I don’t want to assume that when you doxastically adjust in response to a disagreement where you believe p and your interlocutor believes $\neg p$, that you must necessarily and right off the bat grant that $\neg p$ has a considerable degree of credibility. I want to leave open the possibility that you might doxastically adjust to a disagreement about p (by considering your own judgment p again, and seeking to ascertain what reasons and arguments might bear on the question of *whether* p) even though you regard $\neg p$ as *not* very credible. (As we’ll see in more detail later, part of the motivation for this idea will be that you might think $\neg p$ is very unlikely, but that giving your practical circumstances, considering p again and better ascertaining the reasons that bear on the question of *whether* p (or $\neg p$) is nevertheless worth doing.)

C. The Plan

In the following sections, I will address three kinds of concerns that might now be raised about whether and how the practical considerations in disagreements like *Social Math* and *Collective Opinion* really do enhance the propriety of doxastic adjustment. First, we will want to be clear on just *which* practical considerations in these socially embedded disagreements contribute to the (apparent) enhanced propriety of doxastic adjustment. Second, we will want to ascertain whether and to what extent these considerations really

contribute to the enhanced propriety of *doxastic* adjustment, rather than merely to the enhanced propriety of making *practical* adjustments to one's conduct (say, by agreeing to some sort of reasonable compromise). Lastly, we will want to understand why and how the relevant practical considerations enhance the propriety of doxastic adjustment (assuming that they sometimes do). After all, it's not obvious how practical considerations can or do interact with ordinary epistemic considerations to give subjects reasons to adjust their *doxastic* attitudes rather than just their practical conduct.

In the next section (Section II), I address the first of the three questions above. I discuss two kinds of categories of practical considerations that might make a difference to the propriety of doxastic adjustment in cases like *Social Math* and *Collective Opinion* and use these categories to help define a class of disagreements I call *socially problematic disagreements*. In the section following (Section III) I give a preliminary response to the second question above, where I outline the general kinds of conditions under which doxastic adjustment will be at least an important *part* of an overall choiceworthy response to socially problematic disagreements. Finally, in Section IV, I go on to examine the third question above. I'll ask how practical considerations in socially problematic disagreements give parties to the disagreement reason to make doxastic adjustments, trying to highlight how those practical considerations can serve to have this effect via an enhancement of the importance of getting at the truth.

II. Socially Problematic Disagreements and Doxastic Adjustment

In this section, I'll home in on *which* practical considerations in socially embedded disagreements like *Social Math* and *Collective Opinion* may contribute to the enhanced propriety of doxastic adjustment in those cases. I'll divide the practical considerations that

arise in these disagreements into two categories—those that relate to what I'll call the *cost of error* (on the one hand), and those that relate to what I'll call the *social separation costs* of disagreement (on the other hand). I'll call disagreements where practical considerations related to both of these costs are at stake *socially problematic disagreements*, and clarify three different ways in which disagreement may serve as a reason for doxastic adjustment in the context of socially problematic disagreements.

A. Socially Problematic Disagreements: The Cost of Error

Social Math and *Collective Opinion* are both *socially embedded* disagreements in the sense that, in both cases, there is something about the way the disagreements are embedded in the social-practical life of the disagreeing subjects that seems to make a difference to how they should respond to the disagreement. The mere fact that these disagreements are socially embedded doesn't in *itself* make it the case that there is any enhanced propriety to doxastic adjustment, though. To the contrary, there are ways that disagreements can be socially embedded that may make *entrenched steadfastness* even more normatively appealing than it would otherwise be. Consider *Scholarly Opinion*:

Scholarly Opinion: Tess is a tenured historical researcher in a field that prizes novelty and originality of opinion as a means to establishing a well-rounded, multi-layered understanding of historical phenomena. On the basis of some historical archives, she concludes that Truman's personal desire to intimidate the Soviet Union played a significant role in his decision to drop the atomic bomb. Another tenured researcher working with the archives, Terry, concludes that Truman's personal desire to intimidate the Soviet Union did *not* play a significant role in his decision to drop the atomic bomb. Though Tess and Terry disagree, neither of them considers the other any less epistemically competent or informed than themselves. They know there is a new historical museum going up that will draw to some degree on their research, but only in so far as that research has been thoroughly peer-reviewed and accounted for in several literature review articles that will be completed after they have concluded their own studies. They each go about researching arguments for and publishing on our respective views.

As with *Collective Opinion* (and *Social Math*), *Scholarly Opinion* is embedded in the social life of disagreeing subjects in a way that seems to make some difference to how they may appropriately respond to the disagreement. Here, however, the social-practical circumstances of the disagreement don't seem to add any reasons for the disagreeing parties to engage in any conciliatory doxastic adjustment. If anything, it is the opposite: those circumstances may make it even more permissible for Tess and Terry to take up a kind of entrenched steadfast stance towards their respective views. To even a greater degree than in the original case of *Abstract Opinion*, they may be permitted to stand by their respective judgments and (perhaps even) double down on them, specifically seeking out evidence and arguments that would reinforce the case for their own view, rather than pursuing a fresh, equitable inquiry into the contested question of what Truman's motives were.

What accounts for this difference between *Scholarly Opinion* and *Collective Opinion*? There are two kinds of costs that the subjects in these cases are differently situated with respect to, and, I'll suggest, each of these costs make a difference to the propriety of doxastic adjustment verses entrenched steadfastness. First, there is the *cost of error*: the cost that will be sustained if a subject acts on an erroneous judgment about the matter at hand. In *Scholarly Appraisal*, the costs of error that are likely to be incurred by either researcher—let's focus on Tess for now—are relatively low. "Acting on" a judgment for Tess just amounts to publishing and advocating for views in scholarly forums, and we can suppose that, given the institutional context, there is no significant looming cost to acting on an erroneous judgment in this way. There may be some significant costs incurred if those who later come along and do literature reviews on the topic are not sufficiently

thorough, and so if the museum ends up publicizing an unbalanced and incorrect overall understanding of the historical facts, but (we can stipulate) there is a buffer set up between the actions and judgments of the researchers and those costs of error. We can suppose that the scholarly institution is set up to balance and help screen out mistakes of individual researchers like Tess before these costs are incurred; in fact (we can suppose) the institution is set up so that the truth is best apt to come out, and so the relevant costs are least likely to be incurred, when individual researchers each make the best case they can for a variety of opposing views. Since, as a result, there are no critical *costs* to acting on an erroneous judgment that an individual researcher like Tess needs to be overly cautious to avoid, she doesn't need to be overly concerned to avoid making an erroneous judgment. Not so in *Collective Opinion*. In *Collective Opinion*, there is no similar buffer between subjects and the costs of acting on errors of judgment. It is (we can suppose) important to Taylor and Tom that they politically align themselves with the single best candidate in the presidential race, and if they act on an erroneous judgment about Truman's motives, they will fail to do that. Given these costs of acting on an erroneous judgment, they gain added reasons—not applicable to the researchers in *Scholarly Opinion*—to be concerned to avoid making an erroneous judgment.¹⁵

The enhanced costs of error that pertain to the subjects in *Collective Opinion*, as opposed to *Scholarly Opinion*, plausibly contribute to the enhanced propriety of doxastic adjustment in *Collective Opinion*. If the costs to acting on an erroneous judgment are high, I have (*ceteris parabis*) more reason to be concerned to avoid an error in judgment, and if I have more reason to be concerned to avoid an error in judgment, I will also plausibly

¹⁵ I assume here that our judgments act as guides to action.

(*ceteris parabis*) have more reason to withhold judgment until I can get more information or otherwise (at least) reconsider the truth of my judgment while I seek out such information. Doxastically adjusting in one of these ways will generally help someone better avoid an error in judgment. As a result, the heightened costs of error generally give rise to enhanced reasons for doxastic adjustment.

B. Socially Problematic Disagreements: Separation Costs

There is an additional cost at stake in *Collective Opinion* besides the cost of error that plays a further role in enhancing the propriety of doxastic adjustment in the case. This is what I'll call the *social separation costs of disagreement*. These are social costs that will be sustained in virtue of the persistence of a disagreement in judgments about a contested matter, typically (at least for the purposes of this paper) in the form of the worsening of some practical aspect of the disagreeing parties' respective lives or their social relationship to each other. Once again, we can suppose that the social institutions in *Scholarly Opinion* are set up so that the disagreeing researchers, Tess and Terry, will not experience significant separation costs of this kind. They are tenured professors and, if anything, their disagreement will only make their respective lines of research more noteworthy and their mutual conversation together more enlivening and mutually edifying. Their respective careers and the mutual camaraderie between them will, as a result, not be harmed by the disagreement—if anything, they will be helped.

Once again, however, things are different in *Collective Opinion*. If Taylor and Tom persist in their opposing assessments of the motives and political merits of Truman, social separation costs of various sorts are apt to be sustained. First, whatever benefits there were to their cooperation on political matters will now be severely undercut—for instance,

Taylor can't look up directions to the rally while Tom makes posters anymore; each will have to make time on their own to do both for their own favored candidate. Second, if things are bad enough, the mere cognizance of a disagreement on the matter could wear on their relationship; that itself could give rise to reactive attitudes that weaken feelings of mutual affection and solidarity.¹⁶ Third—and even if things are not so bad—Taylor and Tom's different judgments on the matter may lead them to want to take different actions that will either be mutually impossible to take or at least work at cross-purposes with each other. Tom may fervently want the Truman advertisements in the house windows and on the car to be kept up, while Taylor fervently objects to them, and wants to switch them to advertisements for Dewey. Taylor needs the car to get to Dewey's rally, but Tom wants it to get to Truman's rally, and so on. That their actions will work at cross-purposes with each other in these ways means it will be impossible or difficult for both of them to enjoy full self-determination with respect to their different judgments (so long as these judgments conflict). And that in turn may have further negative effects—e.g. even if negative reactive attitudes didn't exist before, they are now easily apt to develop.¹⁷ Though all these kinds of separation costs of disagreement are important, I'll for the most part focus on this last cost (with its associated effects)—that is, the cost of disagreement for parties' *full mutual self-determination* (I'll sometimes drop “full” and “mutual” for brevity). I'll understand full mutual self-determination broadly as the ability of both parties to a disagreement to do what they each think there is most reason to do, *given* their own assessment about the truth

¹⁶ In saying this I draw in part on psychological literature on the effect of value-congruence on trust. See e.g. Cvetkovich, George, and Nakayachi (2007), Edwards and Cable (2009), and Siegrist, Cvetkovich, and Roth (2000).

¹⁷ See e.g. Haidt, McCauley, and Imada (1997), Tetlock (2003), Tetlock et al. (2000), and Tybur, Lieberman, and Griskevicius (2009).

of some underlying contested matter (whatever that is), but *setting aside* the new difficult practical considerations that arise from needing to negotiate through the disputes. If Taylor and Tom could each act on their own assessment of what there is most reason to do in this sense, Taylor would keep up only Dewey posters and use their car to attend Dewey rallies, while Tom would keep up only Truman posters and use their car to attend Truman rallies.¹⁸

The addition of separation costs (like the cost of mutual self-determination) to a disagreement plausibly contributes to the enhanced propriety of doxastic adjustment in response to it. There are two ways in which it might do so, but I focus on the more straightforward one for now. (I'll discuss the other one later, in Sections Three and Four). Since separation costs are incurred at least in part by virtue of the persistence of disagreement, it seems that to the extent that someone has reason to avoid these costs, she also inherits added reasons to try to resolve the disagreement that helps give rise to those costs. If she has added reason to try to resolve the disagreement, she also plausibly has more reason *not* to remain steadfastly entrenched in her initial judgment—since this will only make the disagreement more enduring. Rather, she plausibly has added reason to reserve judgment until she can attempt some inroads into resolving the disagreement, or otherwise (at least) some added reason to reconsider the truth of her underlying contested judgment while she seeks out such inroads to resolution. Doxastically adjusting in one of these ways will plausibly better help parties to a disagreement better resolve the disagreement—at least on the assumption that there are inroads to be made on such

¹⁸ *Mutatis mutandis* for Social Math: the good of full mutual self-determination is at stake because, so long as disagreement persists, circumstances will not permit you both to pay the amount you think you have most reason to pay.

resolution (I will discuss this further in Section Two)—and so heightened separation costs seem to be able to rise to enhanced reasons for doxastic adjustment.

We can start to test this idea that the separation costs of disagreement enhances the propriety of doxastic adjustment by considering this variant of *Collective Opinion*, that keeps the costs of error (previously discussed) more or less fixed but changes the relevant separation costs of disagreement:

Personal Opinion: The situation is as in *Collective Opinion* but this time Taylor lives alone and her disagreement is with an office mate. She takes him to be as politically savvy and conscientious as she is, and she regards his judgment as being generally as dependable as her own. As they are talking one afternoon, she hears that he thinks that Truman's motives were not seriously influenced by a desire to intimidate the Soviet Union. He has also apparently been listening to similar newscasts, and read similar papers and magazines.¹⁹

Here, unlike in *Collective Opinion*, there is nothing at stake for Taylor and the person she disagrees with that depends on whether their disagreement persists; they can each go on in their way freely without any impediment that arises from having different judgments on the matter of Truman's motives. As a result, even if one thinks that there is some normative pressure for doxastic adjustment in *Personal Opinion*—in any case more pressure than in *Abstract Opinion*—that pressure is plausibly weaker than it is when Taylor finds herself at odds with her partner, Tom, in *Collective Opinion*. This is just what we would expect if, as suggested above, not just the costs of error but also the social separation costs of disagreement gives one added reasons for doxastic adjustment.

¹⁹ Here is a similar variant of the *Social Math* case:

Personal Math: You go out for dinner by yourself, but while you're there you happen to see your friend from *Social Math*. You get seats next to each other and eat together, but have separate bills. You are calculating the check at the end. You always tip 20% and you conclude that with that, your bill comes to \$43. You are confident in your answer. Your friend happens to see your calculations and mentions that he thinks you got it wrong—he's pretty confident that with that tip percentage, you owe \$45, not \$43.

C. Disagreement as a Practical and Epistemic Reason for Doxastic Adjustment

I'll term disagreements where both costs of error and separation costs are at stake *socially problematic disagreements*. To close this section, I will review three different ways in which a consideration—such as the fact of disagreement—might constitute a reason for doxastic adjustment in a socially problematic disagreement: first, as an epistemic reason; second, as a practical reason related to social separation costs; and third, as a practical reason related to the costs of error. I specify which of these ways in which disagreement can serve as a reason for doxastic adjustment will be most centrally significant for our purposes (the second), and also point to an important consequence that follows from this way in which disagreement can serve as a practical (rather than just an epistemic) reason for doxastic adjustment.

First, a consideration can figure as a reason for doxastic adjustment with respect to p by serving as an ordinary *epistemic* reason to suspend judgment about p : it can serve as an indicator of the (possible) falsity of p , and so act as a defeater to your epistemic justification for believing p . Disagreement can serve as an epistemic reason for adjustment (in particular, for suspending judgment) in this way. When it does, the extent to which it does so will vary more or less directly with the epistemic credentials of one's interlocutor. To the extent that someone is your epistemic *inferior* with respect to p , his disagreement with you about p will give you a *less weighty* epistemic reason to adjust with respect to p (in particular, it will give you *less* reason to suspend judgment about p), and to the extent that someone is your epistemic *superior* with respect to p , her disagreement with you about p will give you a *more weighty* epistemic reason to adjust with respect to p (in particular, it will give you more reason to suspend judgment about p , or perhaps some reason to switch

to judging that *not p.*) Of course, if someone is your peer, your epistemic reasons for adjustment will fall somewhere in the middle.

The second and third ways in which a consideration may be able to serve as a reason for doxastic adjustment in socially problematic disagreements are practical. On the one hand, a consideration may be able to serve as a practical reason for doxastic adjustment by *giving rise to (or aggravating) separation costs* whose adequate management requires doxastic adjustment.²⁰ Of course in socially problematic disagreements, disagreement does give rise to separation costs (socially problematic disagreements are defined in part by this feature). So, thus understood in its relation to the incurrance of separation costs, disagreement may be able to serve as a *practical* reason for doxastic adjustment. It may give you practical reasons for doxastic adjustment, that is, by serving as a harbinger of relevant social separation costs.

Last, a consideration may be able to serve as a practical reason for doxastic adjustment by *giving rise to (or aggravating) the costs of acting on an erroneous judgment*. Thus, the consideration that *I may get behind the wrong presidential candidate* may be able to serve as a practical reason for Taylor to doxastically adjust with respect to her initial judgment about Truman's motives (as illustrated in *Personal Opinion* and *Collective Opinion*).²¹ It is not obvious off the bat that the fact of disagreement itself can heighten the cost of error and so serve as a practical reason for adjustment in this last way. In the end though, that idea will be very significant to the part of my argument where I argue that parties to a socially

²⁰ By analogy: my falling in the pool serves as a practical reason for the lifeguard to jump in by giving rise to costs whose adequate management requires her to jump in.

²¹ Because a consideration that heightens the cost of error makes it more important that Taylor avoids an erroneous judgment, that consideration can apparently give her added reason to suspend judgment, or at least reconsider and inquire further about her judgment.

problematic disagreement should be more, rather than less, concerned with the truth. To make that point, I will ultimately argue that in socially problematic disagreements, disagreement can raise the cost of error *by* raising separation costs.²²

It is in its relation to social separation costs that disagreement makes its most distinctive contribution to reasons for doxastic adjustment. After all, in its capacity as an epistemic reason for adjustment, disagreement serves as just one among many potential epistemic defeaters for justified belief. Further, there are a great variety of ways in which the costs of error may be raised for a subject, and so a great variety of kinds of considerations that may serve as a reason for her to doxastically adjust by way of raising the cost of error. On the other hand, separation costs are defined in part by their relation to disagreement, as social costs that are incurred in virtue of the persistence of a disagreement. As a result, if we want to determine what is distinctive about the way that socially problematic disagreements give rise to reasons for doxastic adjustment, we will need to focus on the way in which separation costs give rise to reasons for doxastic adjustment. And that is precisely what I will do. From here on out, I will be chiefly interested in the significance of disagreement for doxastic adjustment as it is considered in this capacity, as a practical reason for adjustment related to the incurrence of social separation costs.

²² It is worth making two clarificatory points here. First, the extent to which a consideration can serve as a reason for doxastic adjustment about *p* by raising the costs of error might very well depend on what kinds of ordinary epistemic reasons a subject already has to believe *p*. If she already has absolutely conclusive evidence about the truth of *p*, and there is not any doubt about her flawless evaluation of that evidence that can reasonably be raised, then perhaps a consideration that heightens the cost of error about *p* does little or even nothing by way of giving her reason to doxastically adjust with respect to *p*. For the purposes of the argument in this paper, though, I'll mostly assume that these conditions do not pertain, and so that a consideration that heightens the cost of error does give on *prima facie* reason for adjustment. A second note to make is that epistemologists who advocate for *pragmatic encroachment* think that the cost of error that is at stake for a subject when she considers a (prospective) judgment *p* can bear on whether she is epistemically justified in believing *p*. If they are right, then considerations that relate to the cost of error might also be styled

D. Upshots of Practical Considerations for *Non-Peer* Disagreements

There is one final important point to make about the upshots of the idea that disagreement may serve as a practical reason for doxastic adjustment. In so far as disagreement serves as a practical reason for doxastic adjustment, the strength of the reasons for doxastically adjusting in response to a disagreement will *not* necessarily be correlated with the strength of the epistemic credentials of your interlocutor. Even when the epistemic credentials of your interlocutor would not be enough on its own to give you reason to doxastically adjust, once you find oneself in a socially problematic disagreement with that interlocutor—and so the disagreement takes on added significance as a practical reason for adjustment—you might nevertheless find yourself with reasons for such adjustment. Focusing on disagreement as a practical consideration attached to separation costs, consider the difference between *Solitary Concert* and *Social Concert*:

Solitary Concert: You are going to an orchestra concert you’ve been looking forward to for a long time—but you left your phone in the trunk and so can’t look up directions on it. You came just a couple weeks ago to the same venue and are relatively confident that you remember needing to take a right at the big intersection just ahead. You decide your memory is clear enough for you to settle that you do in fact need to take the right, and put your signal indicator on to turn. Just then you see that the old in car GPS system says to take a left. You don’t use the system much because it’s had some trouble updating maps and you’ve noticed that now you’re right more often than it is when you disagree. “That’s fine,” you think to yourself, “No need to be worried about what that old system says. I trust my memory. I can settle that it’s a right here without needing to stop to check any further.” You judge that you should take a right.

Suppose your memory is functioning well and you remember correctly; further, the old GPS is in fact unreliable (as you thought). If so, then (plausibly) you have enough epistemic support for your judgment that *(R) I should take a right*, to justify you in fully

as epistemic reasons of a sort. For my purposes, however, I’ll keep referring to these considerations as

settling on *R* as your “final answer.” You are justified in remaining steadfastly entrenched in your judgment that *R*, without having to further reserve judgment about *R* or reconsider and inquire about the possibility that *not R*. But now consider whether your epistemic support for your judgment *R* does equally well at giving you justification for steadfast entrenchment on *R* in the same way in a socially problematic version of the same disagreement, where the machine is replaced by your father:

Social Concert: The situation is as before except that the old in car GPS system is off and you are going to the concert with your father, who has also anticipated it and is sitting next to you in the passenger seat. He sees you are getting ready to turn right and tells you that that’s not the way you to go—he was here a couple weeks ago (around the same time you were) and (he says) you should take a left. You’ve noticed you are right more often than he is about directions when you disagree. “That’s fine,” you think to yourself, “No need to be worried about what dad says. I trust my memory. I can settle that it’s a right here without needing to stop to check any further.” You judge that you should take a right.

Plausibly, you are not justified in remaining steadfastly entrenched on *R* in *Social Concert* in the way that you very well might have been in *Solitary Concert*. If you were so justified, then the only compelling reason you would have for stopping the car would be to humor your father—you would have enough epistemic support to settle that you are right, and so you would be justified in closing the question of *whether p* and only engaging in any further “inquiry” about it with your father as a polite pretense. But plausibly there would be something amiss in responding to the disagreement with your father in this way; you appear to now have more reason to reserve your judgment or to at least reconsider it and genuinely inquire further about it, than you did in *Solitary Concert*. This is precisely what our discussion above would predict, since now the reasons the disagreement gives you for doxastic adjustment are not just epistemic but also practical in nature. (In particular, part of

practical reasons throughout.

what seems to be going on here is that you and your father are linked in a way that puts your mutual self-determination at stake: so long as you have different judgments about which way to go, one of you—probably your father in this case—will have to be dragged along into something he judges to be not optimally choiceworthy. And so you have added reasons to doxastically adjust that go beyond the weak epistemic reasons for adjustment associated with the disagreement.)

This result has an important upshot for our analysis of socially problematic disagreements. We started by considering cases of disagreement that epistemologists of disagreement have focused on, which were all cases of *peer* disagreement. But now we have seen that when disagreements are socially embedded in the way *socially problematic disagreements* are, the epistemic credentials of your interlocutor are not necessarily crucial to determining whether you should doxastically adjust. There may be compelling reasons for such adjustment even when you disagree with someone who is your epistemic *inferior*, on account of your *practical* relationship to them.²³ As a result, from here on out my discussion of the importance of doxastic adjustment can be taken to apply not just to socially problematic disagreements between epistemic *peers*, but to socially problematic disagreements more generally.

III. The Purely Practical Approach

I have now surveyed *which* considerations at stake in socially embedded disagreements may contribute to an enhanced propriety of doxastic adjustment. I divided the relevant

²³ This relates back to a point made in Section One about the distinction between doxastic adjustment and open-mindedness. You do not have to see your father's claim as being very credible at all to have reason to doxastically adjust in response to it. Some may think, however, that to be "open minded" about his claim, you do have to accord it some significant degree of credibility.

practical considerations into two categories—those that relate to the *separation costs* of disagreement and those that relate to the *cost of error*—and pointed out that disagreement can potentially serve as an added practical reason for doxastic adjustment by way of its relationship with either of these costs. I said I would focus on the former kind of cost (separation costs) for now and have not gone into much detail with respect to latter kind of cost (costs of error). In the next section, I'll examine *how* practical reasons related to both sorts of costs might interact to give rise to reasons for doxastic adjustment. Here, however, I first stop to address an important concern about *whether* relevant practical considerations really do add reasons for doxastic adjustment. In particular, I query the extent to which the *social separation costs* of disagreement really can give subjects added reasons to make *doxastic* adjustments rather than merely *practical* adjustments to conduct (like, e.g., acting in accord with the terms of a reasonable compromise). To the extent that the separation costs of disagreement can be fully adequately managed by practical adjustments to conduct (like compromise), after all, those costs will not be poised to also give much additional reason for *doxastic* adjustment.

I don't give a full account of necessary and sufficient conditions for when separation costs contribute to compelling reasons for doxastic adjustment. Rather, I here try to provide a review of the kinds of parameters that will make doxastic adjustment more or less important in light of the practical considerations related to separation costs, and on the assumption that parties could also make practical adjustments to conduct. This will leave us with a clearer view of the range of cases in which the practical aspects of a disagreement do plausibly make an all-things-considered difference to the propriety of doxastic

adjustment. I will then focus attention on understanding *how* those practical aspects of disagreement make that difference in the sections following.

A. The Critic: A Purely Practical Approach to Socially Problematic Disagreements

A critic of the idea that social separation costs can give parties to a disagreement significant added reasons for doxastic adjustment might argue as follows. It was suggested that disagreement can serve as a practical reason for doxastic adjustment in so far it incurs social separation costs, but the discussion of how such separation costs are incurred was overly simplistic. In particular, it was pointed out that Taylor and Tom's disagreement about Truman's motives incurs separation costs (and so generates reasons for doxastic adjustment), in so far as it undermines their ability to mutually enjoy full self-determination. But the significance of that cost was overblown, and the counterweighing practical costs that are incurred by doxastic adjustment were left entirely unaccounted for. Once both of these things are taken into account, it becomes apparent that a great deal of the time, the separation costs incurred by a disagreement give parties to the disagreement *only* reasons to make *practical* adjustments to their conduct, and not also to *doxastically* adjust in response to it.

First, a critic might say, the extent of the separation costs incurred by socially problematic disagreements like Taylor and Tom's were overblown. True, so long as they disagree, there will be *some* cost to their *full* mutual self-determination; nevertheless, that cost can be suitably managed without their having to make any adjustments to their doxastic attitudes. This is because the extent of the separation costs incurred by a disagreement is determined not just by whether parties to the disagreement remain entrenched in their original contested beliefs, but also by *how they treat* those contested

beliefs in their practical reasoning about what to do together now that they find themselves in a socially problematic disagreement. If Taylor and Tom are reasonable, then they will believe that in the context of a socially problematic disagreement like their own, one should bracket or discount the salience of one's own contested judgment about the truth as one engages in practical reasoning, and instead give priority to finding a mutually acceptable course of action.²⁴ If Taylor and Tom are reasonable in this way (the critic may continue), then they will find some compromise that they can each deem reasonably choiceworthy (given the circumstances), and—since they each deem the compromise reasonably choiceworthy—neither will be objectionably coerced into some outcome they object to. So each will act in an adequately self-determined way after all when they resolve to act by the terms of that compromise. Thus, for instance, perhaps Taylor and Tom remain steadfastly entrenched in their contested judgments about Truman's motives, but agree to take down the political signs from their front lawn and car, and trade off who gets to use the car each weekend for their favored political event. Granted, neither gets his or her ideal outcome, but each still exercises a fully adequate degree of self-determination. (However the details of “adequacy” work out here, the degree of self-determination that Taylor and Tom enjoy is not one about which they can reasonably protest.²⁵)

²⁴ A problem with this view that I don't get to address later is that, even if you can discount the salience of your own contested judgment about the truth as you engage in practical reasoning, it is arguably much more difficult to discount the salience of your judgment in the formation of your reactive attitudes. Reactive attitudes often end up having social consequences as well, and so discounting one's judgments only in the way proposed by the critic will arguably not do enough to stave off all the important social costs of maintaining those judgments.

²⁵ After all, the critic might continue, this is the real world and disagreements like this are bound to arise, even between reasonable, conscientious people like Taylor and Tom. Lamenting too much about the fact that we don't get our own way all the time is childish. This contention of the critic's bears some resemblance to the contention of Rawls (1993), when he points out that we should see the fact of reasonable pluralism “not as a disaster but rather as the natural outcome of the activities of human reason under enduring free institutions” (p. xxiv-xxv).

Further, the critic might continue, the discussion suggesting that separation costs may give parties to a socially problematic disagreement compelling reasons for doxastic adjustment left out the fact that doxastic adjustment may itself often incur significant costs. When it does, that makes settling for a reasonable compromise all the more choiceworthy as the uniquely best response to a socially problematic disagreement. Given that doxastic adjustment may bring with it its own kinds of costs, a reasonable compromise is not just an *adequate* response to a socially problematic disagreement, it is the single *best* response to the disagreement that parties to the disagreement can take.

The costs generated by doxastic adjustment depend on the kind of doxastic adjustment at issue. Consider first what happens if parties to a socially problematic disagreement settle on *suspending judgment* in response to it (rather than reconsidering or pursuing further inquiry). To begin, at least some cases, the suspension of judgment itself will be experienced as a loss. Perhaps Taylor's identity is partially caught up in the way she is engaged in politics, and if she now remains neutral about who the best political candidate is she will experience some loss of purpose and self. More generally—and whatever one thinks about Taylor's example—consider such socially central disagreements as disagreements about religion. Disagreements about persons' religious commitments figure very prominently in our social world, and they are often socially problematic (they incur social costs of various sorts, including threats to people's full mutual self-determination).²⁶ But if believers respond to the separation costs of disagreement by suspending judgment

²⁶ Disagreement about religious issues sometimes comes at the cost of full mutual self-determination when, for instance, their different religious viewpoints compel different kinds of mutually incompatible public expressions of faith regarding e.g. public prayers and statues, curriculum conduct in public schools, and various conflicting rules for social morality.

about their beliefs, that will come as a great existential loss for them. An even more general problem is that suspension of judgment can easily have costly consequences for action.

Thus, suppose Taylor suspends judgment about Truman's motives. Since (suppose) this is a critical issue in the election, she also ends up suspending judgment about who the better candidate is—so she remains neutral about who to vote for. This neutrality could be costly, especially if she had in fact made the right judgment about the evidence originally.²⁷

Perhaps now she'll fail to vote, or her indifference will be taken as a sign by her friends, who trust her judgment as the most politically savvy person among them. Especially since it is being supposed that *part* of what is at stake in a socially problematic disagreement like Taylor's is that there is some cost of error, there is plausibly also some value to her positively getting at, and acting on the premise of, the *right* judgment about the facts. But if she suspends judgment about *p*, she will fail to do that. Responding to a socially problematic disagreement by suspending judgment, then, can have serious costs both with respect to a subject's experience of her subjective identity and with respect to the costs of neutrality.

Now consider the costs of doxastic adjustment when it involves *reconsideration and inquiry* rather than suspension of judgment. First, for some, reconsideration of their cherished beliefs may be nearly as distressing as suspension of judgment about those beliefs. Further, one must consider the possibility that inquiry may not lead anywhere—in fact, it may sometimes make things even worse between disagreeing parties. A great deal of socially problematic disagreements will likely revolve around religious, moral, and political issues that there has already been long and persistent disagreement amount, and in

²⁷ From within the ranks of epistemologists of disagreement, this kind of worry is most likely to be pressed by someone like Kelly (2010).

these cases it's hard to see how separation costs could give compelling reasons for further inquiry, since it seems such inquiry won't do much to resolve the disagreement.²⁸ If it doesn't, that presents significant opportunity costs, since the time and resources spent on inquiry could have been better used elsewhere. Further, there's reason to think that inquiry may sometimes only *exacerbate* the social separation costs associated with a disagreement.²⁹ Thus, when Taylor and Tom start looking further into Truman's motives for dropping the atomic bomb and debating his fitness for office, they may only become more convinced of the inadequacy of each other's case, and more passionate about the superiority of their own judgment. So responding to a socially problematic disagreement by reconsidering and further inquiring about contested judgments may do little to help, and even some to *hurt*, the social separation costs associated with that disagreement.

Given the costs of doxastic adjustment, and the benefits of a purely practical compromise, the critic might now conclude, it will often be best that parties to a socially problematic disagreement just settle for a purely practical resolution of their disagreement—they should agree to a reasonable compromise but may remain steadfastly entrenched in their own beliefs. Once all things are properly considered, the fact that a socially problematic disagreement incurs separation costs does not make for an *overall* compelling reason for those parties to make doxastic adjustments.

²⁸ Public reason liberals like Rawls (1993) will argue on this basis that we need not pursue further inquiry about disagreements bearing on comprehensive visions of the good. Rawls suggests that the *only* way to generate agreement on a comprehensive vision of the good is via the oppressive use of state power and so, not by means of free discussion and inquiry (p. 37).

²⁹ See e.g. Sunstein (2002, 2007). But on the other side see Mercier and Landmore (2012).

B. The Need for Doxastic Adjustment in the Domain of Reasonable Compromise

I will offer two kinds of response to the critique offered above. First, in this subsection, I'll consider cases where (as the critic points out) there is a practical compromise that parties to a socially problematic disagreement can reasonably be expected to mutually agree to. I'll say that cases of this sort fall in the "domain of reasonable compromise," and grant that in these cases, subjects should generally agree to such a compromise if and in so far as their disagreement persists. Still, I'll outline some conditions under which there are weighty reasons for those parties to doxastically adjust *as well*, so as to see whether they can in the end do better than *just* settle for that compromise. Then in the next subsection, I'll point out that there are cases of socially problematic disagreements where the costs of error are of such a weighty kind that a party to a socially problematic disagreement may not be justified in agreeing to a practical compromise in response to it. I'll say that cases of this sort fall in the "domain of justified coercion." In this domain, I'll suggest, the purely practical approach does little to help address the costs of socially problematic disagreement, and so doxastic adjustment is the only way those costs can be effectively addressed, if indeed they can be effectively addressed at all.

Let's begin with the cases of socially problematic disagreement that fall in the domain of reasonable compromise. The advocate of the purely practical approach suggested that because (on the one hand) (i) making *practical* adjustments in response to socially problematic disagreements can often recoup some of the separation costs of disagreement, and (on the other hand) (ii) making doxastic adjustments itself has costs, subjects in socially problematic disagreements will often do best just by settling on a reasonable practical compromise, without worrying about making any doxastic adjustments to their contending views. I address (i) and (ii) in turn and conclude with a rough characterization

of the conditions under which doxastic adjustment will plausibly be *part* of disagreeing parties' best overall response to a socially problematic disagreement in this domain.

(i) “Left Over” Separation Costs in Reasonable Compromises

Suppose we confront a case of a socially problematic disagreement in which there is a reasonable compromise forthcoming that manages to mitigate some of the social separation costs of disagreement—parties to the compromise can enjoy some adequate level of self-determination even though they lose out on full self-determination. What we want to ascertain now is what determines the extent of the social separation costs that may be left over after such a reasonable compromise is made, since it is the extent of these costs that persist as a remainder after compromise that may go on to make a difference to whether parties should also doxastically adjust in response to the disagreement. The core variable that I will focus on as important to determining the extent of these “left over” separation costs in the context of a reasonable compromise is the variable of the costs of “parallel processing.” Most simply, these are the costs that are accrued to a party to reasonable compromise when she must keep track of and act on the basis of two different sets of action plans in two different practical territories: one set of “personal” action plans that are predicated on the assumption that her (contested) beliefs are true and serve as a guide to action in her private life, where her conduct doesn't affect the other party and so is presumably not affected by the terms of the compromise, and a second set of “public” action plans that don't assume the truth of her contested beliefs and serve as a guide to action in her “public” life, where relevant conduct would objectionably affect the disagreeing party, and where the terms of the compromise are presumably applicable.

The extent to which there are social separation costs of a disagreement “left over” after a reasonable compromise will depend on both the *depth* and the *breadth* of these costs of parallel processing. First, their *depth*: the more Taylor and Tom’s political commitments are deeply central to their identity or what they care about, the more difficult and costly it will be for them to keep their personal and public action plans separate. So, for instance, suppose Taylor and Tom agree to a compromise whereby they take down any signs of political patronage on property in which they have joint ownership (the house, the car, and so on) and trade off who gets the car and other home resources to work on political events. To the extent that their political commitments are deeply central to their identity and what they care about, they will find it more difficult and costly to abide by the terms of the compromise.³⁰ A second factor that will determine how costly parallel processing is for Taylor and Tom is the *breadth* of cases in which private and public action plans must be separated: the more often Taylor and Tom’s interactions bear on political matters where compromises must be made, the more difficult and costly it will be for them to hold up to the requirements of a compromise. Thus, contrast Taylor’s relationship with Tom and her relationship with her officemate in *Personal Opinion*. Her officemate also disagrees with her about Truman but their interactions rarely bear on politics. In this case, even if the instances in which Taylor must compromise are difficult for her because of the depth of her commitment (perhaps she can’t speak freely about politics at lunch the way she really would like to be able to), those difficulties may be isolated enough from each other to be easily sustained. On the other hand, since a great deal of the interactions between Taylor and Tom do bear on politics, the breadth of cases in which compromises must be made will

³⁰ There is a catch twenty-two here, since these issues about identity are part of what would also motivate people to want to compromise rather than doxastically adjust.

contribute to the difficulty of compromise.³¹ As these two kinds of costs of parallel processing go up, practical adjustments like reasonable compromise will be comparatively less effective at addressing the separation costs of a socially problematic disagreement. To the extent that reasonable compromise becomes less effective at managing those separation costs, moreover, those separation costs will increasingly serve in the capacity we suggested that they might serve in Section Two—that is, as reasons for doxastic adjustment aimed to address the underlying contested judgments that give rise to the disagreement with its associated social costs.

(ii) The Costs of Doxastic Adjustment

I've suggested that as the costs of parallel processing grow, *all else equal*, there are more separation costs that attend reasonable compromises, and so more (*prima facie*) reasons for parties to doxastically adjust in response to the disagreement. But as the advocate of the purely practical approach pointed out, not all else always *is* equal, because the costs of doxastic adjustment must also be accounted for. Even if the costs of parallel processing are relatively high in the case of a particular reasonable compromise, the costs of doxastic adjustment may be higher (or just as high)—and if that is so, then there will still not be any compelling overall reason for parties to make doxastic adjustments. It will be better for them (or just as good) to agree to the reasonable compromise and just live with the separation costs associated with parallel processing.

³¹ Or consider again the example of check splitting in *Social Math*. The parties to this disagreement might very easily agree to a compromise if it is a *one time* compromise and the dollar amount difference between their estimates is not very much, so that they don't care much about the difference between what they think they should pay, and what they ultimately do end up paying. But if the difference between their estimates differs more, or the same kind of problem continues to arise again and again when they go out, they may gain more and more reasons *not* to just compromise but rather to look into what if anything is going wrong with their respective bits of reasoning.

Let me respond first to the critic's worries about the costs of suspension of judgment. Not all kinds of suspension of judgment are the same, and in so far as there is reason for parties to a socially problematic disagreement to adjust by suspending judgment, I'll grant that the critic's worries point to the preferability of a kind of moderated, "unsettled" suspension of judgment I'll refer to as *reservation of judgment*, over a hardline, "settled" neutrality. First, even if Taylor does suspend judgment about Truman's motives or who the better candidate for president is, she may nevertheless sometimes act on the basis of her best estimation or guess about the matter (so she may go out and vote and campaign for Dewey, if she thinks he is *probably* the better candidate). In this way, her suspension of judgment may be moderated, and not give rise to total neutrality of action. Second, if Taylor suspends judgment about Truman's motives, that suspension of judgment needn't be a *permanently settled* one. She need not simply throw up her arms and give up forming a judgment on the matter; she may remain actively engaged in inquiry aimed at re-establishing a definite view on the topic. For the purposes of argument from here on out, I will assume that if and in so far as suspension of judgment is a choiceworthy response to a socially problematic disagreement, it will generally be moderated in both these ways. Since I'm particularly concerned with subjects' doxastic responses to socially problematic disagreements, I'll make special note of the assumption that suspension of judgment should generally involve actively engaged inquiry, and so (from here on out) talk about *reservation of judgment* where I previously made reference to *suspension of judgment*. Even if there would be significant costs to someone like Taylor's responding to a socially problematic disagreement by suspending judgment in a hardline, settled way, those costs

will be reduced if she instead merely *reserves* judgment while pursuing further inquiry and (when appropriate) acting on her best guess about the truth.

At this point the critic will rightly press further on the costs of inquiry he pointed to above. Since I'm now assuming that doxastic adjustment with respect to p will always include inquiry about p , the costs of inquiry must figure significantly into any assessment of the propriety of doxastic adjustment as a response to socially problematic disagreements. I grant the critic that the costs of inquiry may sometimes be significant—that inquiry may sometimes, first, not lead to any resolution of disagreement and so incur opportunity costs and, second, aggravate the social separation costs of disagreement by further polarizing disagreeing parties. I'll not give any deeper analysis here of the factors that might make these costs of inquiry too high, but simply specify that in the range of cases of socially problematic disagreements that we address from here on out, where there is supposed to be an enhanced propriety to doxastic adjustment, there must be sufficient reasons in the background to think that inquiry may be fruitful, and that not too much polarization will occur. It should be recalled that I specified earlier that doxastic adjustment involves *equitable* inquiry (rather than inquiry aimed merely to vindicate one's original position), and I expect that when parties are sincerely committed to such inquiry, there will be at least some non-trivial range of cases where they can avoid the potentially polarizing effects of inquiry.

With the above discussion in mind, we can now give a characterization of the conditions under which parties to a socially problematic disagreement are most likely to find themselves with compelling reasons to doxastically adjust, and not *just* agree to a reasonable compromise. Doxastic adjustment will be more apt to be *part* of a fully

adequate response to a socially problematic disagreement in this domain of reasonable compromise to the degree that (i) the costs of parallel processing are relatively high, and (ii) the costs of inquiry are relatively low.

C. The Need for Doxastic Adjustment in the Domain of Justified Coercion

We have seen that there are limits to how far the purely practical approach can take parties to a socially problematic disagreement: even if they agree to discount or bracket their personal judgments so as to abide by the terms of a reasonable compromise, the separation costs associated with parallel processing persist, and may give parties reason to doxastically adjust. The purely practical approach hits an even more severe limit in another domain, though, which I'm calling the domain of "justified coercion." (For reasons we'll shortly see, this label should be taken lightly, since in some of the cases in the domain the "coercive" element amounts only to an "uncompromising" resolve not to support or enable certain kinds of conduct.) In this subsection I describe the kinds of disagreements that fall into this domain of justified coercion, and suggest that here, too, practical compromise fails to adequately address relevant social costs to disagreement. These costs, then (I suggest) continue to give parties to disagreements in this domain added reasons for doxastic adjustment.

(i) The Domain of Justified Coercion

In the domain of *reasonable compromise*, the fact that one party's contested judgment is true, and the other's is not, does not significantly change how each ought to treat her judgment in the context of practical reasoning about how to act towards each other. Regardless of whether a given party's contested judgment is true (it was assumed), she is obliged to treat it in a certain discounted way in the course of deciding what to do; in

particular, she should bracket it in whatever way is required to allow her to abide by the terms of the reasonable compromise. The correctness of her judgment, thus, does not make her justified in impinging on the other party's self-determination or acting in a coercive way towards him.

Though I've granted that there are some cases of socially problematic disagreement that fall in the domain of reasonable compromise, not all do. This is because in some socially problematic disagreements, the cost of acting on an erroneous judgment are very high, or of particularly problematic sort. When this is so, compromise may not be justified. (It may not be justified, in particular, if acting on the terms of the compromise involve acting on the premise of the erroneous judgment and so incurring the relevant high costs of error.) Unlike in cases that fall in the domain of reasonable compromise, in these cases, *if* one party's contested judgment is true, and the other's is not, that *does* affect how each ought to treat her judgment in the context of practical reasoning about what to do. In these cases, the party whose contested judgment is correct may be justified in *not* setting aside her judgment, and in instead acting in a way that is consistent with it—even if that involves some degree of impingement on the other party's self-determination or (even) acting in a coercive way towards him. These are cases of socially problematic disagreements that fall in the *domain of justified coercion*.

There are two primary kinds of cases of socially problematic disagreements that fall in this domain that I will attend to. As in my previous discussion of cases, I will suggest that the separation costs of disagreement in these cases contribute to added reasons for doxastic adjustment. However, here I will begin to introduce the ways in which these separation costs can relate to the costs of error, so as to *together* generate added reasons for doxastic

adjustment. As specified in the last subsection, I will also assume that in all the cases I consider, the costs of doxastic adjustment are not too high.

(ii) Doxastic Adjustment and Disagreements about Justice

One prominent kind of socially problematic disagreement that can sometimes fall in the domain of justified coercion is disagreements about justice. I take it to be plausible that if a party to a disagreement about justice has the correct judgment about justice, she may sometimes be justified in taking coercive action against another party who disagrees with her.³² In these cases, of course, a purely practical approach that recommends reasonable compromise won't have much to offer her by way of guidance in how to respond to the disagreement. However, it seems that she *will* have added reasons for doxastic adjustment. And, in particular, it seems that the separation costs incurred by the disagreement—that is, in this case, the threat of her coercive action—contributes to her reasons for doxastic adjustment. Consider these cases:

Justice Disagreement: Jasmine and Dez get an invitation to a speech from a proponent of Religious Regime X. The speaker will argue that it is important to the good of society that persons dress conservatively, and so justifiable to legally enforce a dress code by charging moderate fines for infractions (comparable to parking tickets). Dez says to Jasmine, “Maybe we should go. Maybe their policy is not so bad; maybe it’s just a different way of doing things.” Jasmine answers, “No thanks—I already know the answer to that. That kind of policy is unjust. I’ve been doing a lot of reading in my political philosophy courses and the harm principle shows that legal enforcement of a dress code is an unjust infringement on people’s liberty. No one should be legally forced to abide by such a conservative dress code.”

³² A great deal more would need to be said about exactly under what conditions she would be so justified—obviously, for instance, the disagreeing party would need to be poised to act on his incorrect judgment about justice in a way that would cause trouble—but I won’t specify all the details; I’ll simply assume some such action may in some cases be justified for her.

Here, Jasmine is plausibly justified in settling on (*J*) *It is unjust to legally enforce a conservative dress code*. She does not have compelling reason to further reserve judgment about *J* or reconsider and inquire about the possibility that $\neg J$ and may in this sense remain steadfastly entrenched in her judgment that *J*. But now consider whether she is equally justified in such steadfast entrenchment on *J* in a socially problematic version of the same disagreement, where she is poised to act on *J* in a way that would have a coercive effect on those she disagrees with:

Social Justice Disagreement: Jasmine's country takes over a city that previously belonged to Religious Regime X in a just war and there is a question about what to do with their dress code policy. Many inhabitants of the city are ambivalent about the regime change, and they are overwhelmingly keen to keep the dress code legally enforced. The issue is up for a general vote and Jasmine and Dez get an invitation to a speech in which the speaker will argue that it is important to the good of society that persons dress conservatively, and so justifiable to legally enforce a dress code by charging moderate fines. Dez says to Jasmine, "Maybe we should go. Maybe their policy is not so bad; maybe it's just a different way of doing things." Jasmine answers, "I've been doing a lot of reading in my political philosophy courses that suggests that legal enforcement of a dress code is unjust. But you're right—after all, we are going to be voting on the fate of the inhabitants pretty soon. We had better go listen to what they have to say in defense of their view before making a final call about this."

Plausibly, *if* Jasmine is right about *J*, she would be justified in voting against the dress code policy, even though that would infringe on the self-determination of the inhabitants of the city. That said, Jasmine is plausibly *not* justified in remaining steadfastly entrenched on *J* in *Social Justice Disagreement* in the way she very well might have been in *Justice Disagreement*. This seems to be at least in part because of the separation costs that are now at stake: it is because she is now poised to act on *J* in the context of a disagreement where her action would infringe on the self-determination of others, that she has added reasons to reserve judgment, or at least reconsider it, while she engages in further inquiry about the

question of *whether J.* More specifically, and as we'll say much more about in the next section, part of what seems to be going on here is an interesting interplay between the costs of error and the social separation costs of the disagreement Jasmine is part of. The costs of acting on erroneous judgment have risen significantly for Jasmine *in part because* of the separation costs at stake: if she acts on an erroneous judgment about the justice of the dress code policy, there is now a great deal more to be lost *because* she will have incurred the relevant separation costs (infringement on another's self-determination), *and incurred them for nothing*. It is as a result more important in *Social Justice Disagreement* than it is in *Justice Disagreement* that she be cautious about the formation of her judgment.

(iii) Doxastic Adjustment and Disagreements about Morality

Another prominent kind of socially problematic practical disagreement with a similar normative profile is disagreements about important matters of morality. To distinguish the cases I have in mind from the kinds of cases discussed in the last subsection, I'll focus on matters of morality that are important, but not important in the way that they would need to be to justify the kind of flat-out coercion that might be justified when matters of justice are at stake. I take it to be plausible that even among disagreements about these kinds of matters of morality, if a party to a disagreement—call her Rosita—has the correct judgment, she may sometimes be justified in “uncompromisingly” not supporting or enabling the conduct of those who disagrees with her.³³ In these cases, once again, a purely practical approach that recommends compromise will have little to offer someone like Rosita. However, as with Jasmine before, it seems that someone like Rosita will have

³³ A great deal more would need to be said about exactly under what conditions she would be so justified—obviously, for instance, the disagreeing party would need to be poised to act on his incorrect

added reasons for doxastic adjustment. And, in particular, it seems that the separation costs incurred by the disagreement—that is the costs incurred by others by her lack of support or enablement—contributes to her reasons for doxastic adjustment. Consider these cases:

Moral Disagreement: Though Rosita believes late term abortions shouldn't be politically legislated, she does believe they are gravely wrong and that, except when the mother's life is at stake, non-medical individuals have a moral obligation not to facilitate or support such abortions. She has considered the issue thoughtfully, and got an A+ on the final paper in her advanced ethics course in which she argued for this view. She becomes pregnant, and late in the pregnancy, finds herself in a difficult situation. She realizes that having an abortion would make her significantly better poised to flourish in her later adult life. In a discussion group, a fellow community member whose character and competence in moral reasoning Rosita deeply respects says she's been thinking about this issue and has come to the conclusion that there are a fair range of cases in which late term abortions are morally permissible and (further) that in these cases you should lend help if you can. She knows Rosita is hard up for money and offers to pay for the abortion. "What do you say, Rosita?" she asks. "Shall I offer you some help?" Rosita responds that on the basis of her own careful thinking about the matter, she disagrees; she is confident that late term abortions (and enabling support of them) is immoral. She will not be having an abortion, and declines the offer of help.

In *Moral Disagreement*, Rosita is plausibly justified in settling on (*M*) *Late term abortion is immoral and should not be enabled*. Even though a respected fellow community member disagrees with her, she has considered the issue carefully enough to justify her standing steadfast on *M*. But now consider whether she is equally justified in such steadfast entrenchment on *M* in a socially problematic version of the same disagreement, where she is poised to act on *M* in a way that affects not herself, but rather someone else who she disagrees with:

Social Moral Disagreement: Years later, Rosita takes an adult education course and argues for the same view she did in college; once again, she gets an A+. Then Rosita finds out her college-aged daughter is pregnant. Rosita deeply respects her

judgment about morality in a way that caused trouble—but I won't specify all the details; I'll simply assume some such action may in some cases be justified for her.

character and competence in moral reasoning. Her daughter says she's found herself in a difficult situation, and she realizes that having an abortion would make her significantly better poised to flourish in her later adult life. She has been thinking carefully about this issue and has come to the conclusion that there are a fair range of cases, including cases like her own, in which late term abortions are morally permissible. She really thinks that it is the best thing for her to do, but she doesn't have money to afford it. She would need Rosita to pay. "What do you say, Mom?" she asks. Rosita is not hard up for money and has always been willing to help out monetarily in difficult situations with her children in the past. She is taken aback, but responds, "I've thought for a long time that it is immoral and that people like me should not enable it. But now I feel like I need to hear more about what you're thinking and reevaluate before making up my mind about what to think about what I should do here."

Plausibly, *if* Rosita is right about *M*, she would be justified in not paying for her daughter's abortion. Still, Rosita plausibly gains some reasons in *Social Moral Disagreement* not to remaining steadfastly entrenched on *M* in the way she very well might have been in *Moral Disagreement*. Like before, this seems to be at least in part because of the separation costs that are now at stake. It is because she is now poised to act on *M* in the context of a disagreement where her action would work against the full self-determination of her daughter, that she has added reasons to reserve judgment, or at least reconsider it, while she engages in further inquiry about the question of *whether M*. Again, part of what seems to be going on is that the cost of acting on erroneous judgment has risen for Rosita *because* of the separation costs at stake: if she acts on an erroneous judgment about the morality of late term abortion here, there is now more to be lost than there was before *because* she will have incurred the relevant separation costs (working against her daughter's self-determination), *and incurred them for nothing*. It is as a result more important in *Social Moral Disagreement* than it is in *Moral Disagreement* that she be cautious about the formation of her judgment.

In this section, I've considered how a purely practical approach to socially problematic disagreements fares, and seen that it suffers from two important limits. First, within the domain of reasonable compromise, it incurs separation costs in the form of costs of parallel processing. I suggested that if these costs are relatively high and the costs of doxastic adjustment—and in particular, of continued *inquiry*—are relatively low, there may well still be compelling reasons for parties to doxastically adjust in response to the disagreement. Second, in the domain of justified coercion, I suggested that separation costs cannot justifiably be addressed by compromise because the costs of error are too high. Nevertheless, significant separation costs still exist, and—it appears—they may continue to give parties to socially problematic disagreements in this domain strong reasons to doxastically adjust.

I have now indicated that, even taking into account the possibility of practical adjustments in response to disagreement, the practical considerations in socially problematic disagreement may still contribute to additional reasons for parties to doxastically adjust. Having shown *that* they plausibly do contribute in this way, I now go on in the next section to discuss *how* they may make this contribution.

IV. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment (Or, Why the Truth Matters to Socially Problematic Disagreements)

In this section, I'll discuss two ways in which practical considerations related to separation costs may give parties to a socially problematic disagreement reasons for doxastic adjustment—one “truth-motivated” and one “truth-indifferent.” I'll suggest that those practical considerations—which I'll simply call “separation considerations”—may be able to function as either truth-motivated or truth-indifferent reasons; however, for reasons

given I will be particularly concerned to show that they can serve in the former capacity, as truth-motivated reasons for doxastic adjustment. Before getting to these points, I will also consider the objection that practical considerations related to p cannot genuinely have any bearing on whether one should suspend judgment with respect to p . This is not an objection I wish to spend too much time on, since a full examination of the objection would take us too far afield. Still, I'll briefly address this objection in Part A. I then move on to examine questions about truth-motivated and truth-indifferent reasons for doxastic adjustment in Parts B and C.

A. Can Practical Considerations Give One Reasons to Withhold?

One question that arises when one considers how separation considerations may give rise to reasons for doxastic adjustment with respect to p is the question of whether those considerations can genuinely constitute reasons for withholding judgment about p . Philosophers have long been concerned about the question of whether practical considerations can constitute reasons for *belief*, and the same kinds of arguments that motivate the claim that practical considerations *cannot* constitute reasons for belief might also be used to argue that those considerations cannot constitute reasons for withholding. Take, for instance, Hieronymi (2005, 2006, 2008, 2009a, 2009b). Briefly, Hieronymi argues that to believe p is to settle the question of *whether p* in the affirmative, and that practical considerations cannot serve as reasons for believing p because they do not bear on the question of *whether p*. A similar worry could be raised for withholding about p . Perhaps withholding about p is a matter of settling the question of *whether p*, not in the affirmative, but (so to speak) “in the neutral.” If this is what withholding about p amounts to, and if Hieronymi is right that practical considerations can't serve as reasons for

believing *p* because they do not bear on the question of *whether p*, then it looks like it will also follow that they cannot serve as reasons for withholding about *p*. For withholding about *p* will just be a matter of having a different kind of answer to the question of *whether p*, and—if Hieronymi is right—practical considerations don't bear on the question of *whether p*.

I won't be too concerned with this objection about whether practical considerations can give reasons for withholding for two reasons. First, the above worry is based on extending one kind of account of believing and of practical considerations' relation to believing, to the case of withholding—but the former account is contested.³⁴ I do not wish to enter that complicated debate here, and so—if necessary—am willing to simply stipulate as an assumption that practical consideration can serve as reasons for belief in whatever way it might be necessary for them to, to also serve as reasons for withholding.³⁵ Second, and more importantly, I doubt that the argument against thinking that practical considerations can serve as reasons for *belief* can successfully be extended to an argument against thinking that practical considerations can serve as reasons for *withholding*.³⁶ This is because—to put things in a way that engages with Hieronymi's framework—I doubt that withholding about *p* is always a matter of settling the question of *whether p* in some “neutral” way. Certainly that may be one way of withholding about *p*, but it strikes me that there is a distinct way of withholding about *p* that is just a matter of holding off on settling the question of *whether p* at all. To withhold about *p* in this sense may involve just settling

³⁴ See Footnote 13.

³⁵ I am also quite happy to say that reasons for doxastic adjustment simply figure as reasons for belief management (exposing yourself to the right kinds of arguments, evidence, etc.) aimed at the relevant kind of doxastic adjustment.

³⁶ For another view related to why this may not work, see Schroeder (2012).

the distinct question, *Shall I make a judgment about p yet?*, and *this* is a question on which practical considerations clearly seem to have a bearing.³⁷ When practical considerations give subjects in socially problematic disagreements reason for reserving judgment about a contested matter, then, I will assume that they do so in this latter way. They give subjects reasons to hold off on making a decisive judgment about the contested matter, rather than by giving them reasons to make a decisive judgment in a particular “neutral” way.³⁸

I will now set aside the question of whether practical considerations can genuinely constitute reasons for withholding. I will assume that they can, and focus my attention on an alternative question that is more interesting by my lights: when a separation consideration gives a party to a socially problematic disagreement reason to doxastically adjust (whether by reserving judgment and inquiring further, or reconsidering judgment and inquiring further), how does that reason relate to the value of truth?

B. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment, Part 1: Separation Considerations as *Truth-Indifferent* Reasons for Doxastic Adjustment

In principle, there are (at least) two ways that a separation consideration *c* might function as a reason for a party to a socially problematic disagreement *S* to doxastically adjust. First, *c* might function as a reason for *S* to doxastically adjust because it gives her reason to avoid

³⁷ See Friedman (2017, forthcoming) for very interesting work on the nature of suspension of judgment and belief.

³⁸ Consider a juror before and after a trial. After the trial, and so after arguments and evidence have been presented to her, she may deliberate about the question of *Is the defendant is guilty?* and settle this question neutrally—she may definitively judge that what arguments and evidence bear on the question aren’t strong enough to settle it in the affirmative or in the negative. This will be one way in which she could come to withhold judgment on the question. But it strikes me that there may be something interestingly different about this kind of withholding judgment on the question that the juror does after the trial, and the kind of withholding of judgment she may do before the trial, where she simply denies having an opinion on the matter of whether he is guilty or innocent at all. Parties to a socially problematic disagreement make simply deny having an opinion, on the grounds that they realize that there is still a great deal of evidence that they

having *a judgment that conflicts with her interlocutor's judgment*.³⁹ Supposing she initially believes *p*, and then finds out that her interlocutor believes an incompatible proposition *q*, *c* gives her a reason to simply avoid having that judgment *p*. In this case, the fact that *c* gives S a reason to doxastically adjust does not depend on any further fact about the importance of getting at the truth: it is enough just that S avoids disagreeing with her interlocutor, regardless of who is right or what the underlying facts about the contested matter are. So, I will say that when a separation consideration operates as a reason for doxastic adjustment in this way, it functions as a *truth-indifferent* reason for doxastic adjustment. A second way in which a separation consideration *c* might give S a reason to doxastically adjust is by giving her reason to avoid having a *false judgment*—in other words, because it gives her reason to believe truly, if she believes at all. In this case, the fact that *c* gives S a reason to doxastically adjust does depend on there being some further fact about it's being important for her to have a true belief, if she has a belief at all. So, I will say that when *c* operates as a reason for doxastic adjustment in this way, it functions as a *truth-motivated* reason for doxastic adjustment.

On first glance, it appears that when a separation consideration gives a party to a socially problematic disagreement, S, reason to doxastically adjust, that consideration must function as a truth-indifferent reason. This is because whether or not two parties to a socially problematic disagreement incur separation costs does not depend on whether either of their opposing beliefs *p* or *q* is *true*; instead, it depends wholly just on whether *they have* those opposing beliefs. Thus, for instance, regardless of whether Taylor or Tom

have not looked at. Of course, whether they can do so will depend on whether that fact about their relationship to the evidence actually pertains.

³⁹ This is how I originally introduced these considerations as functioning in Section One.

is right about Truman's motives, so long as they *have* the views they do, the social separation costs of disagreement will be incurred (at least one of them will not exercise full self-determination, for instance). Since these separation costs will be incurred regardless of who is right, it appears that they do not give (e.g.) Taylor any added reason that she didn't already have before the disagreement to avoid making a *false* judgment about Truman's motives. Rather, it appears that those costs' primary function will be to give her added reasons just to avoid having the judgment she does—for it is by avoiding having that judgment that it seems she will most effectively stave off the incurrence of separation costs. As a result, it seems that separation considerations give parties to a socially problematic disagreement *truth-indifferent* reasons for doxastic adjustment. In so far as they act on those reasons and doxastically adjust on the basis of them, the parties to a socially problematic disagreement will be motivated in a way that is indifferent to the truth about the contested matter at hand.

I don't deny that separation considerations may sometimes give a party to a socially problematic disagreement, S, truth-indifferent reasons for reserving judgment. If this is the only capacity in which those considerations can operate as reasons for doxastic adjustment, however, those reasons will in many cases not go far enough to *justify* S in doxastically adjusting in response to the disagreement. In particular, consider socially problematic disagreements in the domain of coercion, where the cost of error is very high. We considered disagreements about, for instance, matters of justice and important matters of morality, and in these disagreements the cost of acting on a false judgment would mean perpetrating or enabling either an injustice or otherwise gravely immoral action. Since our judgments generally serve as guides to action, the fact that it is very important for S to

avoid *acting* on a false judgment in these contexts will also generally make it the case that it is important that she avoids *having* a false judgment. As a result, in disagreements in this domain, it is important that S regulate her doxastic life in a way that helps her avoid having false judgments. But if S adopts a separation consideration as a truth-indifferent reason for doxastic adjustment, she will generally not do well at achieving the goal of avoiding false judgments, since her adjustments will be aimed primarily at helping her avoid coming to a different conclusion about things than her interlocutor, rather than at helping her avoid erroneous judgment.

Further—and apart from the *consequences* of S's adopting a separation consideration as a truth-indifferent reason for doxastic adjustment—it seems that by adopting that consideration as a truth-indifferent reason for doxastic adjustment, she will be displaying a kind of disloyalty to the truth that is objectionable in the context of such important questions as questions of justice and morality. Consider again for instance Rosita. We said in the previous section that in the socially problematic version of her case where there were separation costs (*Social Moral Disagreement*), she appeared to have added reasons for doxastic adjustment that she did not have in the non-socially problematic version of her case where there were not separation costs (*Moral Disagreement*). That is, it seemed Rosita had added reason to reserve or reconsider her judgment about the moral impermissibility of late-term abortion and to reflect on what her daughter had to say about its permissibility, in virtue of the fact that there were now separation costs at stake for her in making her judgment, like (e.g.) impinging on her daughter's self-determination. Though it *seemed* that she had added reasons for doxastic adjustment, however, we can also easily imagine someone criticizing her for that adjustment. Perhaps Rosita's husband, who has long been a

fellow opponent of late-term abortions, finds out she is reconsidering her view in light of her daughter's request. We can imagine him condemning her for this, on the grounds that she is (purportedly) "caving" to social pressure, trying to make things easier for herself and to ease the uncomfortableness of disagreement, while neglecting the grave importance of the truth they both agreed to just before they found out about their daughter's situation. Though it initially seemed that Rosita had good reasons for doxastically adjusting, if Rosita's doxastic adjustment is a response to separation considerations, considered as *truth-indifferent* reasons for doxastic adjustment, the content of these criticisms from her husband would be more or less on target.⁴⁰ For she would indeed be doxastically adjusting to ease the social difficulties associated with the disagreement, and (to that degree) setting aside the significant value of the truth about the moral permissibility of late-term abortion. In this case and in others like it—where the relevant socially problematic disagreement bears on fundamentally important matters whose truth cannot be easily set aside—it looks like subjects who doxastically adjust on the basis of a truth-indifferent reason may sometimes be rightly criticizable for their adjustment. (In this way, it seems that in the domain of justified coercion, parties who set aside concerns about the truth in the regulation of their *doxastic* life are not all that different from parties who set aside concerns about the truth in the regulation of their *practical* life.) The upshot of this discussion is that, even if separation considerations can sometimes provide truth-indifferent reasons for doxastic adjustment, in the context of socially problematic disagreements where there is a very high cost of error, those considerations may fail to provide *enough* reason to *justify* a subject in doxastically adjusting.

⁴⁰ We can set aside whether it was his place to actually communicate those criticisms.

C. Truth-Motivated and Truth-Indifferent Reasons for Doxastic Adjustment, Part 2: Separation Considerations as *Truth-Motivated Reasons* for Doxastic Adjustment

The conclusion reached in the last subsection is problematic, since in a case like Rosita's, it *doesn't* appear that she must come out as being unjustified in her doxastic adjustment.

When she expresses willingness to reconsidered her view and engage in further inquiry with her daughter, it doesn't seem that what she is doing has to be understood in terms of "caving" or playing fast and loose with the truth. Indeed—at least as the case strikes me—it seems to be just the contrary: she seems to be doing something commendable and is being intellectually responsible rather than intellectually lazy or fickle.

How do we explain this? If we want to understand Rosita's doxastic adjustment as something other than "caving" to social pressure—if we want it to come out as commendable (intellectually responsible, or the like)—we will need a way to understand the separation considerations she is doxastically adjusting in response to as truth-motivated reasons. For, if they can be understood in this way, her doxastic adjustment can be understood as motivated by respect for the significant importance of the relevant truth that is being contested. And indeed—as I'll now argue—it looks like those separation considerations can after all be understood as truth-motivated reasons for doxastic adjustment.

Recall what we said in the analysis of Rosita's case in *Social Moral Disagreement*. I noted that what seems to be going on is that the costs of acting on erroneous judgment had risen for Rosita *because* of the separation costs at stake. In particular, I noted that *if she acts on an erroneous judgment about the morality of late term abortion*, there is more to be lost than there was before she encountered the separation costs at hand, because if she acts on that erroneous judgment now, she will incur those separation costs—e.g. working against

her daughter's self-determination—for *nothing*. In this way, if she acts on an erroneous judgment she will incur a non-trivial moral cost for someone else for the sake of her own false ideal, and so without any compensating good to show for it. This is worse than just acting on an erroneous judgment; indeed it is plausibly even worse than just acting on an erroneous judgment that incurs some non-trivial moral cost for herself, as when she decides not to have a late term abortion of her own in *Moral Disagreement*.

This understanding of what is going on in *Social Moral Disagreement* allows us to understand separation considerations as operating as *truth-motivated reasons* for doxastic adjustment. For consider: the consideration (*c*) *That separation costs will be incurred* now operates to heighten the costs of error, and it gives Rosita a reason to doxastically adjust *because* it heightens the cost of error. Because the costs of acting on an erroneous judgment are heightened, and because she will be acting on whatever judgment she comes up with on the matter, she now has enhanced reason to avoid having an erroneous judgment. That means that when Rosita doxastically adjusts in response to *c*—so understood in its relation to the costs of error—she is responding to *c* in its capacity as a reason to avoid having a false judgment. But this means that she is responding to *c* as a truth-motivated reason for doxastic adjustment. Rosita responds to her daughter's query not because she is playing fast and loose with the truth, but because she realizes all the more the momentousness of its importance, given the costs to her daughter of her getting things wrong.

We've now seen that separation considerations can contribute to truth-motivated reasons for doxastic adjustment in a case of disagreement like Rosita's, which falls under the domain of justified coercion. Before closing, it's worth also going back to cases of disagreement that fall in the domain of reasonable compromise, like the case of Taylor and

Tom in *Collective Opinion*. On first glance, one might think that separation considerations can only figure as truth-*indifferent* reasons for doxastic adjustment in this domain, because of how differently the cost of error figures in disagreements in this domain. Is that so? To what extent do separation considerations contribute to truth-indifferent vs. truth-motivated reasons for doxastic adjustment in this domain? I will not answer this question fully, but offer the following two important observations.

First, a point about why it is more plausible that separation considerations may provide truth-*indifferent* reasons for doxastic adjustment in the domain of reasonable compromise, than in the domain of justified coercion. Recall from Section Three that we said that in disagreements in the domain of justified coercion, the cost of error may make it unjustified for parties to the disagreement to agree to a practical compromise. In cases that fall in the domain of reasonable compromise, by contrast, we observed that the cost of error is generally *not* so weighty as to make practical compromises unjustified (to the contrary, the fact that the cost of error is not as weighty seemed to have the upshot that parties are *obliged* to agree to a reasonable compromise, if it is forthcoming). At least on face—and I will not take time to explore this matter beyond surface appearances here—a parallel observation may be true with respect to the role of the cost of error in making parties to disagreements in the domain of reasonable compromise justified in adopting separation considerations as truth-indifferent reasons for doxastic adjustment. Just as the reduced cost of error makes practical compromise more justified in the domain of reasonable compromise, the reduced costs of error plausibly also make doxastic adjustment on the basis of truth-indifferent reasons more justified in the domain of reasonable compromise.

(Further work will need to be done to ascertain just how far this justification goes in different sorts of scenarios of reasonable compromise.)

This all still leaves the question of whether separation considerations, considered as *truth-motivated* reasons, can justify doxastic adjustment in the domain of reasonable compromise. It seems to me that they can. For, just as in the case of disagreements in the domain of justified coercion, disagreements in the domain of reasonable compromise have costs of error that can be augmented by separation costs. Consider again Taylor and Tom. Suppose they agreed to a reasonable compromise whereby they took down the political advertisements and split use of the car and other home resources. Suppose further that this compromise incurred significant separation costs in the form of costs of parallel processing for Tom and Taylor. *If* Taylor makes an erroneous judgment about Truman's motives, then—in parallel to what we observed in Rosita's case—these separation costs will have been incurred *for nothing*. Taylor and Tom will have gone through all they did with respect their compromise in part for the sake of her being able to act on a judgment that is *false* and so that she presumably would have been better off avoiding on that basis. For similar reasons as pertained in Rosita's case, the upshot of this is that Taylor can adopt separation considerations as *truth-motivated* reasons for doxastic adjustment. Like Rosita, Taylor may be motivated to doxastically adjust, not because she is unconcerned with the truth, but rather because she realizes it's importance: she realizes that if she is going to ask Tom to sustain the costs of parallel processing with her, she better make sure to have a good reason to do so.

V. Conclusion

In this paper, I have surveyed whether, why, and how practical considerations can make a difference to the kinds of doxastic attitudes that parties should take up towards contested claims in the context of socially problematic disagreements. I have argued that practical considerations *can* make a difference to the doxastic attitudes parties should take up in these contexts. In particular, I have argued that the social separation costs associated with these disagreements can give parties to those disagreements reasons for *doxastic adjustment*, where doxastic adjustment with respect to a judgment *p* involves reservation of one's judgment about *p* or reconsideration of *p*, and inquiry into the question of *whether p*. Moreover, I showed that parties who act on these separation cost-related reasons for doxastic adjustment need not be understood as thereby being indifferent to the truth. Rather, I suggested, it can be because they understand the momentous practical weight that is carried by claiming to have the truth in the context of these disagreements, that these subjects are swayed to show all the more humble of a respect for it.

Bibliography

- Adler, J. (2004). Reconciling open-mindedness and belief. *School Field*, 2(2), 127–142.
- Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.
- Arpaly, N. and Schroeder, T. (2013). *In Praise of Desire*. Oxford University Press.
- Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.
- Atran, S., Axelrod, R., & Davis, R. (2007). Sacred barriers to conflict resolution. *Science*, 317, 1039–1040.
- Baehr, J., (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford University Press.
- Beardsley, M. C. (1958). *Aesthetics*. Hackett.
- Benhabib, S. (1992) *Situating the Self: Gender, Community, and Postmodernism in Contemporary Ethics*. Routledge.
- Bennett, J. (1974). The conscience of Huckleberry Finn. *Philosophy*, 49(188), 123–134.
- Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford University Press.
- Brandom, R. (1998). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brennan, T. (2003). Stoic Moral Psychology. In B. Inwood (ed.), *The Cambridge Companion to the Stoics*. Cambridge University Press
- Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169(2), 285–311.

- Byers, S. C. (2012). *Perception, Sensibility, and Moral Motivation in Augustine: A Stoic-Platonic Synthesis*. Cambridge University Press.
- Calvert, J. (2010). *Sayyid Qutb and the Origins of Radical Islamism*. Hurst & Co.
- Christensen, D. (2007). Epistemology of disagreement: The good news. *The Philosophical Review*, 116(2), 187–217.
- (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, 4(5), 756–767.
- (2013). Epistemic modesty defended. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays* (pp. 77–97). Oxford University Press.
- Clarke, R., McKenna, M., and A. Smith. (2015). Introduction. In Clarke, R., McKenna, M., and A. Smith (Eds.), *The Nature of Moral Responsibility* (pp. 1–18). Oxford University Press.
- Clifford, William K. (1999/1877). The Ethics of Belief. In *The Ethics of Belief and Other Essays* (pp. 70–96). Prometheus Books.
- Conee, E. and Feldman, R. (2004). *Evidentialism: Essays in Epistemology*. Clarendon Press.
- Cvetkovich, G., & Nakayachi 1, K. (2007). Trust in a high-concern risk controversy: A comparison of three concepts. *Journal of Risk Research*, 10(2), 223–237.
- Dancy, J. (2004). Enticing reasons. In R. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (Eds.), *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*. Clarendon Press.
- Davidson, D. (1971). Agency. Reprinted in *Essays on Action and Events* (pp. 43–61). Clarendon Press.

- . (1982). Rational animals. *Dialectica*, 36(4), 317–327.
- Daniels, N. (2018). Reflective Equilibrium. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <<
<https://plato.stanford.edu/archives/spr2018/entries/reflective-equilibrium>>>.
- Dreben, B. (2003). On Rawls and political liberalism. In S. Freeman (Ed.), *The Cambridge Companion to Rawls* (pp. 316–346). Cambridge University Press.
- Edwards, J. R., & Cable, D. M. (2009). The value of value congruence. *Journal of Applied Psychology*, 94(3), 654.
- Elga, A. (2007). Reflection and disagreement. *Noûs* 41(3), 478–502.
- Enoch, D. (2015). Against public reason. In D. Sobel, P. Vallentyne, and S. Wall (Eds.), *Oxford Studies in Political Philosophy, Volume 1* (pp. 112–143). Oxford University Press.
- . (2010). Not just a truthometer: Taking oneself seriously (but not too seriously) in cases of peer disagreement. *Mind*, 119(476), 953–997.
- Estlund, D. M. (2009). *Democratic Authority: A Philosophical Framework*. Princeton University Press.
- Fantl, J. and McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford University Press.
- Feldman, R. (2007). Reasonable religious disagreement. In L. Antony (Ed.), *Philosophers without Gods: Meditations on Atheism and the Secular Life*. Oxford University Press.
- Ferry, M. (2013). Does morality demand our very best? On moral prescriptions and the line of duty. *Philosophical Studies*, 165(2), 573–589.

- Fischer, J.M. and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.
- Forst, R. (2013). More than a prehistory: Antiquity and the middle Ages. In *Toleration in Conflict: Past and Present* (pp. 36-95). Cambridge University Press.
- Friedman, J. (2017). Why suspend judging? *Noûs* 50(3), 302–326.
- . (Forthcoming). Inquiry and belief. *Noûs*.
- Freidman, M. (2000). John Rawls and the political coercion of unreasonable people. In V. Davion and C. Wolf (Eds.), *The Idea of a Political Liberalism: Essays on John Rawls* (pp. 16–33). Rowman & Littlefield.
- Fricker, M. (2007). *Epistemic injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Gardner, P. (1993). Should we teach children to be open-minded? Or, is the Pope open-minded about the existence of God? *Journal of Philosophy of Education*, 27(1), 39–43.
- Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.
- Gibson, J. (1977). The theory of affordances. Reprinted in R. Shaw and J. Bransford (eds.), *Perceiving, Acting, and Knowing* (pp. 127–149). Laurence Erlbaum.
- Gaus, G. (2003). *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project*. Sage.
- (2010). *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge University Press.
- (2015). On dissing public reason: A reply to Enoch. *Ethics*, 125(4), 1078–1095.

- Gutmann, A. and D.F. Thompson. (1996.) *Democracy and Disagreement*. Harvard University Press.
- Gutting, G. (1982). *Religious Belief and Religious Skepticism*. University of Notre Dame Press.
- Haidt, J., Rozin, P., McCauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology & Developing Societies*, 9(1), 107–131.
- Hare, W. (2011). Helping open-mindedness flourish. *Journal of Thought*, 46(1-2), 9–20.
- Hieronymi, P. (2005). The wrong kind of reason. *The Journal of Philosophy*, 102(9), 437–457.
- (2006). Controlling Attitudes. *Pacific Philosophical Quarterly*, 87(1), 45–74.
- (2008). Responsibility for Believing. *Synthese*, 161(3), 357–373.
- (2009a). Two Kinds of Agency. In Lucy O’Brien & Matthew Soteriou (Eds.), *Mental Actions* (138–162). Oxford University Press.
- (2009b). Believing at Will. *Canadian Journal of Philosophy*, 39, 149–187.
- Hampton, J. E. (1998). *The Authority of Reason*. Cambridge University Press.
- Horgan, T. and M. Timmons (ms). *Illuminating reasons*.
- Horgan, T. and J. Tienson. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 520–33). Oxford University Press.
- Horgan, T., J. Tienson, and G. Graham. (2004). Phenomenal intentionality and the brain in a vat. In R. Schantz (Ed.), *The Externalist Challenge: New Studies in Cognition and Intentionality* (pp. 297–317). Walter de Gruyter.

- Horgan, T. and G. Graham. (2013). Phenomenal intentionality and content determinacy. In R. Schantz (Ed.), *Prospects for Meaning* (p. 321–344). de Gruyter.
- Inwood, B. (1985). *Ethics and Human Action in Early Stoicism*. Oxford University Press.
- Irwin, B. (2001). Is this the man who inspired Bin Laden? *The Guardian* (November 1, 2001). Retrieved from:
 <<<http://www.guardian.co.uk/world/2001/nov/01/afghanistan.terrorism3>>>.
- Jackson, E. (Forthcoming). Belief and credence: Why the attitude-type matters. *Philosophical Studies*.
- Jackson, F. and Pettit, P. (1995). Moral Functionalism and Moral Motivation. *The Philosophical Quarterly*, 45(178), 20–40.
- Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne & T. Gendler (eds.), *Oxford Studies in Epistemology*, Volume 1 (167-196). Oxford University Press.
- (2010). Peer disagreement and higher order evidence. In A.I. Goldman & D. Whitcomb (eds.), *Social Epistemology: Essential Readings* (pp. 183-217). Oxford University Press.
- (2013). Disagreement and the burdens of judgment. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays* (pp. 31–53). Oxford University Press.
- Korcz, K.A. (2015). The epistemic basing relation. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from:
 <<<https://plato.stanford.edu/archives/fall2015/entries/basing-epistemic/>>>.

- Kriegel, U. (2008). Moral phenomenology: Foundational issues. *Phenomenology and the Cognitive Sciences*, 7(1), 1–19.
- Kruglanski, A. and Webster, D. (1996). Motivated Closing of the Mind: ‘Seizing’ and ‘Freezing.’ *Psychological Review* 103, 264–83.
- Lackey, J. (2010a). A justificationist view of disagreement’s epistemic significance. In A. Haddock, A. Millar, and D. Pritchard (Eds.), *Social Epistemology* (pp. 298–325). Oxford University Press.
- (2010). What should we do when we disagree? In J. Hawthorne & T. Gendler (Eds.), *Oxford Studies in Epistemology, Volume 3* (pp. 274–293). Oxford University Press.
- Leary, S. (2017) In defense of practical reasons for belief. *Australasian Journal of Philosophy*, 95(3), 529–542
- Lehrer, K. & Paxson Jr, T. (1969). Knowledge: Undefeated justified true belief. *Journal of Philosophy* 66 (8), 225–237.
- Leland, R. J., and van Wietmarschen, H. (2012). Reasonableness, intellectual modesty, and reciprocity in political justification. *Ethics*, 122(4), 721–747.
- Levy, I. (2016). Liberty of conscience and freedom of religion in the medieval canonists and theologians. In Shah and Hertzke (Eds.), *Christianity and Freedom, Volume 1*, Chapter 6. Cambridge University Press.
- Loar, B. (2003). Phenomenal Intentionality as the Basis of Mental Content. In M. Hahn and B. Ramberg (Eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (pp. 229–58). MIT Press.

- Macedo, S. (2000). In defense of liberal public reason: Are abortion and slavery hard cases? In R.P. George and C. Wolfe (Eds.), *Natural Law and Public Reason* (p. 11–50). Georgetown University Press.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Penguin.
- Mandelbaum, M. (1955). *The Phenomenology of Moral Experience*. The Free Press.
- Manne, K. (2013). On being social in metaethics. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics, Volume 8* (pp. 50–73). Oxford University Press.
- March, A. F. (2011). *Islam and Liberal Citizenship: The Search for an Overlapping Consensus*. Oxford University Press.
- Markovits, J. (2010). Acting for the right reasons. *The Philosophical Review*, 119(2), 201–242.
- McCormick, M. S. (2014). *Believing against the evidence: Agency and the ethics of belief*. Routledge.
- McGrath, S. (2008). Moral disagreement and moral expertise. *Oxford Studies in Metaethics, Volume 3* (pp. 87–108). Oxford University Press.
- Mele, A. (1995). *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.
- Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, 33(2), 243–258.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT press.
- (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT press.

- Musallam, A. (2005). *From Secularism to Jihad: Sayyid Qutb and the Foundations of Radical Islamism*. Praeger.
- Oddie, G. (2005). *Value, Reality, and Desire*. Oxford University Press.
- Peacocke, C. (1992). *A Study of Concepts*. MIT Press
- Pitt, D. (2004). The phenomenology of cognition, or, what is it like to think that 'p'? *Philosophy and Phenomenological Research*, 69(1), 1–36.
- Plantinga, A. (1993). *Warrant: The Current Debate*. Oxford University Press.
- (1995). *A defense of religious exclusivism*. In Senor (Ed.), *Rationality of Belief and Plurality of Faith*, (pp. 191–215). Cornell University Press.
- (2000). *Warranted Christian belief*. Oxford University Press.
- Prinz, J. (2015). An empirical case for motivational internalism. In G. Bjornsson, C. Strandberg, R. Ollinder, J. Eriksson, and F. Bjorklund (eds.), *Motivational internalism* (61–84). Oxford University Press.
- Quong, J. (2011). *Liberalism without Perfection*. Oxford University Press.
- . (2018). Public reason. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <<<https://plato.stanford.edu/archives/spr2018/entries/public-reason>>>.
- Qutb, Sayyid. (1953/2000). *Social Justice in Islam*. J.B. Hardie and H. Algar (Trans). Islamic Publications International.
- . (n.d./1983). *Islam and Universal Peace* (3rd ed). (Translator unknown). American Trust Publications.

- . (n.d.) *In the Shade of the Quran*. A. Salahi (Trans). Retrieved from
 <<https://archive.org/stream/InTheShadeOfTheQuranSayyidQutb/Volume_9_surah_s_10-11_djvu.txt>>.
- . (n.d/1991). *The Islamic Concept and its Characteristics*. M. Moinuddin Siddiqui, (Trans.) American Trust Publications.
- . (n.d.) *This Religion of Islam*. (Translator unknown). International Islamic Publishing House.
- . (n.d.). *Milestones*. (Translator unknown). Dar Al-Ilm.
- . (n.d.). *Islam: The Religion of the Future*. (Translator unknown). Al Faisal Press.
- Railton, P. (1986). Moral Realism. *The Philosophical Review*, 95(2), 163–207.
- (2009). Practical competence and fluent agency. In D. Sobel and S. Wall (eds.), *Reasons for Action* (pp. 81–115). Cambridge University Press.
- (2014). The affective dog and its rational tale: Intuition and attunement. *Ethics*, 124(4), 813–859.
- (2016). Intuitive guidance. In M.E.P. Seligman, P. Railton, R. F. Baumeister, and C. Sripda *Homo Prospectus* (pp. 33–86; pp. 87–102). Oxford University Press.
- (ms). Toward a unified account of rationality in belief, desire, and action.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- . (1993). *Political Liberalism: Expanded Edition (with “Reply to Habermas” and “The Idea of Public Reason Revisited”)*. Columbia University Press.
- . (2001). *Justice as Fairness: A Restatement*. Harvard University Press.
- Raz, J. (1990). Facing diversity: The case of epistemic abstinence. *Philosophy & Public Affairs*, 19(1), 3–46.

- Riggs, W. (2010). Open-mindedness. *Metaphilosophy*, 41(1-2), 172–188.
- Rinard, S. (Forthcoming). Believing for Practical Reasons. *Nous*.
- Ross, J., & Schroeder, M. (2014). Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, 88(2), 259–288.
- Ryan, T.J. (2017). No compromise: Political consequences of moralized attitudes. *American Journal of Political Science*, 61(2), 409–423.
- Schroeder, M. (2012). Stakes, withholding, and pragmatic encroachment on knowledge. *Philosophical Studies*, 160(2), 265–285.
- Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford University Press.
- Shah, N. (2003). How truth governs belief. *The Philosophical Review*, 112(4), 447–82.
- . (2006). A new argument for evidentialism. *The Philosophical Quarterly*, 56 (225), 481–98.
- Sliwa, P. (2016). Moral worth and moral knowledge. *Philosophy and Phenomenological Research*, 93(2), 393–418.
- Sobel, D., and Wall, S. (2009). Introduction. In D. Sobel and S. Wall (Eds.), *Reasons for Action* (pp. 1–12). Cambridge University Press.
- Sosa, E. (2015). *Judgment and Agency*. Oxford University Press.
- Siegel, S. (2014). Affordances and the contents of perception. In B. Brogaard (Ed.), *Does perception have content?* (pp. 39–76). Oxford University Press.
- Siegrist, M., Cvetkovich, G. T., & Roth, C. (2000). Salient value similarity, social trust, and risk/benefit perception. *Risk Analysis*, 20, 353–362.
- Siewert, C. (1998). *The Significance of Consciousness*. Princeton University Press.
- Simmons, A. J. (1981). *Moral Principles and Political Obligations*. Princeton University Press.

- Smith, M. R. (1994). *The Moral Problem*. Oxford University Press.
- Smith, A.M. (2012.) Attributability, Answerability, and Accountability: In Defense of a Unified Account. *Ethics*, 122(3), 575–589.
- Solum, L. (2004). Legal theory lexicon 028: Concepts and conceptions. Retrieved from <<<http://legaltheorylexicon.blogspot.com/2004/03/legal-theory-lexicon-028-concepts-and.html>>>
- Starkey, C. (2008). Emotion and full understanding. *Ethical Theory and Moral Practice*, 11(4), 425–454.
- Stout, J. (2004). *Democracy and Tradition*. Princeton University Press.
- Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2), 175-195.
- . (2007). Deliberating groups vs. prediction markets (or Hayek's challenge to Habermas). *Episteme: A Journal of Social Epistemology*, 3(3), 192-213.
- Teichman, J. (1975). Mr. Bennett on Huckleberry Finn. *Philosophy*, 50(193), 358–359.
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324.
- Tetlock, P. E., KIRSTEL, O. V., ELSON, S. B., GREEN, M. C., & LERNER, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853–870.
- Thomson, J.J. (2008). *Normativity*. Open Court.
- Twain, M. (1885/2004). *The Adventures of Huckleberry Finn*. Collector's Library.

- Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: individual differences in three functional domains of disgust. *Journal of personality and social psychology*, 97(1), 103.
- Vallier, K. (2014). *Liberal Politics and Public Faith: Beyond Separation*. Routledge.
- Wallace, T.J. (1994.) *Responsibility and the Moral Sentiments*. Harvard University Press.
- Wedgwood, R. (2001). Conceptual role semantics for moral terms. *The Philosophical Review*, 110(1), 1–30.
- . (2007). *The Nature of Normativity*. Oxford University Press.
- Weithman, P. (2010). *Why Political Liberalism? On John Rawls's Political Turn*. Oxford University Press.
- Whiting, D. (2014). Reasons for belief, reasons for action, the aim of belief, and the aim of action. In C. Littlejohn and J. Turri (Eds.), *Epistemic Norms: New Essays on Action, Belief, and Assertion*, pp. 219–38. Oxford University Press.
- van Inwagen, P. (1996). It is wrong, everywhere, always, and for anyone, to believe anything upon insufficient evidence. In E. Stump and M. J. Murray (Eds.) *Philosophy of Religion: The Big Questions* (pp. 273–84). Blackwell.